University of Padua

Department of Information Engineering
Master degree in Computer Engineering

Development of an EEG-based recurrent neural network for online gait decoding

Supervisor
Prof. Emanuele Menegatti

Candidate
Fabio Pancino

Co-Supervisor
Dr. Stefano Tortora

April 14, 2022

Academic Year 2021-2022

Abstract

Recent neuroscientific literature has shown that the use of brain-controlled robotic exoskeletons in walking rehabilitation induces neuroplasticity modifications, possibly leading to a higher likelihood of recovery and maintenance of lost motor functions due to a neural lesion, with respect to traditional rehabilitation. However, the gait decoding from brain signals remains an open challenge.

The aim of this work is to implement and validate a deep learning model for online gait decoding that exploits Electroencephalography (EEG) information to predict the intention of initiating a step, which could be used to trigger the assistance of a lower-limb exoskeleton. In particular, the model exploits a Gated Recurrent Units (GRU) deep neural network to handle the time-dependent features which were identified by analysing the neural correlates preceding the step onset (i.e., Movement-Related Cortical Potentials (MRCP)). The network was evaluated on a pre-recorded dataset of 11 healthy subjects walking on a treadmill. The network's architecture (e.g., number of GRU units) was optimized through grid search. In addition, to deal with the data scarcity problem of neurophysiological applications, I proposed a data augmentation procedure to increase the dataset available to train the model of each subject. With the proposed approach, the model achieved an average accuracy in detecting the step onset of $89.7 \pm 7.7\%$ with just the 15% of the

dataset for each subject (∼70 steps), and up to $97.8 \pm 1.3\%$ with the whole dataset (∼440 steps).

This thesis support the use of a memory-based deep learning model to decode walking activity from non-invasive brain recordings. In future works, this model will be exploited in real time as a more effective input for devices restoring locomotion in impaired people, such as robotic exoskeletons.

Contents

List of Figures

List of Tables

List of Acronyms

AI Artificial Intelligence.

ALS Amyotrophic Lateral Sclerosis.

AMICA Adaptive Mixture Independent Component Analysis.

ANN Artificial Neural Network.

ASR Artifacts Subspace Reconstruction.

BCI Brain Computer Interface.

BF Biceps Femoris.

BP Bereitschaftspotential.

BP' Backpropagation.

BPTT Backpropagation Through Time.

CAR Common Average Reference.

CSP Common Spatial Pattern.

DL Deep Learning.

ECoG Electrocorticographic.

EEG Electroencephalography.

EEP Epidural Electrical Stimulation.

EMG Electromyography.

EP Evoked Potential.

ERD Event-Related Desynchronization.

ERP Event-Related Potential.

ERS Event-Related Synchronization.

FES Functional Electrical Stimulation.

fMRI Functional Magnetic Resonance Imaging.

fNIRS Near-Infrared Spectroscopy.

GD Gradient Descend.

GRU Gated Recurrent Units.

HMI Human Machine Interface.

ICA Independent Component Analysis.

ICs Independent Components.

LDA Linear Discriminant Analysis.

LPP Locality Preserving Projection.

LSTM Long Short Term Memory.

MAFO Motorized Ankle-Foot Orthosis.

MEG Magnetoencephalography.

ML Machine Learning.

MoBI Mobile Brain/Body Imaging.

MRCP Movement-Related Cortical Potentials.

NN Neural Network.

NS Negativity Slope.

PET Positron Emission Tomography.

RNN Recurrent Neural Network.

SCI Spinal Cord Injury.

SCP Slow Cortical Potentials.

SSEP Steady State Evoked Potentials.

SSVEP Steady State Visual Evoked Potentials.

SVM Support Vector Machine.

TA Tibialis Anterior.

VEP Visual Evoked Potentials.

VM Vastus Medialis.

1 | Background

Scientific and technological progress is bringing us a fast-moving world where new scenarios are opening up every day. New applications are replacing previous ones or making feasible what seemed impossible only a few years ago. In the healthcare field [1], new technologies now embrace all areas and offer unexpected possibilities, supporting the figure of the doctor and allowing unprecedented home-care services.

In recent decades there has been a shift in the medical vision regarding rehabilitation: it has gone from a passive vision of the patient in which it is considered the object of intervention, to an active vision in which the patient becomes the protagonist of his own rehabilitation process.

In this perspective we talk about neurological rehabilitation, since, giving an active role to the patient means going to stimulate the process of neuroplasticity [2, 3, 4], the ability that the human brain has to reorganize itself in conditions of peripheral stimulation and learning through repeated stimuli in order to restore the injured part.

The scientific community's growing interest in the field of rehabilitation and healthcare is fuelled by the aim of improving the independence and quality of life of people with motor, sensory or cognitive deficits. Many of the studies in the literature focus on the rehabilitation of patients with degenerative diseases such as Parkinson's [5] or multiple sclerosis [6] or traumatic injuries

such as strokes [7], amputations [8] or spinal cord injuries [9].

We are witnessing increasingly promising results in assisted rehabilitation through the use of robotic devices, results that highlight the opportunity to contribute to the restoration of basic motor functions of people with motor deficits and which clearly were achieved after numerous unsuccessful attempts. However, what has emerged from the most recent reviews of robot-assisted rehabilitation is that this approach has the ability to increase the likelihood of recovery, maintenance, and optimization of the neural connections involved in a patient with respect to traditional methods [10, 11].

With the consolidation of research in the field of *intelligent robotics* and the idea of being able to improve the precision and results achievable in the field of healthcare, **neurorobotics** [12] was born.

Neurorobotics is a vibrant, active interdisciplinary field of neuroscience and engineering, in which knowledge of the human brain is combined with robotics and Artificial Intelligence (AI) [13]. A neurorobot refers to all those robotic devices that can interact with or emulate the nervous system of humans (or other animals). Neurorobotics can be considered interdisciplinary as we see it intersect from purely industrial disciplines (robotic arms) to others related to health (rehabilitation, surgery) or to entertainment applications (painting, video games).

Despite the numerous fields in which we see neurorobotics involved, the focus of this thesis work will be on devices capable of assisting walking movement. The main features of the neurorobots belonging to the walking assistance category are the ability to allow the patient to physically (or virtually) navigate the environment, either indoor or outdoor, and the capability to partially (lower-limb exoskeleton, leg prostheses) or totally support (powered wheelchair) their weight to alleviate fatigue due to possible disorders.

Among the many solutions for restoring natural gait are less invasive robotic exoskeletons used under the guidance of a licensed physiotherapist [14, 15, 16]. Lower limb exoskeletons can be categorised, according to [17], into three categories:

- *assistance exoskeletons*, are those that help users to complete daily activities that they are no longer able to do.
- *rehabilitation exoskeletons*, are designed to restore skills so that patients can live without the device. The goal of these exoskeletons is to help "when needed," to stimulate neuroplasticity through decreasing assistance gradually as the user improves.
- *augmentation exoskeletons*, are designed to increase the capabilities of generally healthy individuals, to provide support for tasks that would require above-average muscle strength.

In this thesis, I will focus mainly on the application of exoskeletons as rehabilitation devices, but the work can also be generalised to assistive or capacity-enhancing devices. Thus, for what concern the use of exoskeletons in rehabilitation, what emerges from the literature is the passive role that the patient assumes during the therapy process which decreases the neuromuscular activity and energy consumption, reducing rehabilitation efficacy [18]. On the contrary, to make rehabilitation a process through which the body is stimulated more at the muscular and neural level, we find rehabilitation devices based on Brain Computer Interface (BCI) [19, 20]. A BCI can be described as a direct communication channel between the brain and an external device without the involvement of motor processes. It therefore becomes clear that BCIs can be a valid means through which it is possible to have stimulation to restore the compromised functional capacity of damaged

neuronal connections. An approach based on this type of technology would make it possible to make the patient's role more active during the rehabilitation process, thus increasing its effectiveness [21, 22].
The work conducted in this thesis lies in the area of gait rehabilitation using BCI.

2 | Introduction

2.1 Brain Computer Interface

Thanks to the miniaturisation of increasingly powerful and low-cost computer equipment, and, given the ever-increasing need to help people with disabilities, we have seen the rise of the first **Human Machine Interface (HMI)**.
As the term implies, when we talk about HMI, we can imagine a communication channel between a person and a machine. For the sake of completeness, there are many types of HMI, but when communication is established between the brain and a computer, this type of interface is known as **Brain Computer Interface**.

The term BCI refers to a system capable of measuring the activity of the brain and translating it into instructions for various types of peripheral devices (communication systems, wheelchairs, prostheses and orthoses, etc.), creating a communication channel that disregards the normal neuromuscular outputs of the central nervous system [23, 24, 25]. The potential of such a system has been the focus of research over the last few decades.
The possibilities provided by BCIs to measure, process and decode brain activity to interpret neuronal signals can be seen as a way to bypass damaged neural and/or motor structures, making it possible to develop equipment that can improve the living conditions of patients suffering from Amyotrophic

Lateral Sclerosis (ALS), Spinal Cord Injury (SCI), stroke, and other forms of paralysis [26].

From a conceptual point of view, we can think of a generic BCI as a closed-loop system consisting of three blocks [27], Figure 2.1:

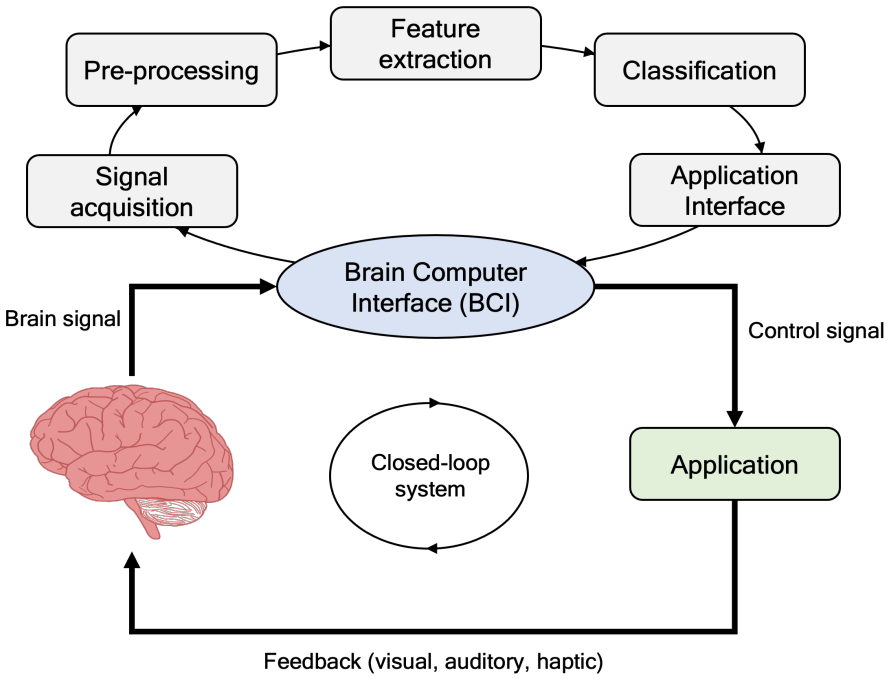

Figure 2.1 BCI schema.

1. Signal acquisition block, in which the brain signals are acquired by using either invasive or non-invasive methods.

2. Processing block, where measured brain signals are pre-processed mainly to remove movement artefacts, features are extracted and finally the signal is classified, thus generating input for external peripherals.

3. Feedback block, which closes the loop by informing the subject on her/his current mental state.

2.1.1 Signal acquisition in BCI

The first categorization of BCI systems can be obtained by analysing the technology for the acquisition of brain signals. In particular, we can identify two macro-categories [28]:

- Invasive systems;
- Non-invasive systems.

Invasive systems involve implanting the electrodes directly into the brain. This type of technology allows the highest signal quality, but on the other hand, the implantation procedure itself has significant risks.
A subcategory of the above-mentioned BCIs is *semi-invasive* BCIs. It is based on Electrocorticographic (ECoG) recording, which uses electrodes implanted subdurally on the surface of the brain to measure electrical signals in the brain cortex [29]. Although being less invasive, also this type of BCIs requires a craniotomy, with all the correlated risks.
On the other hand, *non-invasive* BCI systems are the protagonists of most scientific experiments since they do not require any surgical intervention, and thus not involving any risk situation for the patient himself. Non-invasive BCIs can rely on different types of signal acquisition techniques [30, 31], including:

- Magnetoencephalography (MEG), measures the magnetic field caused by the currents in the brain.

- Positron Emission Tomography (PET), is a nuclear imaging technique used in medicine to observe different processes, such as blood flow, metabolism, neurotransmitters, happening in the body.
- Functional Magnetic Resonance Imaging (fMRI), is a functional neuroimaging procedure using MRI technology that measures brain activity by detecting changes associated with blood flow.
- Near-Infrared Spectroscopy (fNIRS), brain activity is measured through hemodynamic responses associated with neuron behaviour.
- EEG, provides the recording of electrical activity of the brain from the surface of the scalp.

Of those just mentioned, one in particular takes its place in the work to be presented in the next chapters of this thesis, we refer to non-invasive BCI based on EEG.
EEG, as mentioned above, is a non-invasive technique that evaluates the electrophysiological activity of the brain by measuring potentials on the scalp [32, 33]. In particular, the EEG represents the electrical activity of pyramidal neurons in the cerebral cortex, which, due to their synchronisation, arrangement and proximity to the scalp, give rise to electrical fields that effectively add up and are thus detectable at the surface. Despite the low amplitude of this type of signal and high sensitivity to noise, EEG remains by far the most popular neuroimaging modality in BCI, as it offers good temporal resolution, good system portability, relatively low cost, and ease and safety of use [34].
A distinction that needs to be made is between an EEG recorded without external stimuli, called spontaneous EEG, and EEG recorded as a response to external stimuli, called Event-Related Potential (ERP) [35, 36]. There are two fundamental parameters to describe the EEG signal, the amplitude

and the frequency of the oscillations. Typical amplitudes are of the order of 50-200 μV for spontaneous EEG and a few μV for ERP.
From a spectral point of view, the EEG signal occupies a frequency range between 0.1 Hz and 100 Hz. Within this range we can then identify sub-bands of interest because they are related to physiological states (sleep, relaxation, attention, concentration etc) or pathological states (epilepsy, tumours, coma etc), Figure 2.2. Five main EEG rhythms can be identified [37]:

- *Delta rhythms* (0.1 - 4 Hz): In adults, delta waves are associated with states of deep sleep, while a large delta band activity in the waking state is considered pathological. In children, the amplitude of delta waves decreases with increasing age.

- *Theta rhythms* (4 - 8 Hz): Like delta rhythms, theta waves are also more prevalent in children, while in adults they are associated with states of sleep or meditation. In some adults the theta rhythm is also associated with emotional stress, particularly frustration, for example immediately after the sudden removal of a pleasant stimulus.

- *Alpha rhythms* (8 - 13 Hz): These waves are recorded in a waking state, but indicate a state of relaxation. In the occipital areas, for example, the amplitude of alpha waves increases greatly when the eyes are closed, while it decreases drastically when they are reopened, or if mental effort is exerted. In the same frequency range as the alpha rhythm, but with localisation in the area of the motor cortex, we also find the *mu* rhythm, which is particularly interesting because it is strongly correlated with movement.

- *Beta rhythms* (13 - 30 Hz): These waves are recorded in the frontal, central and parietal areas, and occur during the waking state with eyes

open when the subject is involved in a mental activity. Beta rhythms are also associated with motor activity, and are modulated during both real movement and motor imagination.

- *Gamma rhythms* (30 - 100 Hz): gamma rhythms have frequencies greater than 30Hz and indicate a state of deep concentration. Gamma rhythms also occur in relation to certain motor functions and during maximal muscle contraction. Gamma rhythms are less used in EEG-based BCI systems because they are more susceptible to muscular or EEG artefacts.

2.1.2 Control signals in EEG-based BCI systems

BCIs are based on so-called control signals, which are taken directly from the brain. These signals can be categorised into three macro categories which are evoked signals, spontaneus signals and hybrid signals.

Evoked signals

An evoked signal, or Evoked Potential (EP), is a detectable signal as a variation of the EEG that occurs following the presentation of a somatosensory, auditory or visual stimulus. For the sake of clarity, although the terms are sometimes used synonymously, EP and ERP have different meanings, as ERP is associated with enhanced cognitive processing.
However, included in the category of evoked signals we find:

- *Steady State Evoked Potentials (SSEP)*: the brain produces SSEP signals in the cortex when the subject perceives periodic stimuli such as flickering images, modulated sounds, and also when the subject feels certain vibrations [38]. Among the various types of detected SSEP

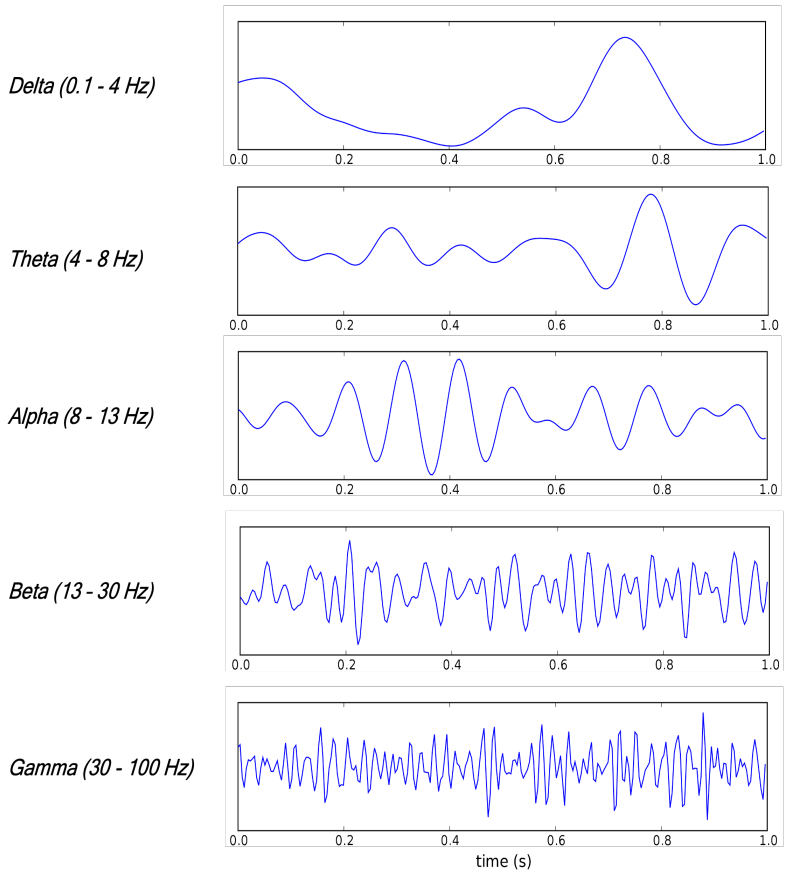

Figure 2.2 EEG rhythms.

signals we can find Steady State Visual Evoked Potentials (SSVEP), somatosensory SSEP and auditory SSEP.

- *P300*: this type of evoked potentials occur as small positive spikes following infrequent visual, somatosensory or auditory stimuli when these are interposed with more frequent or routine stimuli. They are detectable in the parietal cortex approximately 300ms after the infrequent stimulus [39].

Spontaneus signals

Spontaneous signals, as the word implies, are the signals generated spontaneously/voluntarily by the subject without any external stimulation. Part of this family of signals are:

- *Sensorimotor rhythms*: these rhythms are commonly referred to as EEG signal oscillations, and are detectable in areas close to the primary motor and sensory cortex. Sensorimotor rhythms have the characteristic of being modulated in correspondence with any motor task, but the great success of these rhythms in the BCI field derives from the fact that a subject can learn how to control them even without the need for real movement [40].

  Modulations of these rhythms resulting from sensory stimulation, a motor act or imagination can be of two types, called Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS), of alpha and beta bands. ERDs are detected during motor preparation, execution and imagination and can be seen as a correlate of an activated cortical area, in contrast ERSs represent a deactivated cortical area or an inhibited cortical network, at least under certain conditions.

- *Slow Cortical Potentials (SCP)*: these control signals are small, slow changes in cortical potential lasting from 300ms to a few seconds, which can also be detected by EEG [41].

  These slow potential shifts are related to changes in the level of cortical activity, such that negative SCPs correspond to increased activity, whereas positive shifts are associated with reduced activation of the cortex. Through training, it is possible for both healthy and paralysed patients to learn to voluntarily control SCPs, which can then be used

as control signals in BCI systems, e.g. to move a cursor on a screen or control an external device such as an exoskeleton.

A sub-category of SCP that is of interest for the purposes of this thesis work is **MRCP** [42]. These potentials are known to reflect the cortical processes employed in planning and execution of a motor task. These cortical processes represent the pre-intention of a future action that a subject wants to perform, for example the beginning of a step during walking. As shown in Figure 2.3, they begin with a slowly increasing negativity, called the Bereitschaftspotential (BP) [43], and progress to a steeper, later negativity that begins about 300/400 ms (depending on the subject) before the onset of the movement, called the Negativity Slope (NS).
As we will see later, the use of this type of control signal is widely used in the literature to perform tasks such as movement detection and/or classification.

Hybrid signals

When we talk about hybrid control signals we mean the combination of several control signals in order to increase the reliability of the BCI system. Examples of combinations found in the literature are SSVEP and P300 [45], SSVEP and Electromyography (EMG) [46], EEG and EMG [47] and many more.

2.1.3 Synchronous and Asynchronous BCI

BCI systems can be classified as synchronous or asynchronous, depending on how the input data is processed.
Synchronous BCIs analyze brain signals only during predefined time windows, for example, time-locked to a specific event or stimulus. As a consequence, we have that the user during the execution of a task will have to

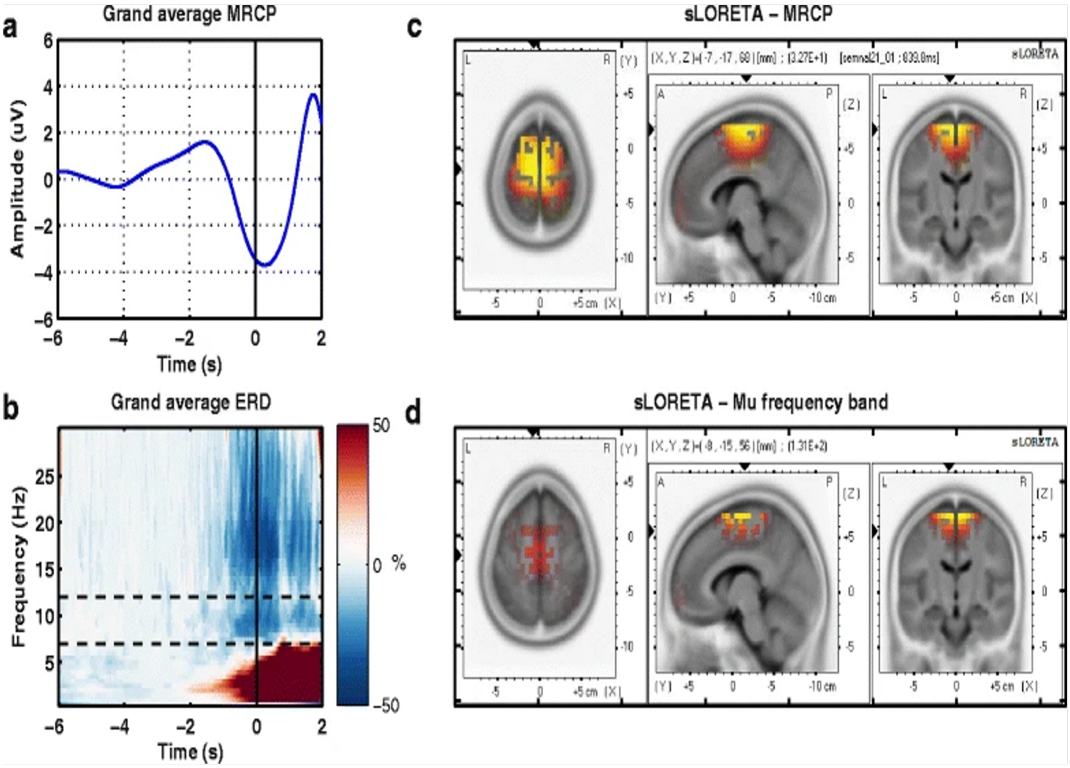

Figure 2.3 Neurophysiological results. a. Grand average MRCP over subjects and sessions for nineteen cases; **b.** Grand average ERD over subjects and sessions. The dashed line marks the mu frequency band later used in classification. In both figures, the black vertical line represents the onset of motion; **c.** and **d.** Source localization results for MRCP temporal features and μ band spectral features at movement onset are displayed from three perspectives: top view, lateral view from the left and back view. From [44].

send commands only during these periods dictated by the BCI itself. One of the advantages of this type of system is its implicit robustness, since the onset of mental activity is known a-priori and is therefore associated with a specific signal. This results in a reduction of motion artifacts and a subsequent simplification of the BCI.

However, the BCIs just mentioned assume that the user is always in the control state, thus causing false interpretations of the brain signal and subsequent false actions in the BCI. Thus, we can derive the need for the system to be able to infer from the EEG whether the user intends to operate the interface (control state) or not to operate (non-control state). A BCI that can handle control states is called asynchronous. From a signal processing point of view, the asynchronous approach allows to a more natural interaction with the user and also a lower latency in the overall system.

2.2 Related Work

2.2.1 BCI for gait decoding

The overall clinical efficacy of BCI-robotics is highly dependent on the correlation between robot motion and predicted motion, which in turn depends on the robustness of the BCI determined by the quality of the brain signal and the performance of the decoding tools. Over the last decade, several BCIs using variety of neural inputs, feedback modalities, and experiment protocols have been reported for resolving the gait decoding problem.

The study carried out in [48] shows how signal features with frequency less than 2 Hz are the most significant in the walk and direction classification phase. In this regard in [49, 50] is shown how the delta band of the EEG signal contains information concerning the kinematics of the walking movements, and it is shown how such information can be decoded through particular Wiener and Kalman filters.

The most common BCIs for the decoding of lower limb movements dealt with the problem of distinguishing basic commands such as ”stop” and ”go” and many of the works found in the literature to do this use ERD (Sezione 2.1.2)

as a feature for decoding. To give some examples, it can be seen in BCIs like [51, 52, 53] the use of this feature for the control of a lower-limb exoskeleton, while in [54] can be found a BCI that control a Functional Electrical Stimulation (FES) system for overground walking. Another aspect that can be explored is the issue of crossing an obstacle, addressed in [55] where the authors decoded the subject's intention to walk or cross an obstacle through a Support Vector Machine (SVM) that predict the intention based on Common Spatial Pattern (CSP) features. In [56], a knee prosthesis is presented that is used for two main functions: sitting and walking. In this case the purpose of the study was to establish the feasibility of manipulating a prosthetic knee by BCI, hence, it can be seen that the solution is based on a simple software that activate the prothesis under a certain threshold condition of ERD signal. Despite being less investigated, recent works showed the potential of decoding the gait activity through MRCP analysis. For example, [57] exploited the EEG delta band to decode the walking kinematics. An interesting study is shown in [58], in which a solution composed by a non-linear dimensionality reduction method, called Locality Preserving Projection (LPP), and a Linear Discriminant Analysis (LDA) classifier were utilized to decode MRCP and control a Motorized Ankle-Foot Orthosis (MAFO).
Nevertheless, one can also find works in which ERD and MRCP features are used together as in [59].

It is worth highlighting that all these works use classical machine learning techniques that do not consider the temporal dependence of the features with respect to the movement they are decoding. However, it has been shown that the brain activity over the motor cortex is characterized by a modulation time-locked with the gait cycle during locomotion activity [60]. Thus, it is of fundamental importance for gait decoding to have an explicit representation

of time inside the classification framework.

2.2.2 Deep learning in BCI

The use of neural networks has never been popular as a choice for solving problems concerning the decoding of neural correlates of EEG signals. This, in the past, was due to too high computation times, scarcity of available datasets and/or problems with the architectures themselves (vanishing/exploding gradients, 3.3).
In recent years, technological progress and the greater availability of large amounts of data have made it possible to develop algorithms based on deep learning that can innovate and improve the performance obtained with the classic methodologies such as statistical analysis or non-linear methods (neural networks). Many of these algorithms, because they are not based on deep learning, require feature engineering as a separate task before performing classification. This leads to disadvantages from the perspective of losing some information because features extraction requires basic knowledge of biology to investigate the neural correlates state through EEG signals. Human experience may help capture features on some particular aspects but it is insufficient in more general conditions. Therefore, an algorithm is required to extract representative features automatically.
Deep learning was created to fill this gap by providing algorithms that can extract features from time series automatically and faster, hence, avoiding the time-consuming feature engineering steps by working directly on possibly pre-processed brain signals to learn distinguishable information through back-propagation.
At present, deep learning in EEG classification, has explored areas such as: Emotion Recognition, Motor Imagery, Mental Workload, Seizure Detection,

Sleep Stage Score, and Event-Related Potential [61].
However, other works recently demonstrated how deep learning has outperformed machine learning in applications like prediction of motion [62] or prediction of epileptic seizures [63].
In this light, it is clear that the thinking of the scientific community is gradually shifting towards the use of deep methodologies to develop better performing BCIs. Bringing back the topic of gait decoding, it can be said that only a few works propose BCIs based on deep learning for solving this problem. In [64], the methodology involves the use of a Long Short Term Memory (LSTM)-based recurrent neural network that exploits the information of SCP to do an offline decoding of a subject's walk. The same authors in [47], have filled the gap due to the single-signal decoding also integrating the EMG information. A last work that explored this field is the one in [65], in which a LSTM network was tested offline to estimate the joint angles of a leg on healthy subjects during treadmill walking. Although also these works show superior performance of deep learning models with respect to traditional machine learning approaches, it is important to note that none of the latter works are compatible with a real-time decoding application.

2.3 Aims and structure

The research work carried out in this thesis aims to develop a novel BCI for decoding the walking intention from EEG signals. In particular, the proposed methodology, to the best of our knowledge, is the first exploiting a deep learning-based approach for gait decoding which is fully compatible with an online application, and thus aims to overcome the limitations found in the literature.

Two main milestones can be identified in this thesis: the first concerns the study of the EEG correlates of the swing phase of the walk. The information extracted from this analysis was essential to be able to propose a sensible and valid approach to detect walking intention. The second milestone concerns the identification and optimisation of a deep learning model able to learn from neural correlates the intention of a subject who is about to walk in order, for example, to trigger the support of a lower limb exoskeleton. Although the focus of this thesis is on the decoding of walking movement, it is worth highlighting that the proposed model can be applied for the decoding of other types of movements (e.g., upper limb reaching).

The thesis is structured in six chapters: Chapter 1 'Background' and Chapter 2 'Introduction' aim to provide an overview of the issues addressed in this work, to describe the landscape of BCI and their inter-facing with deep learning techniques, and to describe the addressed problem. Chapter 3 'Materials and Methods' will list the tools used and describes the proposed approach. In 'Results', Chapter 4, the efficacy and performance of the implemented model will be evaluated on a pre-recorded walking dataset. Finally, in Chapters 5 and 6, respectively 'Discussion' and 'Conclusions', the results obtained will be discussed and the contribution that this work has made in relation to the state-of-the-art will be highlighted.

3 | Materials and Methods

3.1 Summary

As mentioned in 'Aims and structure' (2.3), in this thesis, I aim to show the effectiveness of a deep learning based approach for gait decoding. A summary of the offline processing and classification procedures used in this work is shown in Figure 3.1.

The dataset considered in this thesis is a collection of EEG and EMG record-

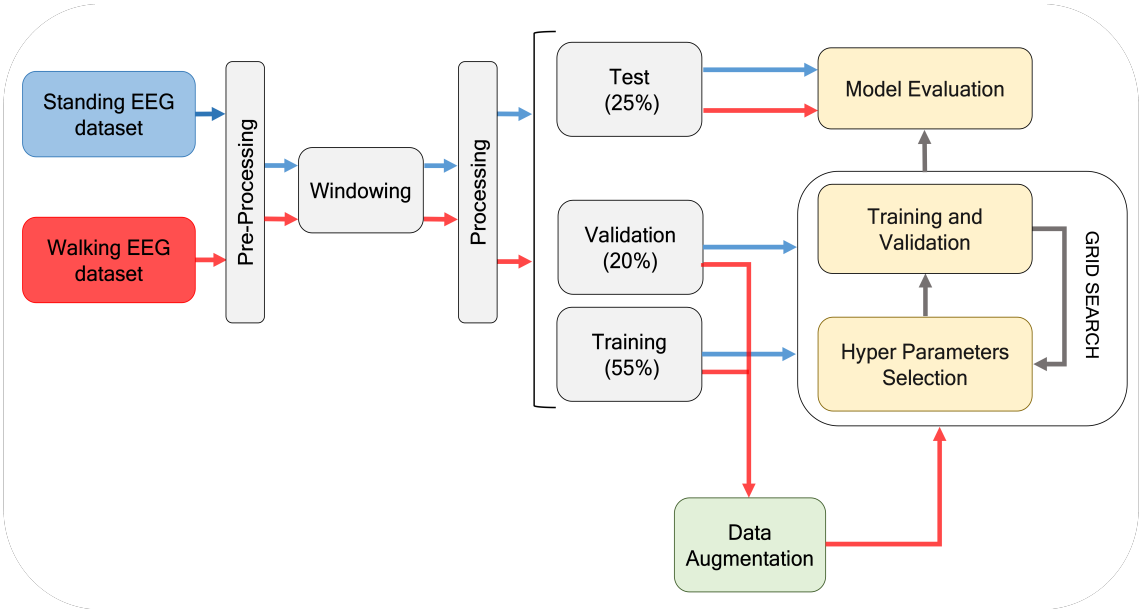

Figure 3.1 Model Evaluation.

ings related to walking. Only the portion related to the EEG signals was used and in particular it was pre-processed, for the reduction of artifact contamination, according to the guidelines of [66], work in which the dataset was acquired using a Mobile Brain/Body Imaging (MoBI) framework.
Considering the promising results of Deep Learning (DL) networks to decode movements (sub section 2.2), and inspired by works like [67], it has been decided to develop a recurrent network model based on GRU layer in order to solve the problem of gait decoding through the detection of pre-movement EEG correlates, MRCP [68].
This type of classifier was optimized and trained with different dataset percentages and with different amounts of data augmentation to get a performance estimation that reflects the real case in which a subject will record a small amount of EEG data per session and the model will progressively learn to be more and more accurate. The aim of this deep learning classifier will be to detect between a future 'swing' and a 'stand' state.

3.2 Deep Learning Fundamentals

DL is a sub-field of Machine Learning (ML) that has achieved enormous success in recent years in various contexts. The applications based on DL that are seen nowadays both in the scientific and industrial fields are varied among which, classification of sounds or images, natural language processing or based on the discussions of this thesis movement detection. When we speak of learning we refer to a procedure that consists in estimating the parameters of the model from the data so that the learned architecture can be used to perform a specific task. In particular, the most used DL architecture is the so-called Artificial Neural Network (ANN), or easily Neural Network (NN).

In the following sections, the architectures and related algorithms taken into consideration in the work to be presented will be described in increasing detail. Part of the proposed material is derived from [69].

3.2.1 Supervised Learning

Among the many DL approaches, in this thesis I will focus only on Supervised Learning [70]. This type of technique uses a-priori labelled data for learning purposes. For the sake of completeness, a model based on supervised learning can be defined as:

- *Model:* seen as a set of hypothesis H.
- *Domain set:* X is the set of object that we may wish to label.
- *Target set:* Y is the set of possible label (target).
- *Training data:* $S = (x_1, y_1), ..., (x_N, y_N)$ is a finite sequence of N pairs in $X \times Y$. This is the input that the model has access to.
- *Predictor:* the model is requested to output a label that act like the function $h \in H$ such that $h : X \to Y$.
- *Loss function:* a function $L : h(x), y \to \mathbb{R}$ that measure how bad is the prediction.
- *Training algorithm:* $A(s)$ that aim to output the best $h \in H$ according to the loss function L.

It can be said that the learning problem that will be addressed can be seen in this way: given a training set S and a model (e.g. a Recurrent Neural Network (RNN) [linked to paragraph]), the aim is to minimize the

loss function L with a training algorithm (e.g. backpropagation) to have the "best" $h \in H$ (which corresponds to the particular weights of the RNN).

3.2.2 Neural Networks

Informally, a NN model is a collection of artificial neurons that receive an input, combine them in their internal state and produce an output using an activation function. The groups of neurons of a network are organised in multiple layers, among which can be distinguished: input layer that receives the data, output layer that produces the results and hidden layer (everything that lies between input and output layer). The interconnections between these artificial neurons are called neural connections, each of these connections has a specific weight that is modified during the training process.

Feedforward and Recurrent Neural Networks

Neural networks can be of two types:

- *Feedforward*, Figure 3.2, in these networks, connections link neurons of one level with neurons of the next level and therefore backward connections or connections to the same level are not allowed.

- *Recurrent*, in recurrent networks there are feedback connections (usually to neurons at the same level, but also backwards). This considerably complicates the flow of information and training, requiring to consider the behaviour in several temporal instants (unfolding in time). On the other hand, these networks are more suitable for the management of sequences (e.g. audio, video, sentences in natural language), because

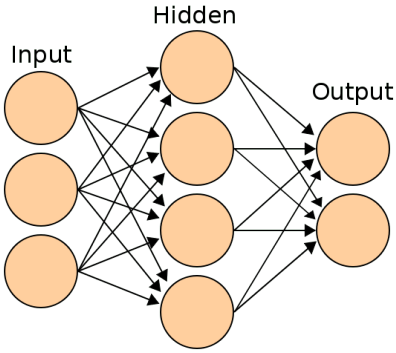

Figure 3.2 Representation of an ANN.

endowed with a memory effect (short term) that at time t makes the information processed at t - 1, t - 2, etc.

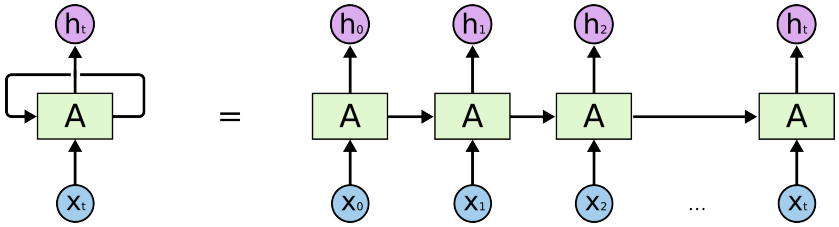

Figure 3.3 Representation of an RNN.

A significant difference in these two types of networks is the fact that feedforward networks have different weights for each node while recurrent neural networks share the same parameter weight in each layer of the network.

Nevertheless, both typologies modify the value of these weights through processes of Backpropagation (BP') and Gradient Descend (GD) to facilitate the reinforcement learning.

In more detail in Section 3.2.2, it will be seen that the recurrent networks, unlike the feedforward ones, use a variant of the algorithm of BP' called Backpropagation Through Time (BPTT).

Backpropagation

The backpropagation algorithm is the algorithm through which an NN has the ability to adapt to a problem. The idea is to test how wrong the network's The idea is to test how wrong the prediction of the network is, and adjust the weights to correct it, repeating the operation many times for each training data. In a simple way the procedure can be encapsulated in a list of steps:

- *Forward step:* Propagate the data through the network and calculate in the output the loss function (this can be done for multiple data and taking the average in the end).
- *Backward step:* Calculate how to adjust the parameters inside the NN through the gradient of the loss function.
- *Output step:* Update the parameters.

Backpropagation Through Time

The application of the backpropagation algorithm to a recurrent neural network applied to sequential data as a time series, as mentioned above, is called Backpropagation Through Time.
In this type of network at each time step presented as input there is a prediction of the output.

It can be seen in Figure 3.4, on the left we have a RNN with a single hidden layer represented compactly.

For convenience a notation can be defined:

- x^t, input layer.
- h^t, hidden layer.

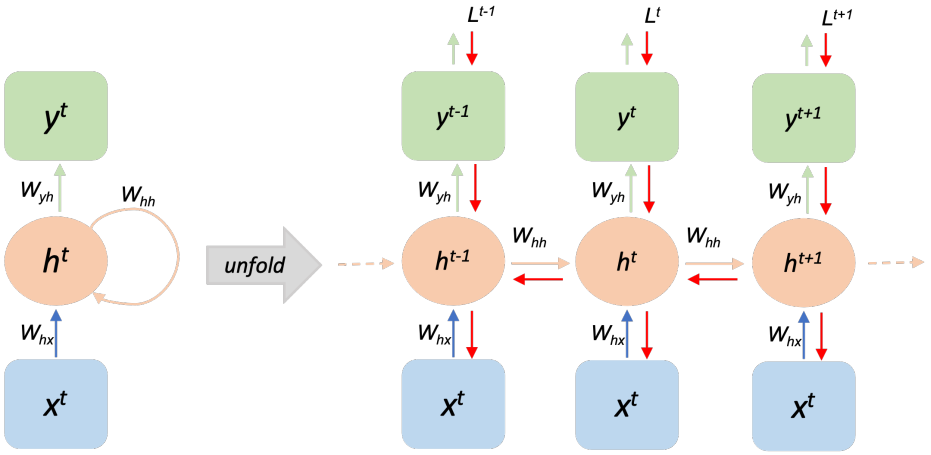

Figure 3.4 RNN Unfold.

- y^t, output layer.
- W_{hx}, matrix weight between input and hidden.
- W_{yh}, weight matrix between output and hidden
- W_{hh}, weight matrix of the hidden layer (not present in feedforward networks).

The algorithm of backpropagation is applied to the unfolded graph of the RNN. You can see from the notation that at each time step are present the same weight matrix as mentioned above. We can then begin to describe the net input of the hidden layer at timestep t through this equation:

$$z_h^t = W_{hx}x^t + W_{hh}h^{t-1} + b_h \tag{3.1}$$

while the activation is given by:

$$h^t = \sigma_h(z_h^t) \tag{3.2}$$

Proceeding, the net input for the output is then given by:

$$z_y^t = W_{yh} h^t + b_h \tag{3.3}$$

with activation:

$$y^t = \sigma_y(z_y^t) \tag{3.4}$$

As it is seen in 3.4, every time step calculates the own loss and the total loss for a given sequence of input values paired with a sequence of output values would be the sum of the losses over all the time steps:

$$L = \sum_{t=1}^{T} L_t \tag{3.5}$$

Using chain rule, the gradient of loss function w.r.t. W_{yh} at t timestep is:

$$\frac{\partial L^t}{\partial W_{yh}} = \frac{\partial L^t}{\partial y^t} \frac{\partial y^t}{\partial W_{yh}} \tag{3.6}$$

The partial derivate of loss function w.r.t. W_{hh} depends on the previous hidden states so if for example we want to calculate the gradiend at time $t = 2$ we have:

$$\frac{\partial L^2}{\partial W_{hh}} = \frac{\partial L^2}{\partial y^2} \frac{\partial y^2}{\partial h_2} \frac{\partial h^2}{\partial W_{hh}} + \frac{\partial L^2}{\partial y^2} \frac{\partial y^2}{\partial h_2} \frac{\partial h^3}{\partial h_2} \frac{\partial h^2}{\partial W_{hh}} \tag{3.7}$$

while for W_{hx} the general formula for gradient calculation is:

$$\frac{\partial y}{\partial W} = \sum_{i=t+1}^{t+N} \frac{\partial y}{\partial h_{t+N}} \frac{\partial h_{t+N}}{\partial h_i} \frac{\partial h_i}{\partial W} \tag{3.8}$$

More details can be found in [71].

Note that as the number of input sequences increases in terms of timeteps,

the number of derivations required for a single weight update also increases considerably. This can cause the weights to disappear or explode (go to zero or overflow) and make learning slow and the model skill noisy.

3.3 Proposed model

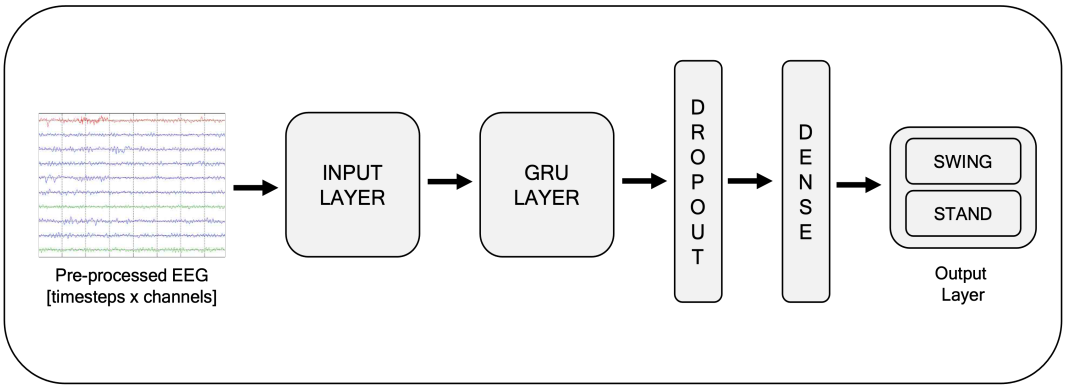

Figure 3.5 Model Architecture.

In the BCI developed in this work, the task of classifying the brain signal is performed by the model in Figure 3.5. It is a model with a simple architecture that exploits the potential of a GRU layer. In particular the architecture of the network is formed by an input layer that receives in input temporal windows of fixed length, a GRU layer necessary for the decoding of neural correlates, a dropout layer to prevent any overfitting and a fully connected layer followed by the output layer with two neurons.

More in detail a GRU is an evolution of the classical architecture of a recurrent neural network that aims to eliminate the problem of the vanishing gradient. According to [72], the problem of the vanishing gradient is related to the intrinsic structure of the usual RNNs. In particular, as explained earlier in subsection 3.2.2, BPTT results in a decrease in the gradient as it

propagates backward in time, and since gradients are values used to update the weights of a neural network, if a gradient value becomes extremely small, it does not contribute to learning. Here the update formula:

$$new\_weight = weight + (learning\_rate * gradient) \tag{3.9}$$

In other words, small or no updates of the gradient lead to a network that does not learn.

GRUs were created to fill this gap and in particular as a solution to short-term memory. To do this, this type of recurrent network uses internal mechanisms called gates which can regulate the flow of information, and in this particular architecture we have:

- *forget gate*, used from the model to decide how much of the past information to forget.

- *update gate*, similar to forget gate, it decides what information to throw away and what new information to add.

As mentioned earlier, such a network is able to maintain information related to previous timesteps. In other words, a GRU is able to understand and maintain in memory the most relevant information and to throw away or give less weight to the others.
Referring to the Figure 3.6, I will now explain the mathematics behind this architecture

First, let's introduce the notations:

- x_t: input at time t

- h_t: hidden unit value at time t

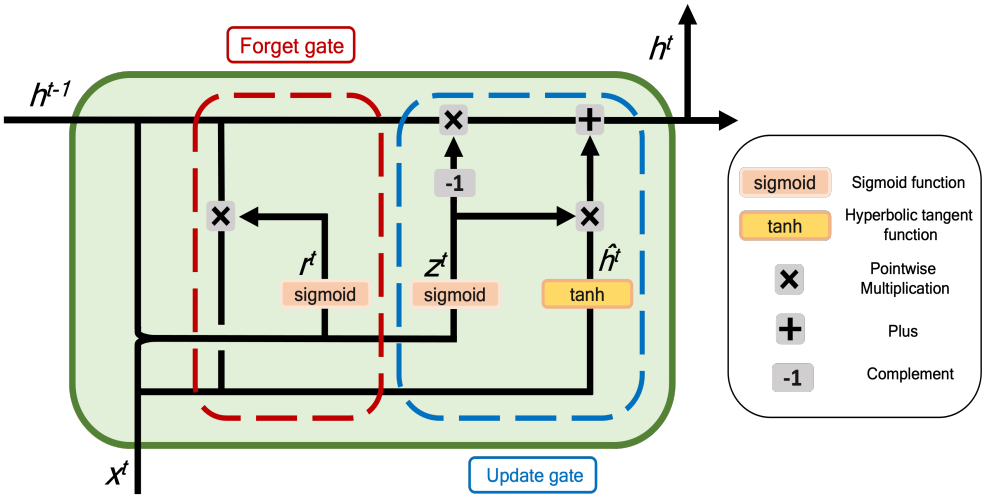

Figure 3.6 Single unit of a GRU layer.

- $\hat{h}_t$: candidate activation vector
- r_t: reset gate array
- z_t: update gate array
- $W^r, W^z, W^h, U^r, U^z, U^h$: weight matrices

First, the array r_t of values in output to the reset gate is computed:

$$r_t = \sigma(W^z x_t + U^z h_{t-1}) \tag{3.10}$$

In this equation we see that the input values x_t to the unit and the hidden unit value at the previous time step h_{t-1} are multiplied by the corresponding weight arrays W^z and U^z. The weighted arrays are then summed and finally the *sigmoid function* is applied. All these operations have the purpose to 'forget' the values that are not relevant for the learning. The sigmoid function being defined in $[0, 1]$ goes to normalise the value of the most important

features to values that are close to 1, while values that must be forgotten close to 0.

Considering now the calculation of the intermediate result given by the update gate z_t, we see from the formula below that essentially the preliminary operation that is carried out is the same of the forget gate with the difference of the weights used.

$$z_t = \sigma(W^r x_t + U^r h_{t-1}) \tag{3.11}$$

A final component we need is the candidate activation vector $\hat{h}_t$ which is calculated as follow:

$$\hat{h}_t = \tanh(W^h x_t + (r_t \odot U^h h_{t-1})) \tag{3.12}$$

Here tanh activation function is used to help regulate the values flowing through the network in the range $[-1, 1]$.
Now that we have all the elements we need to upgrade the hidden state h_t let's calculate its value as follows:

$$h_t = (1 - z_t)h_{t-1} + z_t \odot \hat{h}_t \tag{3.13}$$

To recap, the interacting parts of this formula are r_t which contains the useful information taken from the previous hidden state, $\hat{h}_t$ which stores the information derived from the past and z_t which tells us which information is useful to enrich the new hidden state.

The training strategy for GRU models is Adam [73], which has been widely used in the training of multilayer networks and proved to be efficient in the training process. This model is trained by optimising the cross-entropy loss function. The *learning rate* was set to 0.001, the $\beta 1$ e $\beta 2$ were set at 0.9

and 0.999, respectively, ε to $1 \cdot 10^{-8}$ and *clipnorm* was set to 1.

3.3.1 Network Optimization

When using deep learning models, one must take into account the importance that lies in parameter optimization. The reason is that neural networks are known to be difficult to configure and there are a lot of parameters that need to be set. Thus, in this thesis I also investigated the best set of parameters that would ensure a robust and accurate gait decoding.
To do this, a preliminary grid search was launched to better understand the behaviour of the model as the parameters change, table 3.1:

- *Number of hidden units* of the GRU layer of the model, since there is no definitive rule of thumb on how many hidden units one should choose.
- *Batch size*, the number of samples that will be propagated through the network at each iteration.
- *Percentage of Dropout*, ignoring randomly selected neurons during training usually increases a model's ability to generalise a problem.

Parameter	Tested values
Number of hidden units	50, 100, 150, 200
Batch size	8, 16, 32
Percentage of Dropout	0, 0.1

Table 3.1 Network optimization. Parameters used in the network optimization.

It should be specified that all training performed in this work was done using a balanced dataset with equal numbers of swing and stand samples.

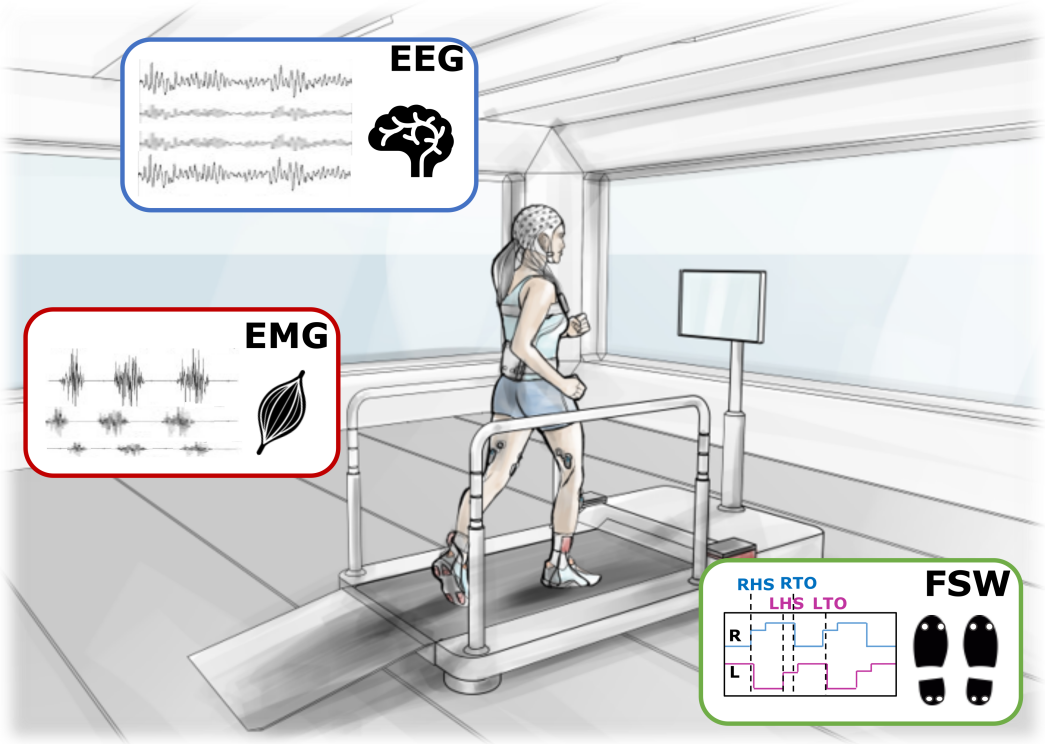

Figure 3.7 Mobile Brain/Body Imaging (MoBI) framework [66]. The experimental set up consisted of a 64-channel EEG and a 6-channel EMG acquisition systems to record neurophysiological signals of participants walking on the treadmill. The Acticap Control Box receiving EEG cables from the cap was securely fastened to the subject's waist and connected to the EEG amplifier via a 2-m cable fastened to the hand rails of the treadmill to minimize cable movements. Four foot-switches (FSW) were placed under the heel and the toe of each foot and wirelessly acquired by the EMG amplifiers to detect gait events.

3.4 Model evaluation

3.4.1 Dataset

The dataset was acquired through the use of the so-called MoBI framework [74], Figure 3.7, which allows the simultaneous acquisition of EEG and EMG with sufficient temporal resolution for motion analysis [75]. Thus, this data collection consists of simultaneous recordings of 64-channel EEG signals and

6 lower extremity EMG signals from eleven able-bodied subjects (mean age 30 ± 4 years) walking on a treadmill at two distinct velocities, 2.5 km/h and 3.5 km/h. For each walking velocity, two acquisitions of 10 minutes each were performed. In this work, only the EEG recordings for each of the subjects were considered, at the velocity of 2.5 km/h.

EEG data were recorded with a custom signal pre-amplifying active electrode cap (acti-CAP, Brain Products GmbH, Germany) and a 64- channel EEG amplifier (SD MRI, Micromed S.p.A., Italy) with a sampling rate of 2048 Hz/channel (bandwidth DC - 1024 Hz). The montage was chosen in accordance with the 5% International 10/20 System [76]. EMG electrodes were placed according to SENIAM guidelines (www.seniam.org) on three muscles of each leg, namely Tibialis Anterior (TA), Vastus Medialis (VM) and Biceps Femoris (BF), which were simultaneously recorded with a wireless EMG system (BTS Free EMG 300) at a sampling rate of 1000 Hz. In addition, the shoes of each subject were equipped with four foot-switches, two under the heel and two under the toes, in order to identify the gait events of each leg (Right/Left Heel Strike—RHS/LHS and Right/Left Toe Off— RTO/LTO). Data from the foot-switches were wirelessly acquired through the EMG system with a sampling frequency of 1000 Hz. A common trigger was sent both to EMG and EEG acquisition systems multiple times at the beginning and end of each session to guarantee robust, minimum-jitter offline synchronization. EEG, EMG and foot-switches data were resampled at 1024 Hz and aligned before further preprocessing.

In order to reduce movement artifacts and to have at a later stage a more faithful decoding of the motion correlates in the EEG data, a pre-processing was done. In essence, it is based on Artifacts Subspace Reconstruction (ASR) and Adaptive Mixture Independent Component Analysis (AMICA) and is

used to remove bad channels, bad epochs and line noise (more details in [64]).

For the purpose of this thesis, the EEG dataset of each subject (both movement and stance) was then divided into time windows (windowing) of 300 ms, a size based on the MRCP literature and the average duration of the stance phase. The Common Average Reference (CAR) filter was then applied to the portion of the signal present in each of the time windows, with the intent of re-referencing the signal with respect to the average electrical activity measured on all scalp channels. Finally, with the aim of the classification phase, z-score normalization were applied on the frequency bands from 1Hz to 8Hz that were extracted for each time window through a 4th-order Butterworth filter. This range was chosen as seen in [64] to give the best results, and in particular as found in [77, 78], it was seen that frequency bands below 10 Hz contain most of the information related to movement.

Thus speaking of windowing, as can be seen in Figure 3.8[A], as far as the walking dataset is concerned, the temporal windows of equal length into which it has been divided include all the temporal instants prior to the swing phase of the walk (movement onset). Regarding the windowing of the stand dataset, Figure 3.8[B], it was carried out by sliding a window of the same length as that used for the walk with an overlap between adjacent windows of 20Hz in order to generate a greater number of samples (sub section 3.4.2).

3.4.2 Data Augmentation

The dataset was divided into 25% test set, 20% validation set, and the remaining training set data. It is well known that deep classifiers require a large number of samples to learn, and it is for this reason that in order to improve performance, especially on smaller cuts of datasets, I used the tech-

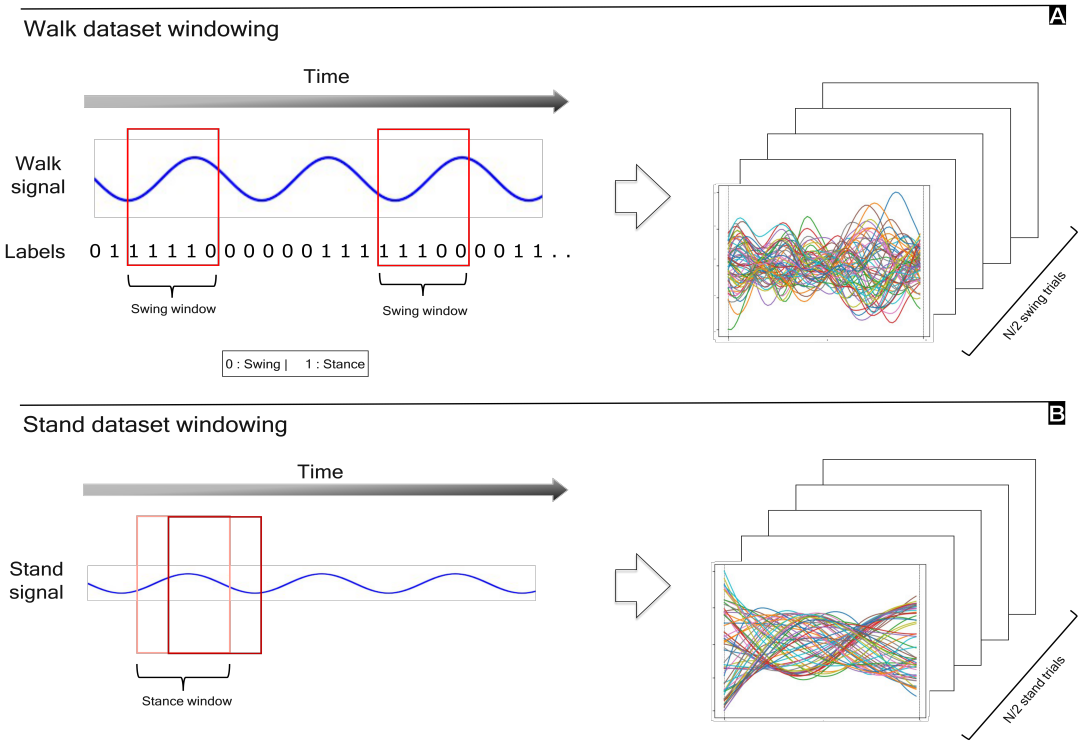

Figure 3.8 Windowing. [A] Windowing on EEG walk dataset. [B] Windowing on EEG stand dataset.

nique of data augmentation to produce new artificial samples useful to the learning process [79].

In this work, the importance of the data augmentation resides in the increase of walking steps available to train the network. Indeed, in a real application with end-users, a subject with a motor deficit is not able to walk for a long period due to early fatigue, thus limiting the amount of training data that can be recorded during the first session.

Recalling that each window of the pre-processed dataset contains the temporal instants antecedent to the beginning of a swing, the process through which the new windows are generated consists of the same procedure described in Section 3.4.1, except that these windows will no longer be taken

strictly from the instant 0, i.e. the beginning of the swing, but rather from a shift of this instant backwards or forwards. In this way, and as it can be seen from the results obtained in the next chapter, I was able to create artificial brain pre-activations that the models during the training phase recognize as new examples.

Considering what was said in Section 3.3.1, network optimization was performed on the 50% of the training set data (220 steps) and to these data, an augmentation on the walking part was performed obtaining about 660 steps. After that, performance was tested by training the model with different portions of training set (Table 3.2) to which four different data augmentations were applied, Table 3.3.

The intention of testing the model with different percentages of datasets served to simulate the performance that can be achieved from the first recording of a patient's data (which is usually around 30-60 steps) to subsequent recordings that improve the overall accuracy of the model. Hence, the study of data augmentation was done in order to evaluate its effectiveness in conditions of data scarcity, e.g. first recordings.

Training set	**# Steps**
7%	30
15%	70
30%	140
50%	220
100%	440

Table 3.2 Training set percentages and number of steps.

	Data Augmentation Shift [ms]
augmentation x1	0 (no data augmentation)
augmentation x2	0, +100
augmentation x3	-50, +50, +150
augmentation x5	-50, 0, +50, +100, +150

Table 3.3 Data Augmentation third grid search.

4 | Results

4.1 Analysis of walking correlates

An important part of this work is the analysis of correlates in order to find an efficient way to decode the gait information from the EEG signal.
The data used belong to the 1-8 Hz frequency band, which as mentioned in Section 3.4.1, is the one that contains most of the motion information. In Figure 4.1, for each subject, it is possible to visualize the median of the signal made on the Cz channel, over a time window of 500 ms, coupled with median absolute deviation (grey area). This channel was chosen for visualization as the legs are somatotopically represented over the motor cortex in the sulcus between the left and right hemisphere. From the temporal plots I can say to be consistent with the literature [42] as the correlate related to pre-intention of movement that emerges, the MRCP, it can be seen in all subjects. In particular, the negative flexion of the MRCP, the BP, is seen to appear about 200/300 ms before movement onset. Figure 4.2 shows the time course of the grand average amplitude of the EEG signals of all the subjects on the C1, C2, Cz, CP1, F2 and Fz channels. The individual channels plotted highlight the course of [42] averaged for all subjects (colored line) and for each of the subjects (gray line). This visualization revealed that the latency and amplitude of the negative peak of the MRCP varies from subject to

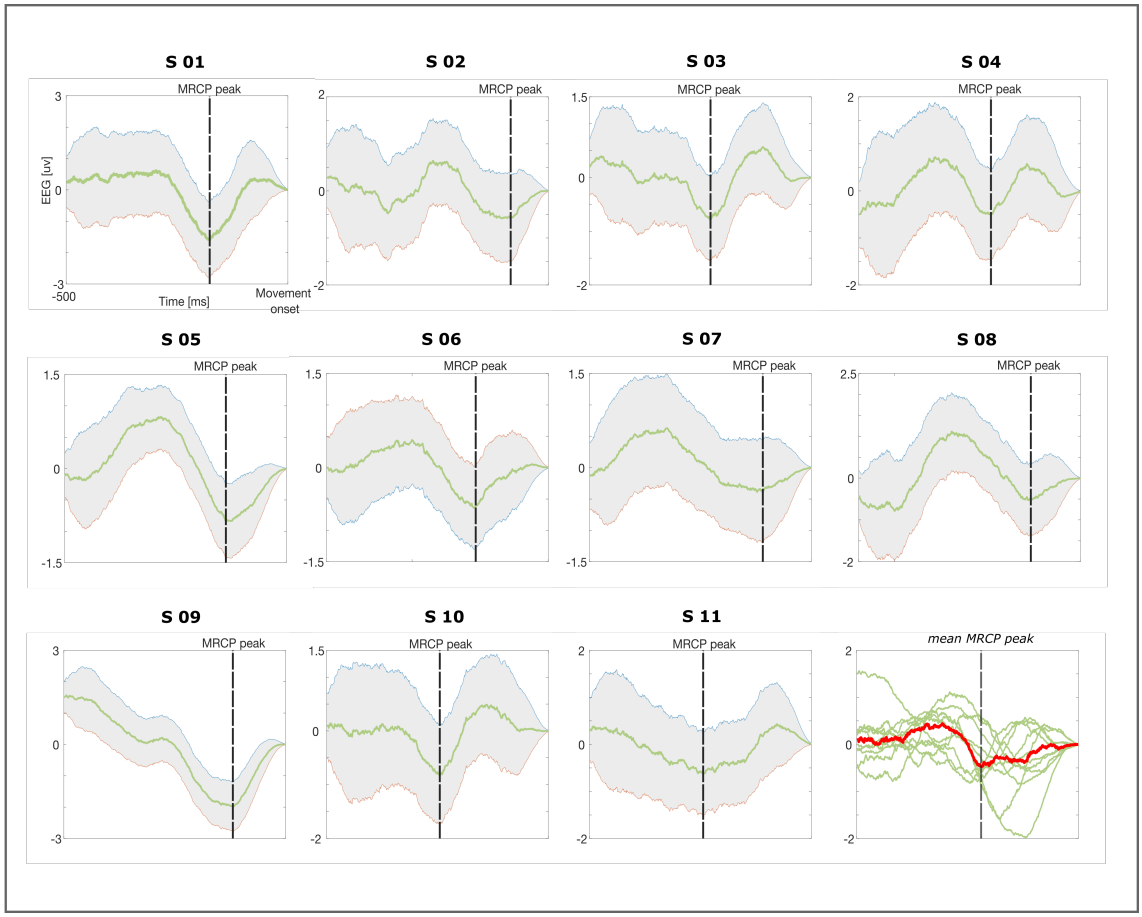

Figure 4.1 Single channel grand average. Median made on Cz channel for each of the 11 subjects considered in this work, the dashed line represents the negative peak of MRCP, while the area around the green line is the median absolute deviation. The information is gathered in the last plot and associated to the median made on all subjects (red line, Cz channel), with the relative line indicating the negative peak of the MRCP (dashed line).

subject and therefore is subject-specific. Also from the grand average of each channel it can be seen that the MRCP appears first in the central channels (e.g., Cz and Fz) than in the lateral channels (e.g., C1 and CP1), consistently with the somatotopic representation of the legs over the sensorimotor cortex. Analysing this behavior on all the recorded channels, it can be seen in the grand average (Figure 4.2, bottom), that in addition to having a specific

subject dependency, the feature presents a specific channel dependency since the MRCP amplitudes and latencies appear to differ from channel to channel. Finally, from the topographic plots below, a spatial view of the behavior just described is shown, in particular in the first two instants prior to the generation of the MRCP, a low activation of the frontal area of the brain can be noticed. While observing the behavior in the instant prior to the MRCP peak, it can be seen a high activation first in the frontal area followed by a propagation of the potential throughout the motor area.

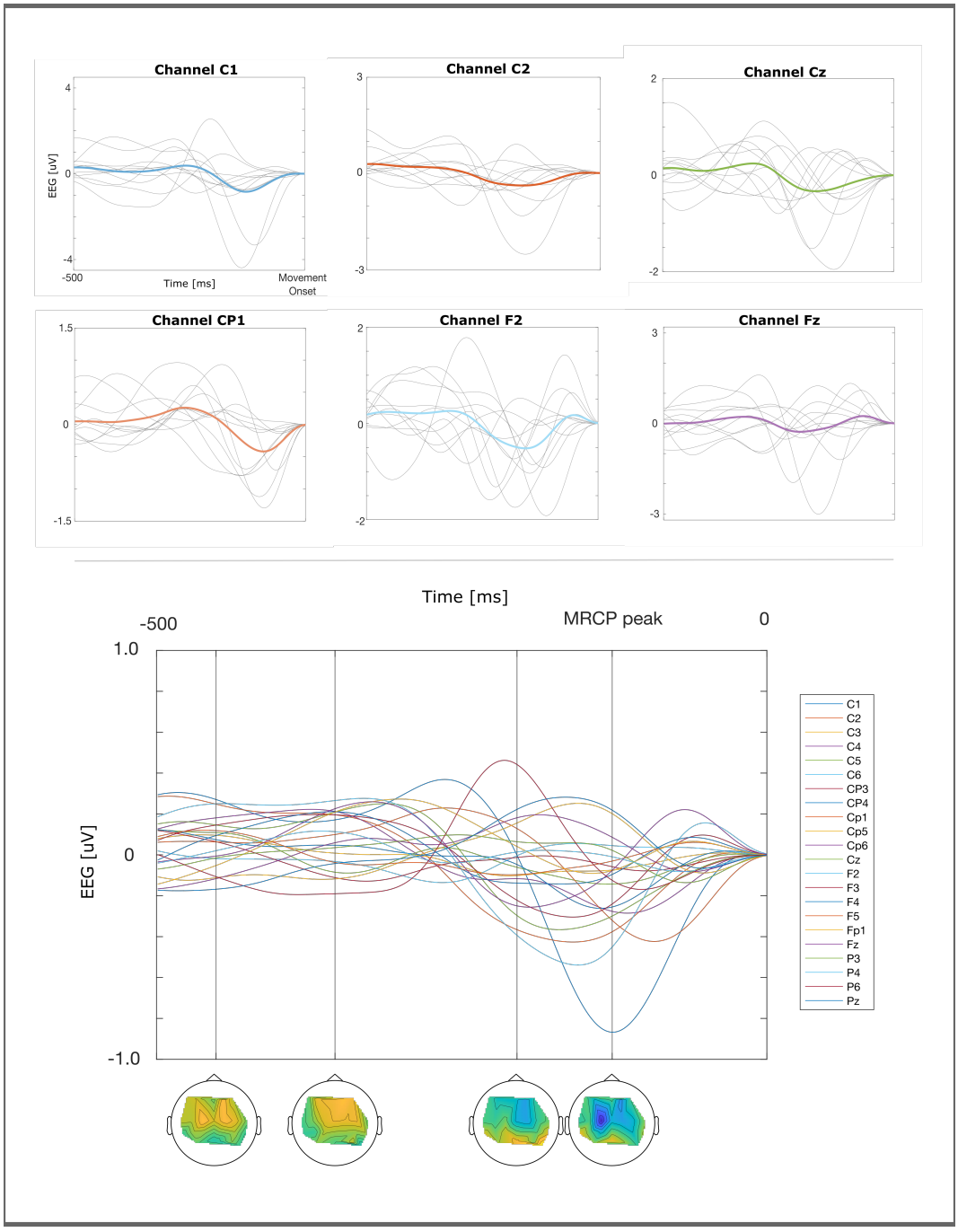

Figure 4.2 Grand-average EEG amplitude time course. On top, grand average of EEG signals on channels C1, C2, Cz, CP1, F2 and Fz across all subjects (colored lines) and for each subject (gray lines).

On bottom grand average across all subjects on the channels in which MRCP is most visible (temporal visualization). In the topographic plots below (spatial visualization) it can be seen how potentials migrate during the instants prior to the onset of MRCP and in its peak of maximum negativity

4.2 Network Optimization

In order to optimize the selection of the hyper-parameter settings of the model presented in Section 3.3, a preliminary grid search was launched. In this first phase, the performance evaluation was carried out through the accuracy and cross-entropy loss metrics, using 50% of the whole dataset (training set) to which a five-fold increase in data augmentation was applied, and was used to predict the validation set, augmented in the same way as the training set. Figure 4.3 (top) shows the results of this first analysis. In the two color maps, the batch size and dropout values are shown and the best overall performance is obtained with values of 8 and 0.1, respectively, obtaining 93.5±4.5% in accuracy and 0.30±0.21 in loss.

Secondly, I analysed the dependency of the classification performance with respect to the network size by setting the batch size and dropout hyper-parameters (8 and 0.1 respectively) and proceeding with a second grid search, based on increasing the number of hidden units of the GRU layer.

The results are shown in Figure 4.3 (bottom). The plots represents the median (solid line) and the subject-specific (dots) performance as the number of hidden units increases, considering a network trained with 15%, 30%, 50% and 100% percentages of the training set. The trend highlighted shows a general increase in validation accuracy and a decrease in validation loss as the number of hidden units increases. This trend flattens around 400 hidden units. A further increase of the number of units leads to a plateau, or even a drop, of the performance. Given the little difference in performance between the networks with 400, 500 and 600 hidden units, it was decided to select as candidate network the one with the lowest number of hidden units, that obtains 96.2±2.9% in accuracy and 0.21±0.16 in loss on average in the model

trained with 50% of training data.

4.3 Data Augmentation Results

The last analysis done in this work concerns the behavior that the identified model assumes with respect to data augmentation.

The color maps in Figure 4.4 represent the performance of the optimal model identified by the previous analyses (400 hidden units, batch size 8 and dropout 0.1) with respect to the percentage of training set used to train the model (7%, 15%, 30%, 50% and 100%) and to the applied data augmentation (no augmentation, 2x, 3x, 5x). The reported results were obtained on the test set, augmented in the same way as the corresponding training dataset.

From the bi-variate analysis (color maps), it is evident that larger amounts of data lead to an improvement in test accuracy and loss. In particular we see that this trend besides being present to the increase of the percentage of the training set used (columns) for the training, datum pertinent to the literature, it is present also when the data on which the model is trained with artificially generated data (rows). In the figures below we can see some aspects in more detail. In the second row we can better see what has just been described, the mean shows the progression of performance as the percentage of training set used increases, a trend that is mirrored in all subjects considered. While, from the results in the third row, see also the Tables 4.1, it emerges the fact that the network trained with a too small percentage of training set (7%) tends to overfit. It is also interesting to note, that between a training without data augmentation and one with a 5x data augmentation the performance has clearly improved, in particular we see that with the use

of a 30% dataset the difference in performance is 11.6 in accuracy and 0.85 in loss. In addition, in a model trained with 100% of the data increased from 3x to 5x there is a difference in loss and accuracy of 0.8 and 0.04 respectively while with 15% of the data with the same augmentation what emerges is difference of 9 in accuracy and 0.61 in loss, so it can be seen that there is no net improvement in performance when using a large dataset.

Accuracy

Training set [%]	**x1**	**x2**	**x3**	**x5**
7%	74.1 ± 13.1	69.5 ± 10.6	67.7 ± 9.8	76.7 ± 10.4
15%	78.8 ± 12.1	79.2 ± 10.8	80.7 ± 9.8	89.7 ± 7.7
30%	83.5 ± 11.5	87.6 ± 8.9	90.4 ± 9.0	95.1 ± 3.4
50%	88.3 ± 8.1	93.9 ± 4.1	94.8 ± 4.1	96.7 ± 2.3
100%	92.1 ± 5.9	96.3 ± 2.7	97.0 ± 1.8	97.8 ± 1.3

Loss

Training set [%]	**x1**	**x2**	**x3**	**x5**
7%	1.86 ± 1.0	2.13 ± 0.77	2.26 ± 0.73	1.47 ± 0.70
15%	1.43 ± 0.80	1.32 ± 0.72	1.17 ± 0.66	0.56 ± 0.45
30%	1.11 ± 0.74	0.80 ± 0.62	0.53 ± 0.51	0.26 ± 0.18
50%	0.69 ± 0.47	0.34 ± 0.25	0.29 ± 0.22	0.16 ± 0.12
100%	0.49 ± 0.4	0.20 ± 0.16	0.16 ± 0.11	0.12 ± 0.07

Table 4.1 Data Augmentation results. Tables that shows the results (accuracy and loss) on the test set (25% of the data) of networks trained with 7%, 15%, 30%, 50% and 100% of the training set data with a specific augmentation. (The test set has the same augmentation)

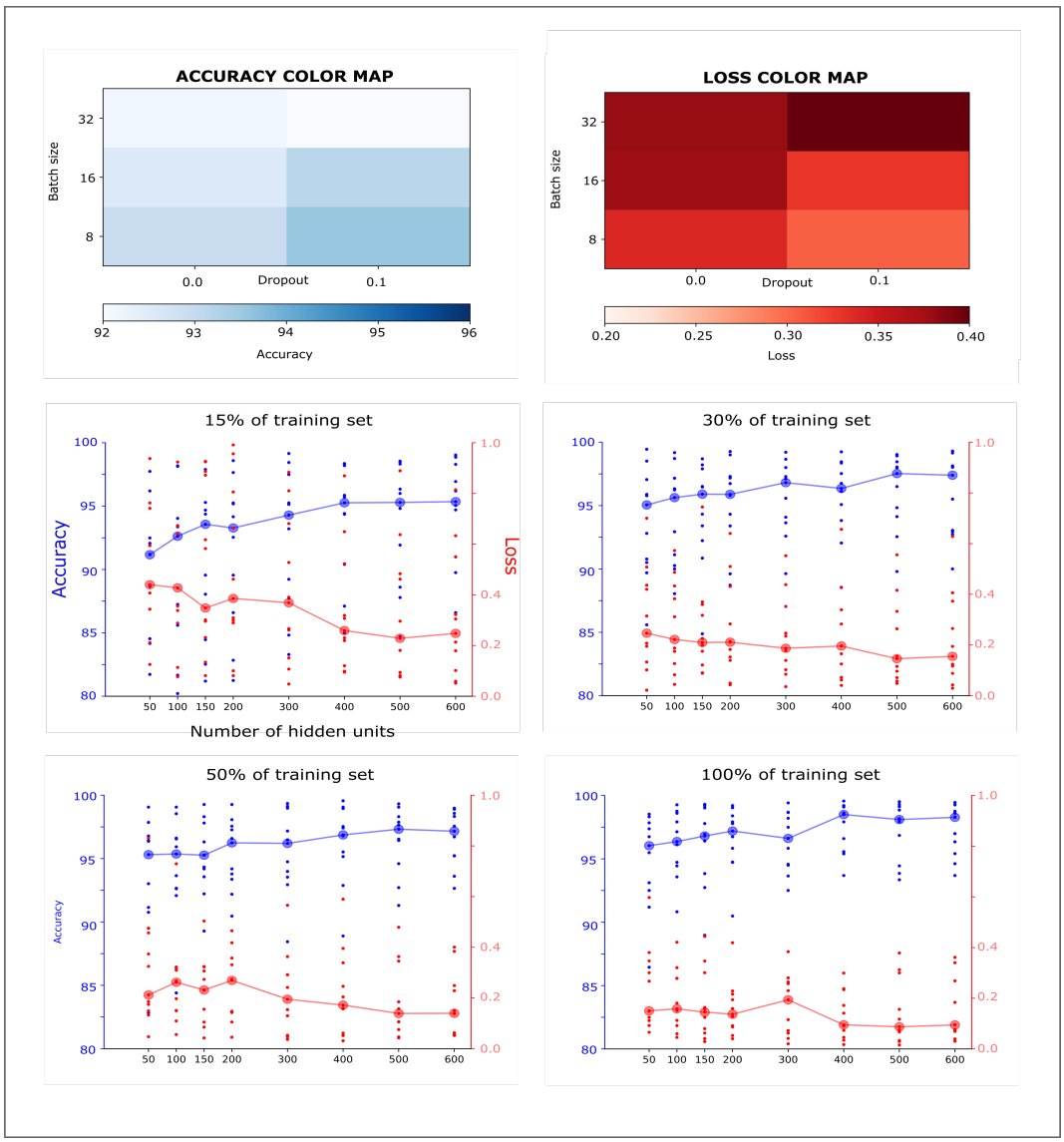

Figure 4.3 Network Optimization results. GRU network performance in terms of validation accuracy (blue) and loss (red) with different architectures (i.e., lot size, dropout, number of hidden units). The color maps represent the average performance of the GRU network, trained with 50% of the dataset augmented (3x) with a number of hidden units equal to 50. The bottom plots, represents the median (solid line) and each point represents the performance of each subject as the number of hidden units varies with a 15%, 30%, 50% and 100% training set, with batch size 8 and dropout 0.1.

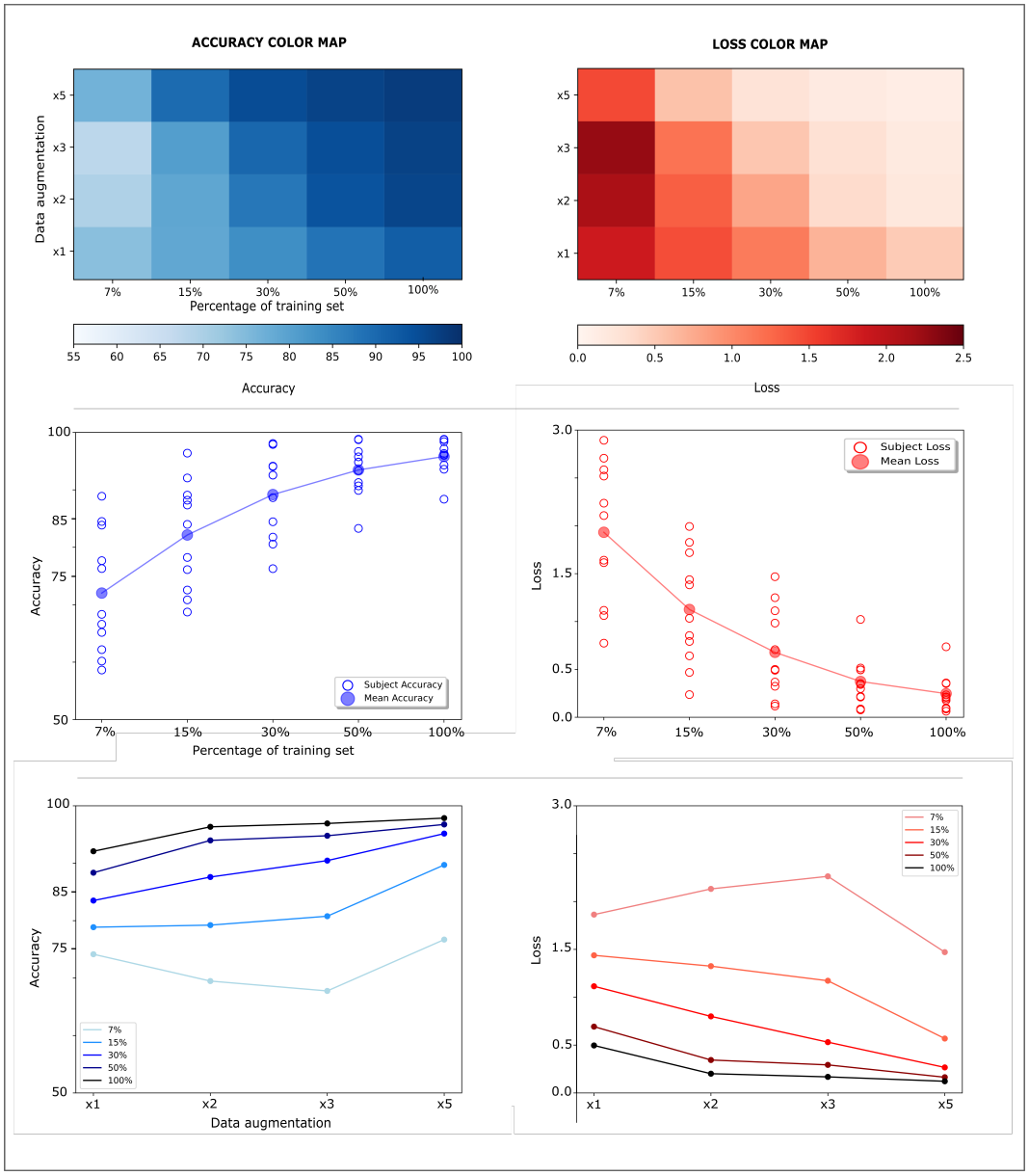

Figure 4.4 Data Augmentation Results. The two colormaps show the grid search results in test accuracy and cross entropy loss obtained with the selected model (400 GRU hidden units, batch size 8 and dropout 0.1) using different percentages of dataset (7%, 15%, 30%, 50%, 100%) and different data augmentation (not increased, duplicated, tripled, quadrupled). In the second row we find a mono-variate analysis, including average performance (test accuracy and loss) on all subjects and performance of individual subjects. In the third line it is present the average performance on all the subjects (for each percentage of training set used for the training) on the various quantities of data augmentation.

5 | Discussion

In this study, I demonstrated the ability to robustly identify step intention through MRCP decoding.

It has been seen that the MRCP is seen in all subjects, in particular as a negative flexion of the signal in the 500 ms prior to the movement onset. However, as it can be seen in Figure 4.1, the characteristics of the MRCP feature appear to be strongly subject-specific and channel-dependent. The MRCP appears earlier in some channels than in others (Figure 4.2), consistently with the neuroscientific literature: the movement intention is generated frontally (e.g., F2, Fz), propagate to the central motor area (e.g., Cz) and then to the rest of the sensorimotor cortex (e.g., C1, CP1) [80, 81]. Given this findings, the selection of a recurrent model has been made to deal with the time and amplitude variability of the feature of interest. Indeed, the GRU layer of the proposed model is capable to encode the temporal dependencies of the input data. In this way the model is able to understand if in a given temporal window there is the presence of the MRCP, and therefore of a pre-intention of movement, or if this correlate is not present and then the subject does not intend to perform any movement.

For the purpose of solving the gait decoding problem using deep learning, several tests and different evaluations were performed in terms of network architecture and hyper-parameters setting. Moreover, to minimize the risk

of overfitting and thus prove the robustness of the proposed approach, three completely disjoint datasets were considered for training, validation and testing [82]. In particular, network performance was evaluated on data that was not available during the network optimization and training processes. Despite the good performance that the GRU model has filed in preliminary tests, for the purpose of a better understanding of its behavior it was necessary to identify the set of hyper-parameters that best generalized our problem. The set of hyper-parameters was selected in two steps: the first grid search allowed the identification of the best model performance with respect to batch size and dropout percentage. Then, a second grid search was adopted to optimize the number of hidden units in the GRU layer, as shown in Figure 4.3. In line with previous works [64], in general, larger networks, in terms of number of hidden units of the recurrent layer, store a greater amount of time-dependent information, thus improving the accuracy of classification, as well as the classification loss, for all the considered subjects.
As it was said in Section 4.2, it was noticed that at a certain point the performance of the networks reached a plateau and in this regard the network chosen was the one with 400 hidden units. A limitation of this work is the fact that networks with more than one layer of GRU have not been tested. It is true that increasing the size of the same layer at a certain point does not increase the performance, and this has been verified, but this does not mean that the performance cannot be improved by increasing the depth of the network in terms of number of layers. However, it has to be said that looking at works like [61, 64], it has been seen that in general more than 2 layers are not recommended and that in general very good performances can be obtained with networks with only one layer.

A concluding part of this work was dedicated to studying the impact that

data augmentation has on the performance of the designated model.
In an EEG-based recurrent neural network for gait decoding the amount of data used for training is an important factor for the correct generalization of the problem.
The study of data augmentation was done to understand the benefits it can give on small amounts of data, which reflect for example the first phase of data acquisition of a patient, in which usually 60 to 70 steps are recorded at most. This detail should not be taken for granted because as mentioned in Section 3.4.2, the effort in performing movements by a patient with motor deficit is to be considered a limitation in terms of amount of data recorded with consequent impact on performance after the first training of the neural network.
From the results in Figure 4.4 (third row), it has been seen that models trained with a too small amount of data tends to overfit with data augmentation because the number of original data available are not sufficient to create a certain variability of the input. However, using a slightly larger dataset (15%), the model obtained satisfying performance, with accurcay close to 90% on average. It is important to highlight the fact that as the percentage of training set increases, the data augmentation becomes less important because the data are already enough, while when the data are few the augmentation is useful because it helps to reduce the performance gap. Indeed, by using a third of the available data (30% dataset) and applying a 5x data augmentation, I achieved performance which are close to that obtained with models trained with 50% and 100% of the dataset.

The optimization of network hyperparameters was done in grand average on all the subjects, while the model was trained subject-specifically. This choice was made to overcome two problems: the risk of overfitting and the

training time. Indeed, since only a few data were available for each subject, customize excessively the network on a specific subject can lead to overfitting. Regarding the training time, the optimization procedure is time consuming, and requires hours of process, which are not compatible with the standard duration of a rehabilitation session (about 1 hour). So with the analysis in grand average proposed, it is possible to create a first network for the first session, with generic hyper parameters that in average performs well for all the subjects, and subsequently, as the data are accumulated in the successive sessions, it is possible to customize the hyper parameters in order to have better performance on the single subjects.

Although the proposed technique is able to effectively classify gait patterns from EEG signals, there are some aspects of this study that need to be addressed in future work. The analysis in my work did not only aim to analyze a network for decoding purposes, rather, the aspect of the amount of data available was also evaluated. This aspect in the field of deep learning and in particular in problems related to bio-informatics, as stated in [61], is more relevant than in classical applications (e.g., image classification, speech decoding). In this regard, the amount of data that we expect from a single session of data recording is closer to the 7% of the dataset used in this work, that showed the worst performance as well as decreasing performance with the proposed data augmentation method. Thus, future work should focus on investigating other data augmentation approaches that would allow more satisfactory performance with smaller dataset. Another aspect, which limits an accurate evaluation is the fact that the performances of the models reported derive from an evaluation made only offline. However, what I can say is that the network is compatible with an online application, in fact, it expects the data in chunk (temporal windows) which are sufficiently short

in order to have an elaboration for an online use. In the future, a real-time version of the proposed model will be implemented and tested to control a lower limb exoskeleton within ROS-Neuro [83], a middleware for neurorobotic applications developed by the University of Padova.
Furthermore, validation of the proposed model was done on a walking dataset acquired by means of a treadmill (Section 3.4.1), in [84] it was pointed out that the features generated in the brain under such conditions might present slightly different characteristics than those resulting from natural walking. In this regard, a future development will certainly be to acquire new data of walking over ground of patients, unlike this dataset, with actual motor disabilities, in order to have a validation of the model that best reflects the real case.

6 | Conclusions

Despite showing promising performance in the literature with respect to traditional machine learning models, only a few works have used deep learning for the gait decoding problem from the sole EEG activity [67, 64], and their are not compatible for an online application.

This work overcomes the limitations of the state-of-the-art by proposing the development of a deep learning model for gait decoding that is suitable for online use.

The proposed approach, exploiting information from movement-related neural correlates, has been shown to successfully decode the pre-intention to take a step during the locomotion.

In summary, my results support the suitability of DL models for decoding time-related information associated to the movement. In particular, it has been demonstrated that the use of data augmentation can be a valid methodology to overcome the problem related to the scarcity of initial data due to the limited ability of patients with motor disabilities to make long EEG recordings.

Future developments will include the integration of a lower limb exoskeleton for real-time testing, and future analyses will evaluate performance on end users (i.e., stroke or SCI patients) in clinical settings and should explore the suitability of the proposed approach for examining cortical plasticity during

rehabilitation.

Bibliography

[1] M. Kyrarini, F. Lygerakis, A. Rajavenkatanarayanan, C. Sevastopoulos, H. R. Nambiappan, K. K. Chaitanya, A. R. Babu, J. Mathew, and F. Makedon, "A survey of robots in healthcare," *Technologies*, vol. 9, no. 1, 2021.

[2] C. S. Mang, K. L. Campbell, C. J. Ross, and L. A. Boyd, "Promoting Neuroplasticity for Motor Rehabilitation After Stroke: Considering the Effects of Aerobic Exercise and Genetic Variation on Brain-Derived Neurotrophic Factor," *Physical Therapy*, vol. 93, pp. 1707–1716, 12 2013.

[3] P. Voss, M. E. Thomas, J. M. Cisneros-Franco, and E. de Villers-Sidani, "Dynamic brains and the changing rules of neuroplasticity: Implications for learning and recovery," *Frontiers in Psychology*, vol. 8, 2017.

[4] E. V. Cooke, K. Mares, A. Clark, R. C. Tallis, and V. M. Pomeroy, "The effects of increased dose of exercise-based therapies to enhance motor recovery after stroke: a systematic review and meta-analysis," *BMC medicine*, vol. 8, no. 1, pp. 1–13, 2010.

[5] M.-G. Kang, S. J. Yun, H. I. Shin, E. Kim, H. H. Lee, B.-M. Oh, and H. G. Seo, "Effects of robot-assisted gait training in patients with parkin-

son's disease: Study protocol for a randomized controlled trial," *Trials*, vol. 20, no. 1, pp. 1–8, 2019.

[6] R. Berriozabalgoitia, B. Sanz, A. B. Fraile-Bermúdez, E. Otxoa, I. Yeregui, I. Bidaurrazaga-Letona, I. Duñabeitia, A. Antigüedad, M. Domercq, J. Irazusta, and A. Rodriguez-Larrad, "An overground robotic gait training program for people with multiple sclerosis: A protocol for a randomized clinical trial," *Frontiers in Medicine*, vol. 7, 2020.

[7] M. Sebastián-Romagosa, W. Cho, R. Ortner, and C. Guger, "Brain computer interface treatment for gait rehabilitation of stroke patients – preliminary results," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2948–2951, 2021.

[8] M. Atzori and H. Müller, "Control capabilities of myoelectric robotic prostheses by hand amputees: A scientific research and market overview," *Frontiers in Systems Neuroscience*, vol. 9, 2015.

[9] F. B. Wagner, J.-B. Mignardot, L. Goff-Mignardot, G. Camille, R. Demesmaeker, S. Komi, M. Capogrosso, A. Rowald, I. Seáñez, M. Caban, *et al.*, "Targeted neurotechnology restores walking in humans with spinal cord injury," *Nature*, vol. 563, no. 7729, pp. 65–71, 2018.

[10] K. Lo, M. Stephenson, and C. Lockwood, "Effectiveness of robotic assisted rehabilitation for mobility and functional ability in adult stroke patients: a systematic review," *JBI Database of Systematic Reviews and Implementation Reports*, vol. 15, pp. 3049–3091, 12 2017.

[11] H. Rodgers, H. Bosomworth, H. Krebs, F. van Wijck, D. Howel, N. Wilson, T. Finch, N. Alvarado, L. Ternent, C. Fernandez, L. Aird, S. Andole, D. Cohen, J. Dawson, G. Ford, R. Francis, S. Hogg, N. Hughes, C. Price,

and L. Shaw, "Robot-assisted training compared with an enhanced upper limb therapy programme and with usual care for upper limb functional limitation after stroke: the ratuls three-group rct," *Health Technology Assessment*, vol. 24, pp. 1–232, 10 2020.

[12] M. Iosa, G. Morone, A. Cherubini, and S. Paolucci, "The three laws of neurorobotics: A review on what neurorehabilitation robots should do for patients and clinicians," *Journal of Medical and Biological Engineering*, vol. 36, 02 2016.

[13] P. Hamet and J. Tremblay, "Artificial intelligence in medicine," *Metabolism*, vol. 69, pp. S36–S40, 2017. Insights Into the Future of Medicine: Technologies, Concepts, and Integration.

[14] C. Chisari, F. Bertolucci, V. Monaco, M. Venturi, C. Simonella, S. Micera, and B. Rossi, "Robot-assisted gait training improves motor performances and modifies motor unit firing in poststroke patients," *European journal of physical and rehabilitation medicine*, vol. 51, 01 2014.

[15] K. J. Nolan, K. K. Karunakaran, K. Chervin, M. R. Monfett, R. K. Bapineedu, N. N. Jasey, and M. Oh-Park, "Robotic exoskeleton gait training during acute stroke inpatient rehabilitation," *Frontiers in Neurorobotics*, vol. 14, 2020.

[16] H. Kim, J.-H. Shin, S. Yang, M. Shin, and S. Lee, "Robot-assisted gait training for balance and lower extremity function in patients with infratentorial stroke: A single-blinded randomized controlled trial," *Journal of NeuroEngineering and Rehabilitation*, vol. 16, 07 2019.

[17] H. Lee, P. W. Ferguson, and J. Rosen, "Chapter 11 - lower limb exoskeleton systems—overview," in *Wearable Robotics* (J. Rosen and P. W. Ferguson, eds.), pp. 207–229, Academic Press, 2020.

[18] V. R. Edgerton and R. R. Roy, "Robotic training and spinal cord plasticity," *Brain research bulletin*, vol. 78, no. 1, pp. 4–12, 2009.

[19] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain–computer interfaces for communication and rehabilitation," *Nature Reviews Neurology*, vol. 12, no. 9, pp. 513–525, 2016.

[20] N. K. Latham, D. U. Jette, M. Slavin, L. G. Richards, A. Procino, R. J. Smout, and S. D. Horn, "Physical therapy during stroke rehabilitation for people with different walking abilities," *Archives of physical medicine and rehabilitation*, vol. 86, no. 12, pp. 41–50, 2005.

[21] S. K. Banala, S. H. Kim, S. K. Agrawal, and J. P. Scholz, "Robot assisted gait training with active leg exoskeleton (alex)," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 17, no. 1, pp. 2–8, 2008.

[22] S. K. Banala, S. H. Kim, S. K. Agrawal, and J. P. Scholz, "Robot assisted gait training with active leg exoskeleton (alex)," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 17, no. 1, pp. 2–8, 2008.

[23] A. Nijholt, D. Tan, G. Pfurtscheller, C. Brunner, J. d. R. Millán, B. Allison, B. Graimann, F. Popescu, B. Blankertz, and K.-R. Müller, "Brain-computer interfacing for intelligent systems," *IEEE Intelligent Systems*, vol. 23, no. 3, pp. 72–79, 2008.

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

[25] J. del R. Millán, "The human-computer connection: An overview of brain-computer interfaces," *Metode Science Studies Journal*, vol. 0, no. 9, 2019.

[26] N. Birbaumer, A. R. Murguialday, and L. Cohen, "Brain–computer interface in paralysis," *Current opinion in neurology*, vol. 21, no. 6, pp. 634–638, 2008.

[27] M. Van Gerven, J. Farquhar, R. Schaefer, R. Vlek, J. Geuze, A. Nijholt, N. Ramsey, P. Haselager, L. Vuurpijl, S. Gielen, *et al.*, "The brain–computer interface cycle," *Journal of neural engineering*, vol. 6, no. 4, p. 041001, 2009.

[28] K. R. Steyrl D. and M.-P. G., "On similarities and differences of invasive and non-invasive electrical brain signals in brain-computer interfacing," *Journal of Biomedical Science and Engineering*, no. 9, 2016.

[29] N. J. Hill, D. Gupta, P. Brunner, A. Gunduz, M. A. Adamo, A. Ritaccio, and G. Schalk, "Recording human electrocorticographic (ecog) signals for neuroscientific research and real-time functional cortical mapping," *JoVE (Journal of Visualized Experiments)*, no. 64, p. e3993, 2012.

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

[31] U. Salahuddin and P.-X. Gao, "Signal generation, acquisition, and processing in brain machine interfaces: A unified review," *Frontiers in Neuroscience*, vol. 15, 2021.

[32] C. D. Binnie and P. F. Prior, "Electroencephalography.," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 57, no. 11, pp. 1308–1319, 1994.

[33] N. Srinivasan, "Cognitive neuroscience of creativity: Eeg based approaches," *Methods*, vol. 42, no. 1, pp. 109–116, 2007. Neurocognitive Mechanisms of Creativity: A Toolkit.

[34] R. Abreu, A. Leal, and P. Figueiredo, "Eeg-informed fmri: a review of data analysis methods," *Frontiers in human neuroscience*, vol. 12, p. 29, 2018.

[35] T. C. Handy, *Event-related potentials: A methods handbook*. MIT press, 2005.

[36] S. Sur and V. Sinha, "Event-related potential: An overview," *Industrial Psychiatry Journal*, vol. 18, no. 1, pp. 70–73, 2009.

[37] J. S. Kumar and P. Bhuvaneswari, "Analysis of electroencephalography (eeg) signals and its categorization–a study," *Procedia engineering*, vol. 38, pp. 2525–2536, 2012.

[38] 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.

[39] J. Polich, "Updating p300: an integrative theory of p3a and p3b," *Clinical neurophysiology*, vol. 118, no. 10, pp. 2128–2148, 2007.

[40] M. D. Golub, S. M. Chase, A. P. Batista, and M. Y. Byron, "Brain–computer interfaces for dissecting cognitive processes underlying sensorimotor control," *Current opinion in neurobiology*, vol. 37, pp. 53–58, 2016.

[41] U. Strehl, U. Leins, G. Goth, C. Klinger, T. Hinterberger, and N. Birbaumer, "Self-regulation of slow cortical potentials: a new treatment for children with attention-deficit/hyperactivity disorder," *Pediatrics*, vol. 118, no. 5, pp. e1530–e1540, 2006.

[42] M. Hallett, "Movement-related cortical potentials.," *Electromyography and clinical neurophysiology*, vol. 34, no. 1, pp. 5–13, 1994.

[43] H. Shibasaki and M. Hallett, "What is the bereitschaftspotential?," *Clinical Neurophysiology*, vol. 117, no. 11, pp. 2341–2356, 2006.

[44] A. I. Sburlea, L. Montesano, R. C. de la Cuerda, I. M. Alguacil Diego, J. C. Miangolarra-Page, and J. Minguez, "Detecting intention to walk in stroke patients from pre-movement eeg correlates," *Journal of neuroengineering and rehabilitation*, vol. 12, no. 1, pp. 1–12, 2015.

[45] E. Yin, Z. Zhou, J. Jiang, F. Chen, Y. Liu, and D. Hu, "A novel hybrid bci speller based on the incorporation of ssvep into the p300 paradigm," *Journal of neural engineering*, vol. 10, no. 2, p. 026012, 2013.

[46] K. Lin, A. Cinetto, Y. Wang, X. Chen, S. Gao, and X. Gao, "An online hybrid bci system based on ssvep and emg," *Journal of neural engineering*, vol. 13, no. 2, p. 026020, 2016.

[47] S. Tortora, L. Tonin, C. Chisari, S. Micera, E. Menegatti, and F. Artoni, "Hybrid human-machine interface for gait decoding through bayesian

fusion of eeg and emg classifiers," *Frontiers in Neurorobotics*, vol. 14, 2020.

[48] P. Velu and V. R. de Sa, "Single-trial classification of gait and point movement preparation from human eeg," *Frontiers in neuroscience*, vol. 7, p. 84, 2013.

[49] A. Presacco, R. Goodman, L. Forrester, and J. L. Contreras-Vidal, "Neural decoding of treadmill walking from noninvasive electroencephalographic signals," *Journal of neurophysiology*, vol. 106, no. 4, pp. 1875–1887, 2011.

[50] T. P. Luu, Y. He, S. Brown, S. Nakagome, and J. L. Contreras-Vidal, "Gait adaptation to visual kinematic perturbations using a real-time closed-loop brain–computer interface to a virtual reality avatar," *Journal of neural engineering*, vol. 13, no. 3, p. 036006, 2016.

[51] J. Choi and H. Kim, "Real-time decoding of eeg gait intention for controlling a lower-limb exoskeleton system," in *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–3, 2019.

[52] S. Y. Gordleeva, M. Lukoyanov, S. Mineev, M. Khoruzhko, V. Mironov, A. Y. Kaplan, and V. Kazantsev, "Exoskeleton control system based on motor-imaginary brain–computer interface," vol. 9, no. 3 (eng), pp. 31–36, 2017.

[53] A. Kilicarslan, S. Prasad, R. G. Grossman, and J. L. Contreras-Vidal, "High accuracy decoding of user intentions using eeg to control a lower-body exoskeleton," in *2013 35th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5606–5609, IEEE, 2013.

[54] C. E. King, P. T. Wang, C. M. McCrimmon, C. C. Chou, A. H. Do, and Z. Nenadic, "The feasibility of a brain-computer interface functional electrical stimulation system for the restoration of overground walking after paraplegia," *Journal of neuroengineering and rehabilitation*, vol. 12, no. 1, pp. 1–11, 2015.

[55] X. Long, D.-X. Liu, S. Liang, Z. Yan, and X. Wu, "An eeg-based bci system for controlling lower exoskeleton to step over obstacles in realistic walking situation," in *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 1609–1614, IEEE, 2018.

[56] D. P. Murphy, O. Bai, A. S. Gorgey, J. Fox, W. T. Lovegreen, B. W. Burkhardt, R. Atri, J. S. Marquez, Q. Li, and D.-Y. Fei, "Electroencephalogram-based brain–computer interface and lower-limb prosthesis control: A case study," *Frontiers in neurology*, vol. 8, p. 696, 2017.

[57] T. C. Bulea, S. Prasad, A. Kilicarslan, and J. L. Contreras-Vidal, "Sitting and standing intention can be decoded from scalp eeg recorded prior to movement execution," *Frontiers in neuroscience*, vol. 8, p. 376, 2014.

[58] R. Xu, N. Jiang, N. Mrachacz-Kersting, C. Lin, G. A. Prieto, J. C. Moreno, J. L. Pons, K. Dremstrup, and D. Farina, "A closed-loop brain–computer interface triggering an active ankle–foot orthosis for inducing cortical neural plasticity," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 7, pp. 2092–2101, 2014.

[59] E. Lopez-Larraz, F. Trincado-Alonso, V. Rajasekaran, S. Perez-Nombela, A. J. Del-Ama, J. Aranda, J. Minguez, A. Gil-Agudo, and

L. Montesano, "Control of an ambulatory exoskeleton with a brain–machine interface for spinal cord injury gait rehabilitation," *Frontiers in neuroscience*, vol. 10, p. 359, 2016.

[60] F. Alton, L. Baldey, S. Caplan, and M. Morrissey, "A kinematic comparison of overground and treadmill walking," *Clinical biomechanics*, vol. 13, no. 6, pp. 434–440, 1998.

[61] 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, p. 031001, 2019.

[62] T. Triwiyanto, I. P. A. Pawana, and M. H. Purnomo, "An improved performance of deep learning based on convolution neural network to classify the hand motion by evaluating hyper parameter," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 7, pp. 1678–1688, 2020.

[63] K. M. Tsiouris, V. C. Pezoulas, M. Zervakis, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using eeg signals," *Computers in biology and medicine*, vol. 99, pp. 24–37, 2018.

[64] 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, p. 046011, jul 2020.

[65] S. Nakagome, T. P. Luu, Y. He, A. S. Ravindran, and J. L. Contreras-Vidal, "An empirical comparison of neural networks and machine learn-

ing algorithms for eeg gait decoding," *Scientific reports*, vol. 10, no. 1, pp. 1–17, 2020.

[66] F. Artoni, C. Fanciullacci, F. Bertolucci, A. Panarese, S. Makeig, S. Micera, and C. Chisari, "Unidirectional brain to muscle connectivity reveals motor cortex control of leg muscles during stereotyped walking," *Neuroimage*, vol. 159, pp. 403–416, 2017.

[67] S. Nakagome, T. P. Luu, Y. He, A. Sujatha Ravindran, and J. Contreras-Vidal, "An empirical comparison of neural networks and machine learning algorithms for eeg gait decoding," *Scientific Reports*, vol. 10, 03 2020.

[68] A. I. Sburlea, L. Montesano, R. C. de la Cuerda, I. M. Alguacil Diego, J. C. Miangolarra-Page, and J. Minguez, "Detecting intention to walk in stroke patients from pre-movement eeg correlates," *Journal of neuroengineering and rehabilitation*, vol. 12, no. 1, pp. 1–12, 2015.

[69] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[70] Q. Liu and Y. Wu, *Supervised Learning*. Springer, 2012.

[71] P. J. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[72] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

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

[74] F. Gennaro and E. D. De Bruin, "Assessing brain–muscle connectivity in human locomotion through mobile brain/body imaging: Opportunities, pitfalls, and future directions," *Frontiers in public health*, vol. 6, p. 39, 2018.

[75] F. Artoni, A. Barsotti, E. Guanziroli, S. Micera, A. Landi, and F. Molteni, "Effective synchronization of eeg and emg for mobile brain/body imaging in clinical settings," *Frontiers in human neuroscience*, vol. 11, p. 652, 2018.

[76] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution eeg and erp measurements," *Clinical neurophysiology*, vol. 112, no. 4, pp. 713–719, 2001.

[77] J. T. Gwin, K. Gramann, S. Makeig, and D. P. Ferris, "Removal of movement artifact from high-density eeg recorded during walking and running," *Journal of neurophysiology*, vol. 103, no. 6, pp. 3526–3534, 2010.

[78] J. T. Gwin, K. Gramann, S. Makeig, and D. P. Ferris, "Electrocortical activity is coupled to gait cycle phase during treadmill walking," *Neuroimage*, vol. 54, no. 2, pp. 1289–1296, 2011.

[79] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, and H. Xu, "Time series data augmentation for deep learning: A survey," *arXiv preprint arXiv:2002.12478*, 2020.

[80] P. Haggard, "Human volition: towards a neuroscience of will," *Nature Reviews Neuroscience*, vol. 9, no. 12, p. 934, 2008.

[81] R. J. Kobler, E. Kolesnichenko, A. I. Sburlea, and G. R. Müller-Putz, "Distinct cortical networks for hand movement initiation and directional processing: an eeg study," *NeuroImage*, vol. 220, p. 117076, 2020.

[82] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, vol. 4. Springer, 2006.

[83] G. Beraldo, S. Tortora, E. Menegatti, and L. Tonin, "Ros-neuro: implementation of a closed-loop bmi based on motor imagery," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2031–2037, IEEE, 2020.

[84] F. Artoni, C. Fanciullacci, F. Bertolucci, A. Panarese, S. Makeig, S. Micera, and C. Chisari, "Unidirectional brain to muscle connectivity reveals motor cortex control of leg muscles during stereotyped walking," *Neuroimage*, vol. 159, pp. 403–416, 2017.

[85] A. Valenti, M. Barsotti, D. Bacciu, and L. Ascari, "A deep classifier for upper-limbs motor anticipation tasks in an online bci setting," *Bioengineering*, vol. 8, no. 2, 2021.

[86] T. Vouga, K. Z. Zhuang, J. Olivier, M. A. Lebedev, M. A. Nicolelis, M. Bouri, and H. Bleuler, "Exio—a brain-controlled lower limb exoskeleton for rhesus macaques," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 2, pp. 131–141, 2017.

[87] F. S. M. Jorquera, S. Grassi, P.-A. Farine, and J. L. Contreras-Vidal, "Classification of stance and swing gait states during treadmill walking from non-invasive scalp electroencephalographic (eeg) signals," in *Converging Clinical and Engineering Research on Neurorehabilitation*, pp. 507–511, Springer, 2013.

[88] D. Liu, W. Chen, Z. Pei, and J. Wang, "A brain-controlled lower-limb exoskeleton for human gait training," *Review of Scientific Instruments*, vol. 88, no. 10, p. 104302, 2017.

[89] C. A. E. Kothe and T.-P. Jung, "Artifact removal techniques with signal reconstruction," Apr. 28 2016. US Patent App. 14/895,440.

[90] F. Artoni, D. Menicucci, A. Delorme, S. Makeig, and S. Micera, "Rel-ica: a method for estimating the reliability of independent components," *Neuroimage*, vol. 103, pp. 391–400, 2014.

[91] J. Pereira, A. I. Sburlea, and G. R. Müller-Putz, "Eeg patterns of self-paced movement imaginations towards externally-cued and internally-selected targets," *Scientific reports*, vol. 8, no. 1, pp. 1–15, 2018.

Acknowledgements

At the end of this thesis work that marks the conclusion of my academic career, I would like to express my thanks to my supervisor Professor *Emanuele Menegatti* for giving me the opportunity to work in a field that was new to me and that fascinated me, and in particular I would like to thank my co-supervisor *Stefano Tortora* for guiding me throughout the process and for giving me useful advice that improved my way of approaching research.
A special thanks goes to my *Family* for supporting me during these years. To my girlfriend *Sofia* for always being there to encourage me to do my best. And last but not least, thanks to all the friends with whom I've had a lot of good times over the years.