



Article

# Applications of Brain Wave Classification for Controlling an Intelligent Wheelchair

Maria Carolina Avelar <sup>1</sup>, Patricia Almeida <sup>1</sup>, Brigida Monica Faria <sup>2,3,\*</sup> and Luis Paulo Reis <sup>1,3</sup>

<sup>1</sup> Faculty of Engineering, University of Porto (FEUP), Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal; lpreis@fe.up.pt (L.P.R.)

<sup>2</sup> ESS, Polytechnic of Porto (ESS-P.PORTO), Rua Dr. António Bernardino de Almeida, 400, 4200-072 Porto, Portugal

<sup>3</sup> Artificial Intelligence and Computer Science Laboratory (LIACC—Member of LASI LA), Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal

\* Correspondence: monica.faria@ess.ipp.pt

**Abstract:** The independence and autonomy of both elderly and disabled people have been a growing concern in today's society. Therefore, wheelchairs have proven to be fundamental for the movement of these people with physical disabilities in the lower limbs, paralysis, or other type of restrictive diseases. Various adapted sensors can be employed in order to facilitate the wheelchair's driving experience. This work develops the proof concept of a brain-computer interface (BCI), whose ultimate final goal will be to control an intelligent wheelchair. An event-related (de)synchronization neuro-mechanism will be used, since it corresponds to a synchronization, or desynchronization, in the mu and beta brain rhythms, during the execution, preparation, or imagination of motor actions. Two datasets were used for algorithm development: one from the IV competition of BCIs (A), acquired through twenty-two Ag/AgCl electrodes and encompassing motor imagery of the right and left hands, and feet; and the other (B) was obtained in the laboratory using an Emotiv EPOC headset, also with the same motor imaginary. Regarding feature extraction, several approaches were tested: namely, two versions of the signal's power spectral density, followed by a filter bank version; the use of respective frequency coefficients; and, finally, two versions of the known method filter bank common spatial pattern (FBCSP). Concerning the results from the second version of FBCSP, dataset A presented an F1-score of 0.797 and a rather low false positive rate of 0.150. Moreover, the correspondent average kappa score reached the value of 0.693, which is in the same order of magnitude as 0.57, obtained by the competition. Regarding dataset B, the average value of the F1-score was 0.651, followed by a kappa score of 0.447, and a false positive rate of 0.471. However, it should be noted that some subjects from this dataset presented F1-scores of 0.747 and 0.911, suggesting that the movement imagery (MI) aptness of different users may influence their performance. In conclusion, it is possible to obtain promising results, using an architecture for a real-time application.

**Keywords:** brain-computer interface; intelligent wheelchair; Emotiv EPOC headset



**Citation:** Avelar, M.C.; Almeida, P.; Faria, B.M.; Reis, L.P. Applications of Brain Wave Classification for Controlling an Intelligent Wheelchair. *Technologies* **2024**, *12*, 80. <https://doi.org/10.3390/technologies12060080>

Academic Editors: Juvenal Rodriguez-Resendiz, Gerardo I. Pérez-Soto, Karla Anhel Camarillo-Gómez, Saul Tovar-Arriaga and Luc de Witte

Received: 9 April 2024

Revised: 15 May 2024

Accepted: 20 May 2024

Published: 3 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Independence and autonomy in mobility are two of the most important conditions for determining the quality of life of people with disabilities or with low mobility capacities [1]. Limited mobility could have origin in a broad range of situations, from accidents to disease to the ageing process. Currently, several mobility-related technologies are designed to achieve independent mobility, in particular powered orthosis, prosthetic devices, and exoskeletons. Notwithstanding these devices, wheeled mobility devices remain among the most used assistive devices [2]. According to the World Health Organization (WHO) [3], approximately 10% of the world's population, or around 740 million people, suffer from disabilities, and, among those people, almost 10% require a wheelchair. Therefore, it is

estimated that about 1% of the total population needs wheelchairs, which translates into 74 million people worldwide [4].

The importance of providing multifaceted wheelchairs that can be adapted to the most diverse conditions of their users is thus emphasised. Different interfaces are being developed, enabling us to overcome existing barriers of use. In particular, special attention has been dedicated to voice control techniques, joysticks, and tongue or head movements. However, hand gesture recognition and brain–computer interface (BCI) systems are proving to be interesting methods of wheelchair control due to their accessible price and non-invasiveness. Therefore, BCIs seem the best option to bridge the users’ will and the wheelchair, as they provide a direct pathway between the “mind” and the external world, just by interpreting the user’s brain activity patterns into corresponding commands [5], and, thus, not requiring neuro-muscular control capabilities whatsoever. Furthermore, people desire to be in charge of their motion as much as possible, even if they have lost most of their voluntary muscle control; therefore, BCIs are an exceptional option [6]. Brain–computer interfaces provide control and communication between human intention and physical devices by translating the pattern of brain activity into commands [6]. The flow of a BCI consists of the acquisition of the information from the brain, followed by the data processing, and ending in the output of a control command [7]. Thus, usually, a BCI can be conceptually divided into signal acquisition, pre-processing, and feature extraction and classification; the last three are the interpretation of the first one.

This paper is structured into six sections, beginning with this introduction. The second section addresses the background and state of the art concerning brain–computer interfaces (BCIs) for acquiring and classifying brain activity. Section 3 details the methods and materials employed in the experimental work. In Section 4, the results obtained from various approaches, including a real-time application, are presented. A discussion of the findings is presented in Section 5, followed by conclusions and suggestions for future work.

## 2. Background and State of the Art of BCI Brain Activity Acquisition Methods

Understanding the acquisition methods for brain activity is crucial for the development of effective brain–computer interfaces (BCIs). There are several methods available, but the most commonly used and well-established method is electroencephalography (EEG). EEG is favoured for its low cost, convenience, standardized electrode placement, and well-documented acquisition techniques. Additionally, EEG offers known filtering methods to address noise and ocular artefacts, making it an attractive option for BCI applications.

### 2.1. Signal Acquisition

There are several methods to acquire brain activity that can be fed into a BCI; however, the most used acquisition method is EEG, as it is low-cost and convenient to use. Other factors that make it such an attractive tool are the standardisation of electrode placement, plentiful and well-documented information on acquisition techniques, and being a well-established method with known filtering [8]. Table 1, adapted from [7], compares the different types of methods used to acquire signals for BCI use.

**Table 1.** Properties of brain activity acquisition methods.

	EEG	MEG	NIRS	fMRI	ECoG	MEA	fTCD
Deployment	Non-invasive	Non-invasive	Non-invasive	Non-invasive	Invasive	Invasive	Non-invasive
Measured Activity	Electrical	Magnetic	Hemodynamic	Hemodynamic	Electrical	Electrical	Hemodynamic
Temporal Resolution	Good	Good	Low	Low	High	High	High
Spatial Resolution	Low	Low	Low	Good	Good	High	Low
Portability	High	Low	High	Low	High	High	High
Cost	Low	High	Low	High	High	High	

I. Magnetoencephalography (MEG) is a neuro-imaging technique, which uses the magnetic fields created by the natural currents that flow in the brain to map the brain activity. To do that, it uses magnetometers. The cerebral cortex's sites, which are activated by a stimulus, can be found from the detected magnetic field distribution [9].

II. Near-infrared spectroscopy (NIRS) is a spectroscopic method that uses the near-infrared (NIR) region of the electromagnetic spectrum (from 780 nm to 2500 nm). NIR light can penetrate human tissues; however, it suffers a relatively high attenuation due to the main chromophore haemoglobin (the oxygen transport red blood cell protein), which is presented in the blood. Therefore, when a specific area of the brain is activated, the localised blood volume in that area changes quickly and, if optical imaging is used, it is possible to measure the location and activity of specific regions of the brain. This is due to the continuous tracking of the haemoglobin levels through the determination of optical absorption coefficients [10].

III. Functional magnetic resonance imaging (fMRI), through variations associated with blood flow, can measure brain activity. This technique relies on the fact that cerebral blood flow and neuronal activation are coupled; thus, when an area of the brain is in use, the blood flow to that region also increases [11].

IV. Electrocorticography (ECoG) is a type of electrophysiological monitoring that records activity mainly from the cortical pyramidal cells (neurons). For that, it requires the electrodes to be placed directly on the exposed surface of the brain so that the recorded activity comes directly from the cerebral cortex [12].

V. Micro-electrode arrays (MEAs) are devices that contain multiple microelectrodes; the number can vary from ten to thousands, through which the neural signals are obtained. These arrays function as neural interfaces that connect neurons to electronic circuitry [13].

VI. Functional transcranial Doppler (fTCD) is a technique that uses ultrasound Doppler to measure the velocity of blood flow in the main cerebral arteries during local neural activity [14]. Changes in the velocity of the blood flow are correlated to changes in cerebral oxygen uptake, enabling fTCD to measure brain activity [15].

However, the robustness of all existing BCI systems is not satisfactory due to the non-stationary nature of non-invasive EEG signals. If a BCI system is unstable, other techniques should be further developed to improve the overall driving performance [6]. Usually, these concerns improve feature extraction and classification as the other option would fall on trading to an invasive approach. Although the range of existent commercial headsets is quite good, most of them lack in the number of available electrodes as they are more turned to improve the user's focus and to help to relax, or be used for gaming. Furthermore, the ones that present better characteristics are the Emotiv EPOC, Emotiv Flex, and the Open BCI [16]. Although the last two do not restrict the electrodes' configuration as Emotiv EPOC does, they are more expensive and complex. As for the open BCI one, it does not offer the same freedom of measurement and comfort as Emotiv ones, as these are wireless with a 12 h, for EPOC, and 9 h, for Flex, lasting battery [17]. The cost of an Emotiv is approximately USD 1000, and Open BCI can cost more than USD 2000 [16]. Hereupon, authors nowadays do not use commercial EEG headsets to obtain the signals that will feed the BCI; they prefer assembling their own EEG set through an amplifier and electrodes, as seen in Table 2.

**Table 2.** EEG headsets used in the literature.

Article	EEG Headset	Principle	Article	EEG Headset	Principle
[18]	12 Ag/Cl electrodes	ERP—P300	[19]	NuAmps and 12 electrodes	ERP—P300
[20]	NuAmps and 15 electrodes	ERP—P300	[21]	gTec EEG (16 electrodes and g.USBamp amplifier)	ERP—P300
[22]	16-channel electrode cap	ERP—P300	[23]	Biopac MP150 EEG system	ERP—P300
[24]	gTec EEG (12 electrodes and g.USBamp amplifier)	ERP—P300	[25]	Neuroscan (15 electrodes' cap)	ERP—P300

Table 2. Cont.

Article	EEG Headset	Principle	Article	EEG Headset	Principle
[26]	BioSemi ActiveTwo system 32 channels	SSVEP	[27]	g.USBamp amplifier with g.Butterfly active electrodes	SSVEP
[28]	8 gold electrodes connected to the g.USBamp amplifier	SSVEP	[29]	gTec EEG with g.USBamp amplifier	SSVEP
[30]	EEG Cap and g.USBamp amplifier	SSVEP	[31]	BrainNet-36 with 12 channels	SSVEP
[32]	BrainNet BNT-36 with 3 channels	SSVEP	[33]	6 channels EEG cap	SSVEP
[34]	NeuroSky mindset	ERD/ERS	[35]	Grass Telefactor EEG Twin3 Machine	ERD/ERS
[36]	G-TEC system with 5 Ag/AgCl electrodes	ERD/ERS	[37]	8 channels EEG cap	ERD/ERS
[38]	5 bipolar EEG channels and a g.tec amplifier	ERD/ERS	[39]	ActiveTwo 64-channel EEG system	ERD/ERS
[40]	Emotiv EPOC	ERD/ERS	[41]	Emotiv EPOC	ERD/ERS
[42]	Emotiv EPOC	ERD/ERS	[43]	Emotiv EPOC	ERD/ERS
[44]	EEG Cap—15 electrodes	ERD/ERS and SSVEP	[45]	Gtec Amplifier (15 channels)	ERD/ERS and SSVEP
[46]	g.BSamp amplifier (5 channels)	ERD/ERS and SSVEP	[47]	NuAmps device (15 channels)	ERD/ERS and ERP—P300
[48,49]	NeuroSky	ERP—P300; Eye Blinking (EMG)	[5]	SYMPTOM amplifier with 10 electrodes	ERP—P300 and SSVEP

ERP—event-related potential; ERS—event-related synchronization; SSVEP—steady-state visual evoked potential.

However, these are usually not wireless options. Nevertheless, there is still a significant group who use Emotiv EPOC, as this one offers a wider range of electrodes when compared with other commercial options, allowing obtaining of the signals from different brain lobes. It is possible to find several public EEG datasets related to motor imagery [50]. These datasets involve recordings of brain activity while subjects imagine performing specific motor tasks, such as moving a limb or making a particular gesture. These datasets are essential for studying motor control, brain–computer interfaces (BCIs), and rehabilitation. These datasets cover a wide range of tasks and experimental paradigms. With varying numbers of subjects, electrode configurations, and recording parameters, each dataset offers insights into different aspects of brain function and behaviour. From collections like the largest SCP data of motor imagery, with extensive EEG recordings spanning multiple sessions and participants, to focused datasets like the imagination of right-hand thumb movement, capturing specific motor imagery tasks, these datasets [50] serve as valuable resources for exploring the neural correlates of motor control, emotion processing, error monitoring, and other cognitive processes.

## 2.2. Signal Processing

The signal-processing module is divided into different parts [51]. The steps vary depending on whether the stage is training or testing; however, the training steps are broader than the testing ones, and, hence, these will be the ones to be discussed. The first step is to pre-process the signal, and it is further subdivided into band-pass and spatial filtering; afterwards, the features are extracted and selected. Finally, the classification is carried out, and the performance is evaluated. To perform this, techniques of machine learning must be applied, and thus the brief explanation of this concept.

Machine learning (ML) is based on data analytics that automates analytical model building. By using algorithms that iteratively learn from data, the computer can find

hidden insights without being explicitly programmed where to look. This approach is used when the problem is complex and can be described by many variables. It creates an unknown target function that models the input into the desired output [52]. The learning algorithm receives a set of labelled examples (inputs with corresponding outputs) and learns by comparing its predicted output with correct outputs to find errors, modifying the model accordingly. The resulting model can predict future events. When exposed to new data, the model can adapt itself. In theory, if the algorithm works properly, the larger the amount of data there are, the better are the predictions. However, they are limited by bias in the algorithm and in the data, which can produce systematically skewed predictions. Therefore, the complexity of the learning algorithm is critical and should be balanced with the complexity of the data [52].

### 2.2.1. Pre-Processing

The EEG signal, per se, is very noisy, which is due to several aspects such as the low signal-to-noise ratio—as it is collected from the individual’s scalp surface, the low spatial resolution, and other sources such as artefacts or interfering frequencies [53]. Artefact removal involves cancelling or correcting the artefacts without distorting the signal of interest and can be implemented in both the temporal and spatial domains [54]. Usually, the pre-processing concerns two types of filtering, in the frequency and the spatial domain [51]: band-pass filtering consists of removing some frequencies, or frequency bands, from the signal [53], outputting the frequency range of interest; and spatial filtering, which consists of combining the original sensor signals, usually linearly, which can result in a signal with a higher signal-to-noise ratio than that of individual sensors [51]. It combines the electrodes, which leads to more discriminating signals [54]. According to Pejas [55], approaches that rely on spatial filtering not only provide more true positives but also allow more flexibility when choosing the electrode placement. Spatial filters that linearly combine signals acquired from different EEG channels can extract and enhance the desired brain activity; thus, usually, it is enough to place the electrodes somewhere in the desired area and not in the exact location.

### 2.2.2. Feature Extraction and Classification

There are different types of features according to the domain from where they are extracted: time, frequency, or spatial. Different methods are used to extract the features from the EEG signal and further classify them so that the control commands can be obtained. Table 3, partly adapted from [56], presents a group of techniques used by different authors. It comprises several examples referring to the different principles: ERPs, SSVEP, ERD/ERS, and Hybrid.

**Table 3.** Summary of different authors’ BCIs regarding the used EEG headset, the neuro-mechanism, the extracted features, the classification methods, the outputted commands, and accuracy.

Article	EEG Headset	Principle	Features	Classifier	Control	Accuracy
[18]	12 Ag/Cl electrodes	ERP—P300	Signal averaging and standard deviation	2 class Bayesian	L/R/F/B (45° or 90°)/S	95%
[19]	NuAmps and 12 electrodes	ERP—P300	Data vectors of concatenated epochs	BLDA (Bayesian)	9 destinations	89.6%
[20]	NuAmps and 15 electrodes	ERP—P300	Raw signal	SVM	7 locations, an ‘application button’ and lock	90%
[21]	gTec EEG (16 electrodes and g.USBamp amplifier)	ERP—P300	Moving average technique	LDA	15 locations, L/R and validate selection	94%
[22]	16-channel electrode cap	ERP—P300	Signal averaging	Linear classifier	6 for the IW (not specified)	92%
[23]	Biopac MP150 EEG system	ERP—P300	Signal averaging	Linear classifier	F/B/L/R	—

Table 3. Cont.

Article	EEG Headset	Principle	Features	Classifier	Control	Accuracy
[24]	gTec EEG (12 electrodes and g.USBamp amplifier)	ERP—P300	Optimal statistical spatial filter	Binary Bayesian	F/B/L/R (45° or 90°/S)	88%
[25]	Neuroscan (15 electrodes' cap)	ERP—P300	Signal averaging	SVM	37 locations, validate or delete selection, stop and show extra locations	—
[26]	BioSemi ActiveTwo system 32 channels	SSVEP	Peaks in the frequency magnitude	—	L/R	>95%
[27]	g.USBamp amplifier with g.Butterfly active electrodes	SSVEP	Frequency band power (PSD)	SVM	L/R/F/S	95%
[28]	8 gold electrodes connected to the g.USBamp amplifier	SSVEP	—	LDA	L/R/B/F/S	90%
[29]	gTec EEG with g.USBamp amplifier	SSVEP	Frequency band power (PSD)	Threshold method not specified	L/R/B/F	93.6%
[30]	EEG Cap and g.USBamp amplifier	SSVEP	CCA	Bayesian	F/L/R/turn on/off	87%
[31]	BrainNet-36 with 12 channels	SSVEP	Frequency band power (PSD)	Decision trees	L/R/F/S	Qualitative evaluation
[32]	BrainNet BNT-36 with 3 channels	SSVEP	Frequency band power (PSD)	Statistical maximum	L/R/F/B	95%
[33]	6 channels EEG cap	SSVEP	FFT and CCA	CCA coefficient	L/R/F/B/S	>90%
[34]	NeuroSky mindset	ERD/ERS	Frequency band power (PSD)	NN	Game	91%
[35]	Grass Telefactor EEG Twin3 Machine	ERD/ERS	Coefficients from the wavelets	Radial basis function NN	L/R/F/B/rest	100%
[36]	G-TEC system with 5 Ag/AgCl electrodes	ERD/ERS	Common spatial frequency subspace decomposition (CSFSD)	SVM	L/R/F	91–95%
[37]	8 channels EEG cap	ERD/ERS	Mean, zero-crossing and energy from different levels of the DWT	ANN	L/R/F/S	91%
[38]	5 bipolar EEG channels and a g.tec amplifier	ERD/ERS	Logarithmic frequency band power	LDA	L/R	75%
[39]	ActiveTwo 64-channel EEG system	ERD/ERS	Frequency band power (PSD) and CSP	SVM	Exoskeleton control LH/LF/RH/RF	84%
[40]	Emotiv EPOC	ERD/ERS	PCA and average power of the wavelets' sub-bands	NN w/BP	L/R/F/B	91%
[42]	Emotiv EPOC	ERD/ERS	—	Emotiv program	L/R/F/S	70%
[56]	Emotiv EPOC	ERD/ERS	Frequency components	SVM	L/R/F/B/S	100%
[56]	Emotiv EPOC	ERD/ERS	Frequency components	NN	L/R/F/B/S	100%
[56]	Emotiv EPOC	ERD/ERS	Frequency components	Bayesian	L/R/F/B/S	94%
[56]	Emotiv EPOC	ERD/ERS	Frequency components	Decision trees	L/R/F/B/S	74%
[42]	Emotiv EPOC	ERD/ERS	Frequency band power (PSD)	LDA	L/R	70%
[43]	Emotiv EPOC	ERD/ERS	Metrics from the EEG signal	Decision trees	L/R	82%

Table 3. Cont.

Article	EEG Headset	Principle	Features	Classifier	Control	Accuracy
[57]	Emotiv EPOC	ERD/ERS	CSP	SVM	L/R	60%
[58]	Emotiv EPOC	ERD/ERS	—	LDA	L/R	60%
[59]	Emotiv EPOC	ERD/ERS	PSD, Hjort parameters, CWT and DWT—PCA for feature reduction	K-NN	L/R	86–92%
[60]	Emotiv EPOC	ERD/ERS	Energy distribution from the DWT	SVM	L/R/T/N	97%
[61]	Emotiv EPOC	ERD/ERS	CSP	LDA	L/R	68%
[61]	Emotiv EPOC	ERD/ERS	CSP	SVM	L/R	68%
[61]	Emotiv EPOC	ERD/ERS	CSP	Nu-SVC RBF Kernel	L/R	68%
[44]	EEG Cap—15 electrodes	ERD/ERS—(L/R) and SSVEP	CSP (ERD/ERS); CCA (SSVEP)	SVM (ERD/ERS); Canonical correlation coefficient (SSVEP)	L/R/A/DA, maintain an uniform velocity and turn on/off	—
[45]	Gtec Amplifier (15 channels)	ERD/ERS—(L/R) SSVEP-(Des)accelerate	CSP (ERD/ERS); CCA (SSVEP)	SVM	L/R/A/DA	—
[46]	g.BSamp amplifier (5 channels)	ERD/ERS and SSVEP	Frequency band power (PSD)	LDA	L/R	81%
[47,48]	NuAmps device (15 channels)	ERD/ERS and ERP—P300	CSP	LDA	L/R/A/DA	100%
[48,49]	NeuroSky	ERP—P300 and Eye Blinking (EMG)	Changes in the level	Threshold	L/R/F/B/S	—
[5]	SYMPTOM amplifier with 10 electrodes	ERP—P300 and SSVEP	PCA (ERP); PSD (SSVEP)	LDA	ERP—9 destinations SSVEP—confirm	99%

L—left; R—right; F—forward; B—backward; S—stop; A—accelerate; DA—decelerate; H—hand; F—foot; T—tongue; N—no imaging.

The ERD/ERS neuro-mechanism is a widely used one and has been producing noticeable results. This corresponds to a change in the power of specific frequency bands since the user is imagining or visualising a certain motor movement. The best combination is obtained with SVMs or NN as classifiers. Authors such as Abiyev et al. [57] and Khare et al. [35] achieved an impressive accuracy of 100%. The extracted features were all in the frequency domain, mostly from the frequency coefficients, band power, or spatial filtering. SSVEP BCIs can originate accurate outcomes regardless of the classifier. This is probably due to the neuro-mechanism itself, as it is linked to a specific frequency, facilitating the extraction of the feature vector. However, contrarily to the ERD/ERS BCIs, these require some sort of hardware, usually flashing buttons (each one at a unique frequency rate), which will act as the stimulus for the user. The latter will focus on the button, which represents the desired direction; hence, proportionally amplifying the EEG signal band corresponding to the button frequency. The extracted features fall in the frequency domain and regard the power in specific frequency bands (corresponding to the respective button). ERPs are short amplitude deflections in the brain signal that are timestamped to an event. They are identified by the triggering event, direction of deflection, observed location, and latency [7]. That is why these BCIs usually use temporal features, whereas ERD and SSVEP BCIs employ frequency features [6]. Concerning the used classifier, the BCI performance does not seem to depend upon this choice. Regarding hybrid BCIs, it can be deduced that methods that aim to decompose the signal are preferentially used to extract the features. Concerning the classification, the used classifiers are mainly SVMs and LDA. It is possible to conclude that, depending on the chosen neuro-mechanism, the type of extracted features will differ.

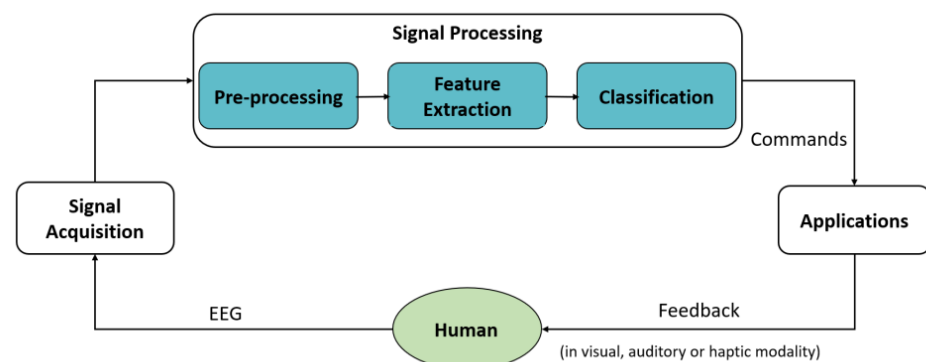
However, for the classifiers, the same cannot be applied, although it is possible to infer that some classifiers have a better performance than others have, namely, SVMs, NN, and LDA.

A BCI provides control and communication between human intention and physical devices by translating the pattern of the brain activity into commands. The goal is to use this as a way of controlling an IW, which will eventually lead to an increase in the quality of life of people with disorders and limitations. A BCI has different main blocks: signal acquisition, signal processing, and the application of the output commands. The first one aims to collect the brain signals to feed them to the signal processing unit. There are several ways of achieving this, with EEG being the most common, affordable, and well-documented way. To make it even more accessible and portable for the patient, the EEG headset should be wireless; hence, the Emotiv EPOC is the chosen one.

Moreover, the aim is to use the Emotiv EPOC headset as a way to record the user's brain activity, as it is rapidly installed and portable. Although many authors have already proposed several solutions, none of them meet the required criteria to be commercialised, either by the lack of portability or the lack of accuracy. Therefore, the final goal would be to have a portable, comfortable, affordable, and reliable solution for an end-user consumer, so that the system would ideally be prepared for an out-of-the-laboratory application. Thus, this work contributes to the conceptualisation of the BCI system, regarding its architecture and algorithms.

### 3. Materials and Methods

A motor imagery (MI) neuro-mechanism is proposed, as it allows the user to focus on the path instead of focusing on the user interface, as the last two are stimulus-dependent neuro-mechanisms. Figure 1 presents the overall scheme of the BCI architecture along with its constituent parts.



**Figure 1.** Scheme of the BCI architecture and its parts.

Three classes for the commands are used, namely left (0), right (1), and neutral (2). The first two correspond to changes in the direction, whereas the last one implies that the subject wishes to maintain the same direction. This choice relies on the fact that the left and right are the basic commands to control a moving device and, since the system is working in a continuum, the necessity of a neutral class to maintain the direction of movement arises. According to Tang et al. [39], some subjects present a better ability to distinguish between the feet and hands, rather than the left hand from the right one. Consequently, three different runs are tested, where the subject can substitute one of the hands for the thought of feet. More specifically, the subject may have a better performance while differentiating the left hand from the feet, and it may be advantageous to use the thought of the feet to turn to the right.

Moreover, the experiments are divided into two main parts: the validation of the concept and the corresponding execution or testing. Regarding the first part, two datasets are used, dataset 2a from the BCI competition IV (dataset A) [62] and another one acquired



in our laboratory using the Emotiv EPOC headset (dataset B). Concerning the execution of the algorithm, a real-time acquisition from the headset is attempted and evaluated.

### 3.1. Datasets

Dataset A contains a four-class MI for different body parts: the left and right hand (LH/RH), feet (F), and tongue. This dataset corresponds to dataset 2a of the BCI competition IV and comprises 2 sessions of 288 trials from 9 different subjects. In each session, there were 6 smaller sessions of 48 trials, each separated by breaks. It also encompasses an evaluation dataset with the same characteristics as the previously described one. For this work, the tongue MI was discarded, as it was not of interest.

The acquisition protocol for each trial can be seen in Figure 2 and it is a sequence composed of a fixation cross (2 s), followed by an arrow representing the desired MI (1.5 s), a period of blank screen for the subject to imagine the asked cue (2.75 s), and it finishes with a break (~2 s). Furthermore, there is a sound alerting for the beginning and ending of the MI period (4 s).

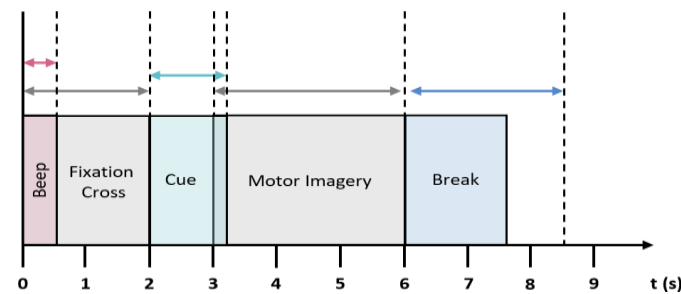


Figure 2. Acquisition protocol for dataset A.

The signals were obtained using 22 Ag/AgCl electrodes, which were positioned following the 10/20 system shown in Figure 3a. These were placed mostly at the central part of the cortex, where the sensorimotor part is located.

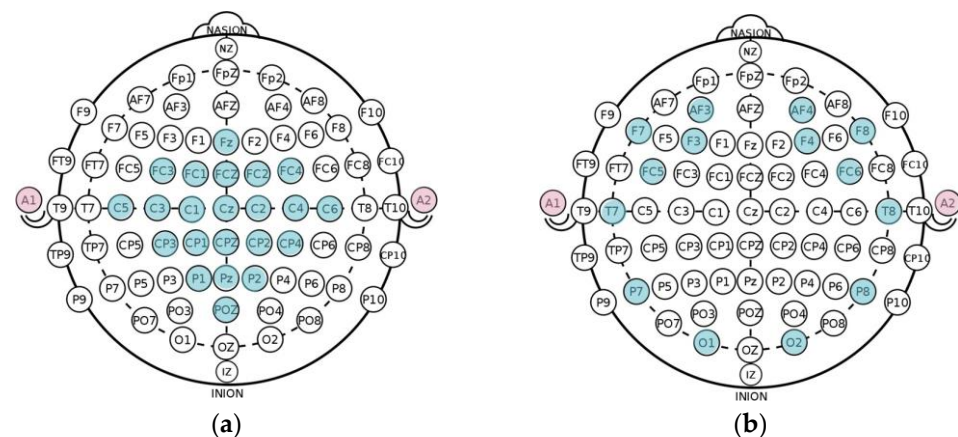
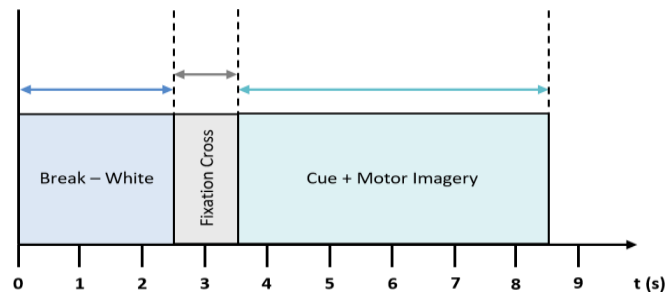


Figure 3. Electrodes' placement, according to the 10/20 system, for both datasets: (a) dataset A; (b) dataset B.

The acquisition protocol for dataset B was approximately the same as for dataset A, with two differences inspired by Tang et al. [39], Dharmasena et al. [42], and Stock and Balbinot [63]. More specifically, in the MI cue, the arrows were displayed on the screen for the whole period, as shown in the diagram presented in Figure 4. Furthermore, the indication of the start of a cue was not used to simplify the process. There were three different cues: right hand (right arrow), left hand (left arrow), and foot (down arrow). The set of sessions comprised 360 trials, 120 for each MI.



**Figure 4.** Acquisition protocol for dataset B.

In total, signals from nine different healthy subjects were acquired, where subjects 4 and 6 are left-handed, while the others are right-handed. All subjects are below 25 years old, except subject 5, who is 51. During acquisition, the subjects were seated comfortably in a chair, in a quiet room, with their hands on the top of the table while looking at the screen; they were also asked to keep their movements, such as eye gazing, sniffing, or coughing, to a minimum. All the procedures were performed under the ethical standards of the 1964 Helsinki Declaration.

The electrode placement can be seen in Figure 3b. Although none of these match the placement for dataset 2a, they still cover part of the central, parietal, and frontal locations of the cortex, which are known for contributing to the MI [14]. However, it is expected that the results will not be as satisfactory, as the centre of the cortex is not covered [61]. It is also important to state that the recorded points may also depend on the format of the subject's head, as electrode placement on narrower heads will not be the same as for wider ones, because the electrodes in the headset are fixed.

### 3.2. Data Processing

The two datasets were divided into training and test as follows:

- Dataset A: the training data supplied by the BCI Competition IV were used as a train and the evaluation one as a test. The duration of the epochs was two seconds, as explained in [62].
- Dataset B: 100 trials of each MI were used as training data, and the remaining 20 were used as tests. Usually, each subject would do a 20-trial session, which results in 5 sessions for training and 1 session for testing. For each visual cue and motor imagery moment, as these had a duration of 5 s, two epochs of two seconds each were extracted, allowing to double the data, ending up with 240 epochs, in total, for each class.
- A subject-oriented approach was followed, requiring the model to undergo training specific to each subject before being tested. However, it should be noted that the sessions utilised for testing differed from those used for training purposes.

#### 3.2.1. Pre-Processing and Feature Extraction

Filtering the EEG signal is already enough to remove noise and ocular artefacts, which are the most common. The first comprises high frequencies, which are discarded, as these are not included in the bands of interest. Moreover, ocular artefacts mainly appear in the theta band, which, once again, is not a band of interest for the MI paradigm. Thus, for every feature extraction approach, presented in the next section, a filtering step is always applied to eliminate these artefacts. The main methods for feature extraction regarding the MI paradigm are spatial filtering using the common spatial pattern (CSP) approach and the use of the signal's frequency band or the frequency coefficients as features. The different approaches were tested, but with some variations. The next steps, feature selection and classification, were the same for all the approaches.

##### 1. Filter Bank Common Spatial Pattern I

As dataset A is from a competition, the first approach was to develop an algorithm based on the winning method, denoted the filter bank common spatial pattern (FBCSP),

as described in [62]. The goal is to maximise the best band for each user, which results in dividing the alpha and beta bands into nine sub-bands, from 4 Hz to 40 Hz. Although Ang et al. [62] use a Chebyshev II filter, in this work a Butterworth filter of order five and zero phase was applied. This choice lies in the facts that this filter is known for being the flattest in the passing band, the zero phase provides zero group distortion, and the order five is a nice compromise with respect to speed. The FBCSP algorithm applies the CSP procedure to each sub-band of the signal. The algorithm generates a linear filter, which is used to extract features that best discriminate between classes, by maximising the ratio between their covariance matrices [64].

## 2. Filter Bank Common Spatial Pattern II

This approach follows the same principles as the first approach but, after obtaining a spatial filter, the average power of  $Z$  is computed and used as features.

## 3. Power Spectral Density

The signal is filtered using a Butterworth filter, for the reasons previously enunciated, from 4 to 35 Hz to comprise the alpha and beta bands. Afterwards, epochs of two seconds are obtained and normalised. The latter consists of centring each channel to have zero mean. For that, the mean of each epoch for each channel is calculated and then subtracted [57]. Afterwards, the Welch method, with a Hanning window, is applied to obtain the power spectral density for each epoch, which is used as the features vector. The Welch method consists of dividing the signal into overlapping segments, which are further windowed. Then, the signal periodogram, which is an estimate of the signal spectral density, is calculated, resorting to the discrete Fourier transform. Windowing the segments, for example with the Hanning window, allows for mitigating spectral leakage. This is because the Fourier transform assumes that the signal is periodic, and non-periodic signals lead to sudden transitions that have a broad frequency response [65]. Different methods for choosing the most significant features were tested, namely a method based on a mutual information criterion, the ANOVA F test, and the extra trees classifier, to compute the features' importance. The first measures the dependency between two random variables and relies on non-parametric methods based on entropy estimation, such as from K-nearest neighbours, to improve the selection. The second assesses the amount of variability between each class mean, in the context of the variation within the different classes, to determine whether the mean differences are statistically significant or not. Finally, the extra trees classifier is used to compute the importance of the features, allowing the irrelevant ones to be discarded. For either of the methods, only the K best features are selected. This is performed by a 10-fold cross-validation, using 5–70% of the features. The 70% limit is imposed to prevent overfitting.

### 3.2.2. Classification

The classifiers were trained to differentiate between three different classes. Due to slower computational time and the fact that they might generate overfitting, non-linear classifiers were not used as a first approach [56]. Thus, four classifiers were trained: Gaussian Naive Bayes (GNB), linear discriminant analysis (LDA), linear support vector machines (LSVMs), and logistic regression (LR). Using these four classifiers, different combinations were tested, as represented in Table 4.

**Table 4.** Testing a combination of classifiers.

Number of Classifiers	Type	Number of Classifiers	
1	Non-probabilistic	2—Ensemble	Voting Hard
	Probabilistic—F1		Voting Soft
2	Non-probabilistic		AdaBoost
	Probabilistic—F1		

When using a single classifier to predict the result, there are two main approaches: predict a class or predict the probability of belonging to each class. In the latter approach, the ideal value for the probability threshold can be obtained through different metrics. The F1-score was the chosen one, as it considers both the precision and the recall of the classifier. When using two classifiers to predict the final command, different approaches were applied, which are further explained:

- Two classifiers: if both classifiers predict that the class is 0, then the class is 0; the same is applied for classes 1 and 2. However, if they do not agree with the classification, then the trial is classified as 2 in order to decrease the number of false positives, which in this case are the trials miss-classified as MI to the left or right;
- Two classifiers with variable probability: the idea behind this approach is the same as before; however, the output of each classifier is a probability and not a class label. Thus, a threshold is estimated for each one of the classifiers to output a label, and then the same method is applied, as explained for the two classifiers;
- Ensemble methods: these methods are already developed and are widely used to combine the different predictions so that a more generalised and robust model can be obtained. These methods can be divided into two main groups: averaging and boosting. Regarding the first one, the different classifiers are built independently and only after that are their outputs combined to reduce the variance. Concerning the boosting methods, the classifiers are built sequentially so that the next classifier can try to decrease the bias of the combined. Voting classifier: this combines the predictions of the different classifiers and outputs a final prediction as the result of a majority vote. This majority vote can be hard or soft. Hard: each classifier predicts the class, and the final prediction is the one that most of them predicted. The final prediction can be obtained using a weighted averaging procedure if the classifiers have different weights. Soft: each classifier has a weight and predicts the probability of each class, and then the final prediction is obtained using a weighted averaging procedure. AdaBoost: considers several initial classifiers, called weak learners, and combines their predictions through a weighted majority vote. This process is repeated, and at each iteration/boost the data are modified. Each sample starts with a weight, and if it is incorrectly classified its weight increases for the classifier to notice it more; on the other hand, correctly classified samples have their weights reduced. After several iterations, the overall classifier, or strong learner, is expected to be better than the individual ones.

To further improve results, several non-linear classifiers were also tested. These include the K-nearest neighbours (K-NNs), kernel support vector machines (KSVMs), decision trees (DTs), neural networks (NNs), and, finally, random forest (RF). Similarly to linear classifiers, the same combinations of classifiers were tested as well. As these classifiers require more data to obviate overfitting, for each trial of MI, which had a duration of five seconds, more epochs were extracted. For each trial, two epochs of two seconds were extracted. The first second of the signal was not used, as a preventive way, since the image of the arrow could act as a stimulus for the pretended direction. Hereupon, this time is sufficient for the person to assimilate which MI must do. Nevertheless, the approach of doubling the number of epochs ended up also being used for the linear and statistical classifiers. This

was to ensure that both methods used the same amount of data. Table 5 summarises all the optimised hyper-parameters for the respective classifiers, as well as a brief description of their function. The optimisation process was carried out using a grid search with a 5-fold cross-validation decided by the F1-score.

**Table 5.** Optimised hyper-parameters for the different classifiers.

Supervised Classifier	Hyper-Parameter	Grid Search Space	Description
LR	C	logspace −4 to 6, step size 1	Regularisation parameter, which has a significant effect on the generalisation performance of the classifier
K-NN	n_neighbors	1 to 50, step size 10	Number of neighbours to use
SVM	Kernel	rbf—Gaussian kernel function	Function used to compute the kernel matrix for classification
	gamma	logspace −3 to 6, step size 1	Regularisation parameter used in RBF kernel, which has significant impact on the performance of the kernel
	C	logspace −3 to 7, step size 1	Regularisation parameter, which has a significant effect on the generalisation performance of the classifier
DT	max_depth	1 to 20, step size 2	The maximum depth of the tree. If none, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples.
	min_samples_split	10 to 500, step size 20	Minimum number of samples required to split a node
	min_samples_leaf	1 to 10, step size 2	Minimum number of samples required in a newly created leaf after the split
NN	hidden_layers	5 to 55, step 10	The <i>i</i> element represents the number of neurons in the <i>i</i> hidden layer
	activation	relu—rectified linear unit function	Activation function for the hidden layer
	solver	adam—stochastic gradient-based optimiser	The solver for weight optimisation
	learning_rate	constant	Learning rate schedule for weight updates. If 'constant', the learning rate is given by learning_rate_init
	learning_rate_init	logspace −4 to 4, step 1	The initial learning rate used. It controls the step size in updating the weights.
	alpha	logspace −4 to 4, step 1	L2 penalty (regularisation term) parameter.
RF	n_estimators	10 to 100, step 20	Number of trees in the forest
	max_depth	None or 2 to 10, step size 2	The maximum depth of the tree. If none, then nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples.
	min_samples_split	10 to 500, step size 20	The minimum number of samples required to split a node
	min_samples_leaf	1 to 10, step size 2	The minimum number of samples required in a newly created leaf after the split

LR—logistic regression; K-NN—K-nearest neighbour; SVM—support vector machine; DT—decision tree; NN—neural networks; RF—random forest.

### 3.2.3. Evaluation

To evaluate the results from the different approaches on the two datasets, the F1-score, the kappa score, and the false positive (FP) rate were used. The F1-score is the average of the precision and recall, and it reaches its best value at 1 (perfect precision and recall) and worst at 0. The kappa score expresses the level of agreement between two annotators. Although it is not usually used to compare a prediction with a ground truth, it was the only

metric provided by the IV BCI Competition. A kappa value between  $-1$  and  $0$  denotes a random classifier, while a value near  $1$  means a perfect one. Concerning the FP rate, a new metric was developed, since it is more important, for the final application, to penalise the FPs from classes  $0$  and  $1$ , than from class  $2$ . Nevertheless, a high rate of true positives is still desirable, independently of the class. Thus, the false positive rate was used for this evaluation. For an FP rate higher than  $1$ , the classifier produces more false positives than true positives; hence, a rate smaller than  $1$  is desirable. For each subject, the best run out of the three was obtained based on the F1-score. Then, for that run, the respective kappa score and FP rate are also presented. Furthermore, for all the approaches, the feature selector was the extra trees. Table 6 contains the used linear and statistical classifiers (0–3), and the non-linear classifiers (4–8).

**Table 6.** Labels of the classifiers.

Number	Name
0	Gaussian Naive Bayes (GNB)
1	Linear discriminant analysis (LDA)
2	Linear support vector machines (LSVMs)
3	Logistic regression (LR)
4	K-nearest neighbours (K-NNs)
5	Kernel support vector machines (KSVMs)
6	Decision trees (DTs)
7	Neural networks (NNs)
8	Random forest (RF)

The first step consisted of only applying the linear and statistical classifiers. Afterwards, with the intent of improving even more the performance of the approach that presented the best results, the non-linear classifiers, along with the first set of classifiers, were only applied to the correspondent approach. This is because these classifiers take longer to run and optimise.

### 3.3. Hardware and Software

Python version 3 was the programming language used for the experimental work related to signal processing and classification, along with Numpy, Pandas, Seaborn, and Scikit-Learn libraries. For real-time testing, an interface was required to deliver the raw data acquired by the headset to Python. Given that the headset is an Emotiv EPOC, the pyemotiv Python library was applied. This library interfaces with the Emotiv EPOC research SDK, provided by the distributor, enabling the output of raw EEG data for the experimental setup [17].

## 4. Results

This section presents the results obtained from the different approaches previously explained, as well as the outcomes of a real-time application.

### 4.1. Filter Bank Common Spatial Pattern I—FBCSP I

In this approach, only the linear and statistical classifiers are used to build the different classifier combinations, since the obtained results were not the best.

#### 4.1.1. FBCSP I Approach Using Dataset A

Table 7 presents the obtained F1-score for the different combinations, using only linear or statistical classifiers. The row “Best” corresponds to the best score for each subject. Most subjects presented a preferable run, regardless of the combinations, except for subject 1,

who chose, at least once, one of the three runs. The highest F1-score, on average, was obtained by the ensemble voting hard, which can be explained as it corresponds to the major vote between the two best classifiers; thus, by combining their predictions, it comes out as more accurate.

**Table 7.** F1-score, kappa, and FP rate for dataset A and FBCSP I, using the first set of classifiers.

		Average	1	2	3	4	5	6	7	8	9
F1-score	1 Classifier	0.729	0.763	0.560	0.906	0.708	0.615	0.601	0.903	0.774	0.733
	Prob. F1	0.729	0.797	0.663	0.883	0.714	0.603	0.575	0.896	0.662	0.768
	2 Classifiers	0.712	0.782	0.496	0.920	0.679	0.610	0.546	0.848	0.776	0.751
	Prob. F1	0.738	0.792	0.623	0.932	0.689	0.634	0.567	0.874	0.768	0.760
	Soft	0.730	0.843	0.597	0.950	0.692	0.563	0.592	0.848	0.749	0.735
	Hard	0.748	0.801	0.632	0.928	0.719	0.647	0.615	0.869	0.759	0.758
	Ada	0.682	0.694	0.588	0.856	0.616	0.551	0.509	0.894	0.676	0.750
	Best	0.765	0.843	0.663	0.950	0.719	0.647	0.615	0.903	0.776	0.768
Kappa	1 Classifier	0.587	0.632	0.333	0.854	0.556	0.410	0.396	0.847	0.660	0.597
	Prob. F1	0.573	0.694	0.493	0.819	0.556	0.299	0.354	0.840	0.472	0.632
	2 Classifiers	0.583	0.653	0.236	0.875	0.500	0.403	0.546	0.764	0.660	0.611
	Prob. F1	0.590	0.674	0.438	0.896	0.535	0.382	0.347	0.792	0.646	0.604
	Soft	0.586	0.764	0.389	0.924	0.521	0.347	0.354	0.771	0.618	0.590
	Hard	0.607	0.701	0.361	0.889	0.569	0.472	0.417	0.785	0.639	0.632
	Ada	0.535	0.542	0.493	0.785	0.424	0.326	0.264	0.840	0.514	0.625
	Best	0.656	0.764	0.493	0.924	0.569	0.472	0.546	0.847	0.660	0.632
Winner	0.570	0.680	0.420	0.750	0.480	0.400	0.270	0.770	0.750	0.610	
FP rate	1 Classifier	0.295	0.206	0.475	0.092	0.355	0.565	0.527	0.031	0.228	0.177
	Prob. F1	0.271	0.134	0.294	0.111	0.178	0.878	0.423	0.123	0.200	0.098
	2 Classifiers	0.264	0.090	0.613	0.071	0.250	0.415	0.607	0.066	0.144	0.119
	Prob. F1	0.195	0.095	0.415	0.045	0.282	0.165	0.516	0.027	0.145	0.063
	Soft	0.334	0.126	0.578	0.044	0.361	0.582	0.699	0.098	0.280	0.242
	Hard	0.271	0.145	0.734	0.060	0.188	0.414	0.500	0.038	0.189	0.172
	Ada	0.386	0.307	0.685	0.157	0.474	0.571	0.755	0.109	0.260	0.154
	Best	0.159	0.090	0.294	0.044	0.178	0.165	0.423	0.027	0.144	0.063

Table 7 presents the respective kappa score, corroborating with the ensemble voting hard being the best combination. The kappa score from the winner of the IV BCI competition is also presented. However, the competition involved the classification of four classes: left hand, right hand, foot, and tongue; as for this work, there are only three classes: left, right, and neutral. Thereafter, the results from the competition are here exposed just as a qualitative comparison. Hereupon, the obtained kappa value of 0.604 is in the same order of magnitude as the result from the competition, 0.57, and thus higher than 0.5, which surely reflects the no randomness of the classifiers. Moreover, the FP rate had its lowest value, on average, for the one probabilistic classifier, whose threshold was decided based on the maximisation of the F1-score. This result is logical, as, by maximising the F1-score, there is an implicit maximisation of the precision and the recall, thus minimising the FP rate. However, the lowest FP rate was expected to belong to the ensemble voting hard because it was the combination with the highest F1-score. These different approaches correspond to different combinations of several classifiers. Table 8 presents the best ones for the different approaches and subjects. It can be concluded that the best algorithms correspond to the Gaussian Naive Bayesian classifier (0), linear discriminant analysis (1) and logistic regression (3), which was not presumed, as the LR was seldom mentioned during the literature review. Concerning the linear SVM (2), it was never picked, suggesting that it is not a good classifier for this dataset, using these features, as it is not capable of accurately distinguishing the three classes.

**Table 8.** Best classifiers, from the first set, for each combination for dataset A and FBCSP I.

	1	2	3	4	5	6	7	8	9							
1 Classifier	0	0	3	3	3	3	3	3	3							
Prob. F1	0	3	3	3	3	0	0	1	3							
2 Classifiers	3	0	3	1	3	1	1	3	0	3	1					
Prob. F1	0	1	3	1	0	3	3	1	0	3	1	0				
Soft	0	3	3	1	3	1	3	0	3	1	0	3				
Hard	3	0	0	1	0	3	3	0	3	0	3	0	1	3	3	1

0—Gaussian Naïve Bayesian; 1—linear discriminant analysis; 3—logistic regression.

#### 4.1.2. FBCSP I Approach Using Dataset B

Similarly to what was presented for dataset A, Table 9 introduces the F1-score for the best run in each approach. Contrary to A, several subjects picked all three runs at least once as their best. Only subjects 5, 6, and 7 picked one or two. This already suggests that the extracted features were not very strongly indicative of the class. Once again, the best F1-score was obtained by the ensemble voting hard approach, followed by the two classifiers. However, since the F1-score varies from 0 to 1, the obtained result is not satisfactory as it stays in the bottom half of the spectrum. Similarly, for the kappa score, the value of 0.218 is closer to 0 than to 1, indicating that the classifier is closer to random than to perfect, as presumed. The FP rate is quite high, reaching almost 1, that is to say, the number of FP is almost the same as TP, thus manifesting that this approach is not adequate for the ultimate goal of controlling an IW. A very low FP rate is mandatory to maintain the safety of the IW driver. Nevertheless, subjects 3 and 1 presented a better performance than the others did, presenting scores equivalent to dataset A, which corroborates that people have different aptness regarding MI [40]. Nevertheless, it is also important to consider inter-individual differences, such as distinct brainwave patterns, cognitive abilities, and learning speeds, among others. This major difference between subjects may also be due to the positioning of the headset, as the electrodes are fixed, which may lead to more coverage of the motor cortex in some subjects than in others.

Table 10 contains the chosen classifiers for the different methods. Similarly to A, LR (3) and LDA (1) presented the best performance. However, the Gaussian Naive Bayes (0) did not perform well enough to be chosen. Once again, the linear SVM (2) was not picked.

**Table 9.** F1-score, kappa, and FP rate for dataset B and FBCSP I, using the first set of classifiers.

		Average	1	2	3	4	5	6	7	8	9
F1-score	1 Classifier	0.444	0.507	0.386	0.636	0.400	0.375	0.405	0.456	0.390	0.439
	Prob. F1	0.438	0.470	0.416	0.662	0.316	0.492	0.334	0.437	0.361	0.452
	2 Classifiers	0.448	0.442	0.456	0.622	0.391	0.419	0.361	0.442	0.404	0.496
	Prob. F1	0.427	0.427	0.353	0.640	0.444	0.403	0.333	0.442	0.303	0.496
	Soft	0.435	0.485	0.358	0.592	0.430	0.455	0.401	0.406	0.330	0.453
	Hard	0.484	0.597	0.401	0.626	0.410	0.547	0.386	0.470	0.483	0.440
	Ada	0.428	0.417	0.375	0.625	0.392	0.433	0.383	0.358	0.375	0.492
	Best	0.507	0.597	0.456	0.662	0.444	0.547	0.405	0.470	0.483	0.496
Kappa	1 Classifier	0.160	0.262	0.075	0.450	0.075	0.063	0.100	0.175	0.075	0.162
	Prob. F1	0.150	0.175	0.000	0.488	0.075	0.250	0.012	0.125	0.050	0.175
	2 Classifiers	0.133	0.162	0.188	0.438	0.037	0.125	0.050	0.000	0.137	0.063
	Prob. F1	0.121	0.137	−0.012	0.438	0.037	0.150	0.000	0.137	−0.038	0.238
	Soft	0.146	0.225	0.037	0.387	0.137	0.175	0.113	0.075	−0.012	0.175
	Hard	0.218	0.387	0.100	0.438	0.088	0.325	0.063	0.200	0.200	0.162
	Ada	0.142	0.125	0.063	0.438	0.088	0.150	0.075	0.037	0.063	0.238
	Best	0.253	0.387	0.188	0.488	0.137	0.325	0.113	0.200	0.200	0.238



Table 9. Cont.

		Average	1	2	3	4	5	6	7	8	9
FP rate	1 Classifier	0.908	0.672	1.239	0.355	1.370	0.867	0.646	0.852	1.435	0.736
	Prob. F1	1.202	1.056	2.000	0.380	1.587	0.767	1.610	1.240	1.273	0.907
	2 Classifiers	0.796	0.811	0.727	0.373	0.744	0.680	0.955	1.250	1.000	0.622
	Prob. F1	0.888	0.804	1.949	0.240	0.209	0.596	0.975	1.176	1.486	0.559
	Soft	0.918	0.793	1.116	0.451	1.059	0.611	0.857	1.043	1.590	0.741
	Hard	0.730	0.521	1.125	0.440	0.511	0.439	1.156	0.839	0.821	0.717
	Ada	0.916	0.720	1.222	0.373	1.426	0.692	1.065	1.302	1.000	0.441
	Best	0.543	0.521	0.727	0.240	0.209	0.439	0.646	0.839	0.821	0.441

Table 10. Best classifiers, from the first set, for each combination for dataset B and FBCSP I.

Classifiers	1	2	3	4	5	6	7	8	9
1 Classifier	3	1	3	3	3	3	1	3	3
Prob. F1	1	3	3	3	3	3	3	3	3
2 Classifiers	1	3	3	1	3	1	3	1	3
Prob. F1	3	1	3	1	3	1	3	1	3
Soft	3	1	3	1	3	1	3	1	3
Hard	3	1	3	3	1	3	1	3	3

0—Gaussian Naïve Bayesian; 1—linear discriminant analysis; 3—logistic regression.

#### 4.2. Filter Bank Common Spatial Pattern II—FBCSP II

Next, we describe the implementation of the FBCSP II approach using dataset A, providing detailed analysis and results of the classification performance. Following that, we examine the application of the same approach using dataset B, shedding light on its efficacy and comparative performance metrics.

##### 4.2.1. FBCSP II Approach Using Dataset A

Table 11 presents the F1-score for dataset A, regarding the use of only linear or statistical classifiers. Most of the subjects present a preference concerning the run; for others, such as 3 and 5, it is not clear, as the three runs were chosen as the best one, at least once. Globally, the different classifier combinations presented resulted more or less in the same ranking and behaved as expected. The best F1-score was obtained, once again, for the ensemble voting hard approach, followed by the one classifier approach, which was not prospected, as it is intuitive that the output of two classifiers would be more accurate than just one. The worst score corresponds to the AdaBoost, thus clearly excluding it as a recommended approach.

The best kappa was from the ensemble voting hard approach, 0.693, which is higher than the one from the previous approach, 0.607, and in the same order of magnitude as 0.57, the kappa score of the competition's winner. It was already prospected that both the two classifier approaches would present a lower FP rate because they prevent the FP for classes 0 and 1. This is also one of the reasons why their F1-score is slightly lower than for the other approaches. AdaBoost presented the lower F1-score and kappa, and therefore presented the highest FP rate. It can be concluded (Table 12) that the best algorithms correspond once again to linear discriminant analysis (1) and logistic regression (3). The Gaussian Naïve Bayesian classifier (0) was chosen a few times; however, it was almost always the second-best classifier, whereas the linear SVM (2) was never picked. Therefore, it is possible to conclude that the latter is not a good classifier for this dataset using these features, as it is not capable of accurately distinguishing the three classes.

**Table 11.** F1-score, kappa, and FP rate for dataset A and FBCSP II, using the first set of classifiers.

		Average	1	2	3	4	5	6	7	8	9
F1-score	1 Classifier	0.793	0.856	0.697	0.950	0.838	0.606	0.668	0.898	0.782	0.846
	Prob. F1	0.792	0.854	0.643	0.928	0.849	0.629	0.663	0.923	0.788	0.855
	2 Classifiers	0.785	0.840	0.643	0.933	0.840	0.604	0.686	0.859	0.842	0.817
	Prob. F1	0.783	0.840	0.643	0.930	0.840	0.644	0.686	0.859	0.793	0.817
	Soft	0.787	0.878	0.682	0.921	0.825	0.662	0.608	0.888	0.805	0.818
	Hard	0.797	0.829	0.698	0.949	0.833	0.618	0.679	0.924	0.774	0.871
	Ada	0.724	0.801	0.620	0.894	0.769	0.616	0.542	0.819	0.745	0.713
	Best	0.818	0.878	0.698	0.950	0.849	0.662	0.686	0.924	0.842	0.871
Kappa	1 Classifier	0.683	0.785	0.535	0.924	0.757	0.382	0.500	0.833	0.667	0.764
	Prob. F1	0.685	0.764	0.465	0.928	0.764	0.431	0.479	0.882	0.674	0.778
	2 Classifiers	0.596	0.681	0.361	0.896	0.611	0.292	0.382	0.639	0.764	0.743
	Prob. F1	0.658	0.757	0.458	0.889	0.750	0.396	0.521	0.757	0.681	0.715
	Soft	0.623	0.757	0.472	0.875	0.736	0.493	0.306	0.826	0.701	0.438
	Hard	0.693	0.743	0.542	0.924	0.750	0.424	0.514	0.882	0.653	0.806
	Ada	0.586	0.701	0.431	0.840	0.653	0.424	0.313	0.729	0.618	0.569
	Best	0.720	0.785	0.542	0.928	0.764	0.493	0.521	0.882	0.764	0.806
Winner	0.570	0.680	0.420	0.750	0.480	0.400	0.270	0.770	0.750	0.610	
FP rate	1 Classifier	0.218	0.108	0.383	0.029	0.127	0.646	0.347	0.016	0.238	0.071
	Prob. F1	0.161	0.038	0.317	0.040	0.066	0.403	0.255	0.025	0.237	0.065
	2 Classifiers	0.192	0.153	0.532	0.035	0.131	0.395	0.276	0.024	0.132	0.050
	Prob. F1	0.150	0.083	0.370	0.080	0.067	0.186	0.279	0.011	0.194	0.080
	Soft	0.238	0.127	0.529	0.081	0.152	0.336	0.581	0.042	0.197	0.097
	Hard	0.195	0.145	0.340	0.044	0.156	0.429	0.295	0.020	0.247	0.080
	Ada	0.260	0.116	0.463	0.109	0.229	0.474	0.624	0.068	0.118	0.136
	Best	0.121	0.038	0.317	0.029	0.066	0.186	0.255	0.011	0.132	0.050

**Table 12.** Best classifiers, from the first set, for each combination for dataset A and FBCSP II.

Classifiers	1	2	3	4	5	6	7	8	9	
1 Classifier	3	3	3	3	3	3	3	3	3	
Prob. F1	3	3	1	3	3	3	3	3	3	
2 Classifiers	3	1	3	1	3	1	3	1	3	1
Prob. F1	3	0	3	1	1	3	3	1	3	1
Soft	0	3	3	0	3	1	3	0	3	0
Hard	3	0	3	1	3	1	3	0	3	1

0—Gaussian Naïve Bayesian; 1—linear discriminant analysis; 3—logistic regression.

Due to their promising results, the use of non-linear classifiers was also tested. However, the results, contrary to what was expected, did not improve; instead, they stayed roughly the same. Furthermore, the best combination was not the ensemble voting hard but the two classifiers, which reflected in a lower FP rate. Regarding the kappa score, its average value is very close to the one previously obtained for the first group of classifiers.

#### 4.2.2. FBCSP II Approach Using Dataset B

Similarly to what was presented for dataset A, Table 13 introduces the F1-score for the best run in each approach. Most of the subjects presented a preference regarding a run or two. The best classifier was the AdaBoost, with an average F1-score of 0.504, immediately followed by the ensemble voting soft, with a score of 0.497. Moreover, the highest F1-score value was lower than for A but higher than for the previous approach, 0.484. Once again, it is important to emphasise that subject 1 and subject 3 presented scores equivalent to dataset A.

**Table 13.** F1-score, kappa, and FP rate for dataset B and FBCSP II, using the first set of classifiers.

		Average	1	2	3	4	5	6	7	8	9
F1-score	1 Classifier	0.461	0.612	0.451	0.712	0.438	0.515	0.479	0.472	0.000	0.471
	Prob. F1	0.478	0.556	0.431	0.710	0.394	0.506	0.467	0.456	0.316	0.463
	2 Classifiers	0.454	0.606	0.435	0.697	0.431	0.527	0.471	0.468	0.000	0.455
	Prob. F1	0.457	0.605	0.433	0.700	0.437	0.506	0.490	0.456	0.000	0.489
	Soft	0.497	0.624	0.425	0.716	0.414	0.526	0.481	0.481	0.337	0.471
	Hard	0.486	0.612	0.433	0.705	0.408	0.515	0.422	0.472	0.332	0.471
	Ada	0.504	0.517	0.417	0.742	0.475	0.500	0.508	0.467	0.342	0.567
	Best	0.524	0.624	0.451	0.742	0.475	0.527	0.508	0.481	0.342	0.567
Kappa	1 Classifier	0.243	0.400	0.175	0.550	0.137	0.275	0.225	0.200	0.000	0.225
	Prob. F1	0.200	0.262	0.100	0.513	0.088	0.225	0.200	0.188	0.000	0.225
	2 Classifiers	0.232	0.387	0.150	0.513	0.125	0.288	0.213	0.200	0.012	0.200
	Prob. F1	0.233	0.375	0.150	0.513	0.137	0.225	0.238	0.188	0.025	0.250
	Soft	0.244	0.425	0.137	0.563	0.113	0.288	0.225	0.213	0.013	0.225
	Hard	0.224	0.400	0.150	0.538	0.100	0.275	0.125	0.200	0.000	0.225
	Ada	0.256	0.275	0.125	0.613	0.213	0.250	0.262	0.200	0.012	0.350
	Best	0.285	0.425	0.175	0.613	0.213	0.288	0.262	0.213	0.025	0.350
FP rate	1 Classifier	0.515	0.306	0.741	0.167	0.51	0.597	0.69	0.804	0.05	0.776
	Prob. F1	0.459	0.361	0.563	0.136	0.511	0.241	0.804	0.727	0.15	0.638
	2 Classifiers	0.49	0.296	0.731	0.148	0.5	0.524	0.667	0.714	0.024	0.804
	Prob. F1	0.444	0.373	0.563	0.173	0.549	0.241	0.695	0.618	0.071	0.717
	Soft	0.544	0.311	0.882	0.188	0.592	0.556	0.724	0.772	0.098	0.776
	Hard	0.561	0.306	0.827	0.169	0.646	0.597	0.8	0.804	0.125	0.776
	Ada	0.513	0.468	0.68	0.124	0.404	0.55	0.77	0.804	0.293	0.529
	Best	0.385	0.296	0.563	0.124	0.404	0.241	0.667	0.618	0.024	0.529

Furthermore, the best value for the average kappa score, 0.256, is both lower than the one for dataset A and also lower than the value of the competition; however, it is higher than the value of 0.218 from FBCSP I. Nevertheless, is still not a desirable value as it is too close to 0. As presumed from the previous scores, the FP rate is higher for dataset B than for dataset A. The lowest value, 0.444, was obtained with the one probabilistic classifier approach. The FP rate for the best approach, AdaBoost, was 0.513, which is high but still lower than the best for approach FBCSP I, 0.730. Concerning the selected classifiers, it is clear, from Table 14, that LDA and LR were preferred over the other two. Moreover, the Gaussian Naive Bayes classifier was not chosen as the best one for either of the combinations. Despite what was concluded from dataset A, the linear SVM was picked for this one, even if only once, which endorses that LDA and LR are the most suitable classifiers for this approach.

**Table 14.** Best classifiers, from the first set, for each combination for dataset B and FBCP II.

	S1	S2	S3	S4	S5	S6	S7	S8	S9		
1 Classifier	3	1	2	3	1	3	1	1	3		
Prob. F1	3	1	3	3	3	3	1	1	3		
2 Classifiers	3	1	1	3	3	1	3	1	3	3	1
Prob. F1	3	1	1	3	3	1	3	1	3	3	1
Soft	3	1	1	3	3	1	3	1	3	3	1
Hard	3	1	1	3	3	1	3	1	3	3	1

1—Linear discriminant analysis; 2—linear support vector machines; 3—logistic regression.

Because the formerly exposed results were not satisfactory, in the sense that they are not acceptable for a real application, further approaches demanded to be tested. Moreover, as FBCSP II was better than FBCSP I, non-linear classifiers were added to the previous classifiers to be trained and tried out. Table 15 contains the F1-score from this method. As expected, the results improved compared with the linear and statistical classifiers approach.

Two combinations obtained the same average F1-score, the one classifier, and the ensemble voting hard, which was not foreseen as it was presumed that two classifiers would predict a more accurate result rather than just one.

**Table 15.** F1-score, kappa, and FP rate for dataset B and FBCSP II, using both sets of classifiers.

		Average	1	2	3	4	5	6	7	8	9
F1-score	1 Classifier	0.651	0.747	0.595	0.911	0.604	0.660	0.706	0.604	0.517	0.515
	Prob. F1	0.648	0.718	0.631	0.912	0.628	0.667	0.737	0.642	0.450	0.447
	2 Classifiers	0.588	0.712	0.586	0.773	0.582	0.601	0.613	0.545	0.460	0.418
	Prob. F1	0.599	0.712	0.560	0.771	0.619	0.650	0.658	0.623	0.374	0.418
	Soft	0.646	0.758	0.626	0.892	0.589	0.644	0.678	0.622	0.485	0.519
	Hard	0.651	0.747	0.595	0.911	0.604	0.660	0.706	0.604	0.517	0.515
	Ada	0.668	0.758	0.631	0.912	0.628	0.667	0.737	0.642	0.517	0.519
	Best	0.651	0.747	0.595	0.911	0.604	0.660	0.706	0.604	0.517	0.515
Kappa	1 Classifier	0.426	0.463	0.375	0.863	0.387	0.488	0.525	0.387	0.125	0.225
	Prob. F1	0.433	0.625	0.175	0.863	0.425	0.488	0.563	0.463	0.100	0.200
	2 Classifiers	0.359	0.550	0.313	0.625	0.400	0.390	0.413	0.325	0.050	0.162
	Prob. F1	0.413	0.575	0.325	0.638	0.413	0.475	0.488	0.438	0.175	0.188
	Soft	0.447	0.625	0.450	0.838	0.375	0.463	0.450	0.425	0.150	0.250
	Hard	0.443	0.600	0.375	0.863	0.387	0.488	0.538	0.387	0.125	0.225
	Ada	0.478	0.625	0.450	0.863	0.425	0.488	0.563	0.463	0.175	0.250
	Best	0.426	0.463	0.375	0.863	0.387	0.488	0.525	0.387	0.125	0.225
FP rate	1 Classifier	0.471	0.519	0.314	0.101	0.690	0.468	0.422	0.648	0.080	1.000
	Prob. F1	0.411	0.322	0.111	0.101	0.595	0.506	0.400	0.506	0.125	1.036
	2 Classifiers	0.314	0.190	0.200	0.078	0.486	0.178	0.288	0.561	0.068	0.774
	Prob. F1	0.399	0.370	0.364	0.110	0.534	0.397	0.466	0.440	0.071	0.836
	Soft	0.471	0.300	0.461	0.121	0.700	0.506	0.500	0.581	0.135	0.933
	Hard	0.471	0.519	0.314	0.101	0.690	0.468	0.422	0.648	0.080	1.000
	Ada	0.290	0.190	0.111	0.078	0.486	0.178	0.288	0.440	0.068	0.774
	Best	0.471	0.519	0.314	0.101	0.690	0.468	0.422	0.648	0.080	1.000

Regarding the kappa score, it was higher for the ensemble voting hard, which is better than the score for just the linear and statistical classifiers, but still lower than for dataset A, which corroborates with that previously stated about the headset used to acquire these signals. The FP rate was lower than previously, as was anticipated due to the rise of the F1-score. Despite these results, it is still important to mention that subjects 1, 3, and 6 produced results comparable to the ones from dataset A, even if with slightly higher FP rates than the ones from dataset A. Table 16 presents the chosen classifier(s) for each combination. These encompass mainly the kernel SVMs, followed by the neural networks. In some cases, the K-NN is chosen as the second-best classifier. None of the linear or statistical classifiers were chosen, which indicates that this dataset requires more complex models to predict better results.

**Table 16.** Best classifiers, from both sets, for each combination for dataset B and FBCSP II.

Individual (I)	I1	I2	I3	I4	I5	I6	I7	I8	I9
1 Classifier	5	5	5	5	5	5	5	5	5
Prob. F1	5	5	5	5	5	5	5	5	5
2 Classifiers	5 7	5 7	5 4	5 7	5 7	5 7	5 7	5 7	5 7
Prob. F1	5 7	5 4	5 4	5 7	5 7	5 7	5 7	5 4	5 7
Soft	5 7	5 4	5 4	5 8	5 7	5 7	5 7	5 4	5 4
Hard	5 7	5 7	5 4	5 7	5 7	5 7	5 7	5 4	5 7

4—K-nearest neighbours; 5—kernel support vector machines; 7—neural networks; 8—random forest.

### 4.3. Power Spectral Density

The next subsections present the findings and insights from the analysis of results obtained using power spectral density.

#### 4.3.1. Power Spectral Density Approach Using Dataset A

As mentioned before, since the best results with linear classifiers were not obtained with the PSD approach, non-linear strategies were not employed. Moreover, as no single linear classifier produced satisfactory results on its own, combinations thereof were not evaluated either. The obtained F1-score was 0.510, which is much lower than the F1-score of 0.793 for the combination of one classifier and FBCSP II. Moreover, the FP rate is also much higher than the one for FBCSP II, thus endorsing the idea that this approach is not the best for this dataset.

#### 4.3.2. Power Spectral Density Approach Using Dataset B

Similarly to dataset A, only the one classifier approach was tested due to unsatisfactory results. However, the difference between this and the FBCSP II was not as large as the one for dataset A. The average F1-score, 0.443, was quite similar to the one for dataset A, 0.510, which did not happen for the other approaches. However, the kappa score and the FP rate were worse. The latter presents a high value of 0.844, which lies close to 1, that is to say, there is almost more FP than TP, which is not the goal.

### 4.4. Real-Time Application

The subject with the best performance for dataset B was chosen to perform the real-time testing. The subject was asked to sit still and maintain movements to a minimum, similar to the training phase. An external person experimented and asked the subject to imagine a certain MI. Every 2 s an epoch was sent to the system and a class was predicted. The person conducting the experiment waited for ten predictions to appear before asking for the next one, as a way of allowing the system to stabilise. Again, due to stabilisation, the first three outputs after a new MI were discarded. As subject 3 was the one who presented the best results, the steps described previously were applied in a real-time scenario, leading to the results in Table 17. The final system consisted of applying the FBCSP II approach, which produced the best score. Then, 70% of the features were extracted and fed to an ensemble voting hard classifier built with the major vote between the kernel SVM and the K-NN, where the vote percentage was 2 to 1, respectively.

**Table 17.** Cues and respective outputs from subject 3's real-time applications.

MI	Output										Majority	%	MI	Output										Majority	%
N	2	1	1	2	2	2	0	0	2	2	2	60%	R	1	2	1	1	1	1	1	1	1	1	1	90%
R	1	1	0	1	2	1	1	0	0	1	1	60%	N	2	2	2	2	1	2	0	0	2	2	2	70%
N	2	0	2	2	2	2	2	0	0	2	2	70%	L	2	1	1	1	1	1	0	0	1	0	1	30%
L	0	0	0	1	0	0	1	1	2	2	0	50%	N	2	2	2	0	2	2	2	2	2	2	2	90%
L	0	1	0	0	0	1	0	1	2	2	0	50%	R	1	0	1	1	1	0	1	1	1	2	1	70%
N	2	0	2	2	0	2	2	2	0	2	2	70%	N	2	2	2	2	2	2	2	2	2	2	2	100%
R	1	0	1	2	1	1	2	1	1	0	1	60%	L	0	0	2	0	2	0	2	0	2	2	0	50%
L	0	0	2	0	0	0	1	1	0	1	0	60%	N	2	1	2	2	2	1	2	2	2	2	2	80%
N	2	2	2	1	2	2	1	2	0	2	2	70%	R	1	2	1	2	1	2	0	2	0	2	2	30%
R	1	1	1	1	0	0	2	1	1	1	1	70%	N	2	2	2	2	2	2	2	2	2	2	2	100%
N	0	2	2	0	2	2	0	0	2	2	2	60%	L	0	0	0	1	0	0	0	0	2	2	0	70%

Overall, the results were satisfactory and as anticipated from the performance previously analysed. From twenty-two cues, there were only two cues incorrectly classified, represented in red, one for the left and another for the right. The left one was misclassified

as right, which is a problem since, if it was an IW, it would go in the opposite direction. The right cue was classified as neutral. Despite being misclassified, it is not the worst since a hypothetical IW would have maintained the same direction. Nevertheless, none of the neutral cues were incorrectly classified. Although some limitations related to the accuracy, these results show that the system is evolving towards the right direction, suggesting that a new headset and a refinement of the algorithm would deliver promising results.

## 5. Discussion

Overall, dataset A, independently of the approach, produced better results than dataset B. The only exception might be for the PSD approaches, where the average F1-score value for both datasets was very similar. There are several possibilities for this discrepancy in the results, such as dataset A was obtained using a stable headset, with movable electrodes. This leads to the electrodes always assuming the same known position throughout the different sessions. Moreover, the results were obtained by professionals in a carefully controlled environment. Dataset B was collected using an Emotiv EPOC. There are several problems regarding the EPOC, such as it has fixed electrodes. This leads to the absolute position being different for all the subjects, and even varies from session to session. Moreover, it did not fit everyone's head, and four subjects did not make it into this dataset due to limitations in positioning the electrodes correctly. Some sensors were starting to be oxidised, which led to noisy acquisitions, which hampered the already difficult EEG processing, as the signal is very sensitive. The Emotiv EPOC does not cover the motor cortex, which is critical for the tasks in this study. While the literature suggests it can work for parietal and frontal areas, its performance is not optimal for motor cortex tasks. Nevertheless, the Emotiv EPOC was chosen for its balance of cost, ease of use, and functionality, offering a reasonable number of electrodes, wireless operation, extended battery life, and affordability. Due to the fact that the process of training is very time-consuming and this work was merely a proof of concept, only the subject with the best performance for dataset B was chosen for the real-time testing.

Comparing the different approaches, it was already expected that the best method would be related to the CSP, as it was the winning method of the competition. This suggests that spatial methods perform better than the others do, which may be related to the elimination of existent artefacts in the bands of interest. However, it was interesting that FBCSP II produced slightly better scores than FBCSP I, implying that feeding the whole spatial filtered signal to the feature selector works better than feeding a transformed version of the signal filtered by just the columns of the spatial filter. Although the results from the competition are merely qualitative, the results from FBCSP II also indicate that using the extra tree classifier to obtain the features' importance and the ensemble voting hard, employing the LR and the GNB, or LDA, to classify the epochs, represents a valuable update. This led to a final average kappa score of 0.69, which is 20% higher than the winner value of 0.57. Despite the 20% not being a real quantitative evaluation, the value of 0.69 already suggests that the algorithm is considerably better than a random classifier and can correctly classify the epochs, presenting an F1-score of 0.797 and the smallest FP rate of all the tested approaches, 0.150. Similarly, dataset B also presented better results for the FBCSP II than for the FBCSP I. Moreover, contrarily to dataset A, dataset B improved its results by allowing the use of non-linear classifiers. Fakhruzzaman et al. [66] and Muñoz et al. [61] used the Emotiv EPOC headset and the CSP method as a features extractor. Fakhruzzaman et al. [66] obtained an average accuracy of 60%, whereas Muñoz et al. [61] obtained an average accuracy of 67.5% using the LDA classifier, 68.3% using the SVM, and 96.7% using Nu-SVC RBF kernel. Overall, the result of 65% from dataset B regarding the FBCSP II approach with the ensemble voting hard classifier falls within the mean values presented by these authors, except for the last method of Muñoz et al. [61]. The latter is greatly higher than the others are, suggesting that this classifier is indicated for this type of feature, and should be considered for further implementation in future work. Furthermore, it is important to state that the signals in dataset B had constraints in AF3 and AF4 electrodes,

which may be important electrodes according to Lin and Lo [60] and Muñoz et al. [61], thus decreasing the obtained accuracy and the F1-score. The average accuracy of other authors using the EPOC and the magnitude of frequency components or the power spectral density (square of the magnitude) as features was 74–100% for Abiyev et al. [57], 70% for Hurtado-Rincon et al. [59], and 86–92% for Lin and Lo [60], and Siribunyaphat and Punsawad [67]. More recent works [49,68,69] have also achieved important F1-scores, using different EEG headsets. This reflects a promising area to explore for controlling intelligent wheelchairs.

## 6. Conclusions and Future Work

The goal of this work was being able to decode MI intentions from the users, using an Emotiv EPOC as the headset to extract the EEG signals. The intentions were left, right, and neutral, which would be further translated into control commands for an intelligent wheelchair. This headset has higher constraints in terms of accessing data in a less controlled environment; however, overall, this work allowed the development of a proof of concept for future projects and a thorough study regarding the different algorithms. Although the real-time results are still not suitable for the actual application, they validate the concept and the developed architecture to connect the different parts of the system. For future work, utilising a broader and more diverse dataset may contribute to enhancing the model's generality. Another interesting future work could be applying different methods for noise removal, such as independent component analysis. Additionally, conducting long-term usage and testing across diverse environments will be essential for assessing system stability and applicability.

**Author Contributions:** Conceptualization, L.P.R. and B.M.F.; methodology, M.C.A., P.A., L.P.R. and B.M.F.; software, M.C.A. and P.A.; validation, M.C.A. and P.A.; formal analysis, M.C.A., P.A. and L.P.R.; investigation, M.C.A., P.A. and L.P.R.; resources, L.P.R. and B.M.F.; data curation, M.C.A. and P.A.; writing—original draft preparation, M.C.A. and P.A.; writing—review and editing, L.P.R. and B.M.F.; visualization, M.C.A., P.A., L.P.R. and B.M.F.; supervision, L.P.R. and B.M.F. project administration, L.P.R. and B.M.F.; funding acquisition, L.P.R. and B.M.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was financially supported by: Base Funding—UIDB/00027/2020 of the Artificial Intelligence and Computer Science Laboratory (LIACC) funded by national funds through the FCT/MCTES (PIDDAC) and IntellWheels2.0: Intelligent Wheelchair with Flexible Multimodal Interface and Realistic Simulator (POCI-01-0247-FEDER-39898), supported by NORTE 2020, under PT2020.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Optimizer/LITEC (under project POCI-01-0247-FEDER-39898 approved on 28 May 2019).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data from the BCI competitions are available at <https://www.bbc.de/competition/iv/#datasets> first accessed on 1 June 2019. Other data are available on request from the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Davies, A.; Souza, L.H.; Frank, A.O. Changes in the quality of life in severely disabled people following provision of powered indoor/outdoor chairs. *Disabil. Rehabil.* **2003**, *25*, 286–290. [CrossRef] [PubMed]
2. Koontz, A.M.; Ding, D.; Jan, Y.; Groot, S.; Hansen, A. Wheeled mobility. *BioMed Res. Int.* **2015**, *2015*, 138176. [CrossRef] [PubMed]
3. Guidelines on the Provision of Manual Wheelchairs in Less Resourced Settings. Available online: <https://www.who.int/publications/i/item/9789241547482> (accessed on 5 January 2024).
4. Worldwide Need—Wheelchair Foundation. Available online: <https://www.wheelchairfoundation.org/fth/analysis-of-wheelchair-need> (accessed on 5 January 2024).

5. Xin-an, F.; Luzheng, B.; Teng, T.; Ding, H.; Liu, Y. A brain–computer interface-based vehicle destination selection system using p300 and ssvp signals. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 274–283. [[CrossRef](#)]
6. Luzheng, B.; Xin-An, F.; Liu, Y. EEG-based brain-controlled mobile robots: A survey. *IEEE Trans. Hum.-Mach. Syst.* **2013**, *43*, 161–176. [[CrossRef](#)]
7. Gürkök, H.; Nijholt, A. Brain–computer interfaces for multimodal interaction: A survey and principles. *Int. J. Hum.-Comput. Interact.* **2012**, *28*, 292–307. [[CrossRef](#)]
8. Major, T.C.; Conrad, J.M. A survey of brain computer interfaces and their applications. In Proceedings of the SOUTHEASTCON 2014, Lexington, KY, USA, 13–16 March 2014; pp. 1–8. [[CrossRef](#)]
9. Hämäläinen, M.; Hari, R.; Ilmoniemi, R.J.; Knuutila, J.; Lounasmaa, O.V. Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* **1993**, *65*, 413. [[CrossRef](#)]
10. Ferrari, M.; Quaresima, V. A brief review on the history of human functional near-infrared spectroscopy (fnirs) development and fields of application. *Neuroimage* **2012**, *63*, 921–935. [[CrossRef](#)] [[PubMed](#)]
11. Logothetis, N.K.; Pauls, J.; Augath, M.; Trinath, T.; Oeltermann, A. Neurophysiological investigation of the basis of the fmri signal. *Nature* **2001**, *412*, 150. [[CrossRef](#)] [[PubMed](#)]
12. Hashiguchi, K.; Morioka, T.; Yoshida, F.; Miyagi, Y.; Nagata, S.; Sakata, A.; Sasaki, T. Correlation between scalp-recorded electroencephalographic and electrocorticographic activities during ictal period. *Seizure* **2007**, *16*, 238–247. [[CrossRef](#)]
13. Buitengeweg, J.R.; Rutten, W.L.C.; Marani, E. Geometry-based finiteelement modeling of the electrical contact between a cultured neuron and a microelectrode. *IEEE Trans. Biomed. Eng.* **2003**, *50*, 501–509. [[CrossRef](#)]
14. Trambaiolli, L.R.; Falk, T.H. Hybrid brain–computer interfaces for wheelchair control: A review of existing solutions, their advantages and open challenges. In *Smart Wheelchairs and Brain-Computer Interfaces*; Elsevier: Amsterdam, The Netherlands, 2018; pp. 229–256. [[CrossRef](#)]
15. Hage, B.; Alwatban, M.R.; Barney, E.; Mills, M.; Dodd, M.D.; Truemper, E.J.; Bashford, G.R. Functional transcranial doppler ultrasound for measurement of hemispheric lateralization during visual memory and visual search cognitive tasks. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control.* **2016**, *63*, 2001–2007. [[CrossRef](#)] [[PubMed](#)]
16. Open Bci. Available online: <https://openbci.com/> (accessed on 5 January 2024).
17. Emotiv Epoc. Available online: <https://www.emotiv.com/> (accessed on 10 January 2024).
18. Pires, G.; Castelo-Branco, M.; Nunes, U. Visual p300-based bci to steer a wheelchair: A bayesian approach. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada, 20–25 August 2008; pp. 658–661. [[CrossRef](#)]
19. Zhang, R.; Wang, Q.; Li, K.; He, S.; Qin, S.; Feng, Z.; Chen, Y.; Song, P.; Yang, T.; Zhang, Y. A bci-based environmental control system for patients with severe spinal cord injuries. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 1959–1971. [[CrossRef](#)]
20. Rebsamen, B.; Burdet, E.; Guan, C.; Teo, C.L.; Zeng, Q.; Ang, M.; Laugier, C. Controlling a wheelchair using a bci with low information transfer rate. In Proceedings of the 2007 IEEE 10th International Conference on Rehabilitation Robotics, Noordwijk, The Netherlands, 13–15 June 2007; pp. 1003–1008. [[CrossRef](#)]
21. Iturrate, I.; Antelis, J.M.; Kubler, A.; Minguez, J. A noninvasive brainactuated wheelchair based on a p300 neurophysiological protocol and automated navigation. *IEEE Trans. Robot.* **2009**, *25*, 614–627. [[CrossRef](#)]
22. Alqasemi, R.; Dubey, R. A 9-dof wheelchair-mounted robotic arm system: Design, control, brain-computer interfacing, and testing. In *Advances in Robot Manipulators*; InTech: Rijeka, Croatia, 2010. [[CrossRef](#)]
23. Shin, B.; Kim, T.; Jo, S. Non-invasive brain signal interface for a wheelchair navigation. In Proceedings of the International Conference on Control Automation and Systems, Gyeonggi-do, Republic of Korea, 27–30 October 2010. [[CrossRef](#)]
24. Lopes, A.C.; Pires, G.; Nunes, U. Assisted navigation for a brain-actuated intelligent wheelchair. *Robot. Auton. Syst.* **2013**, *61*, 245–258. [[CrossRef](#)]
25. Zhang, R.; Li, Y.; Yan, Y.; Zhang, H.; Wu, S.; Yu, T.; Gu, Z. Control of a wheelchair in an indoor environment based on a brain–computer interface and automated navigation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *24*, 128–139. [[CrossRef](#)] [[PubMed](#)]
26. Wang, Y.; Wang, R.; Gao, X.; Hong, B.; Gao, S. A practical vepbased brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2006**, *14*, 234–240. [[CrossRef](#)] [[PubMed](#)]
27. Dasgupta, S.; Fanton, M.; Pham, J.; Willard, M.; Nezamfar, H.; Shafai, B.; Erdogmus, D. Brain controlled robotic platform using steady state visual evoked potentials acquired by eeg. In Proceedings of the 2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, USA, 7–10 November 2010; pp. 1371–1374. [[CrossRef](#)]
28. Prueckl, R.; Guger, C. Controlling a robot with a brain-computer interface based on steady state visual evoked potentials. In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010; pp. 1–5. [[CrossRef](#)]
29. Mandel, C.; Lüth, T.; Laue, T.; Röfer, T.; Gräser, A.; KriegBrückner, B. Navigating a smart wheelchair with a brain-computer interface interpreting steady-state visual evoked potentials. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, MO, USA, 10–15 October 2009; pp. 1118–1125. [[CrossRef](#)]



30. Xu, Z.; Li, J.; Gu, R.; Xia, B. Steady-state visually evoked potential (ssvep)-based brain-computer interface (bci): A low-delayed asynchronous wheelchair control system. In *Neural Information Processing*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 305–314. [[CrossRef](#)]
31. Müller, S.M.; Bastos, T.F.; Filho, M.S. Proposal of a ssvep-bci to command a robotic wheelchair. *J. Control. Autom. Electr. Syst.* **2013**, *24*, 97–105. [[CrossRef](#)]
32. Diez, P.F.; Müller, S.M.; Mut, V.A.; Laciari, E.; Avila, E.; Bastos, T.F.; Filho, M.S. Commanding a robotic wheelchair with a high-frequency steady-state visual evoked potential based brain-computer interface. *Med. Eng. Phys.* **2013**, *35*, 1155–1164. [[CrossRef](#)]
33. Duan, J.; Li, Z.; Yang, C.; Xu, P. Shared control of a brain-actuated intelligent wheelchair. In Proceedings of the 11th World Congress on Intelligent Control and Automation (WCICA), Shenyang, China, 29 June–4 July 2014; pp. 341–346. [[CrossRef](#)]
34. Larsen, E.A. Classification of Eeg Signals in a Brain-Computer Interface System. Master's Thesis, Institutt for Datateknikk og Informasjonsvitenskap, Trondheim, Norway, 2011.
35. Khare, V.; Santhosh, J.; Anand, S.; Bhatia, M. Brain computer interface based real time control of wheelchair using electroencephalogram. *Int. J. Soft Comput. Eng. IJSCE* **2011**, *1*, 41–45.
36. Choi, K. Control of a vehicle with eeg signals in real-time and system evaluation. *Eur. J. Appl. Physiol.* **2012**, *112*, 755–766. [[CrossRef](#)]
37. Barbosa, A.O.; Achancaray, D.R.; Meggiolaro, M.A. Activation of a mobile robot through a brain computer interface. In Proceedings of the 2010 IEEE International Conference on Robotics and Automation (ICRA), Anchorage, AK, USA, 3–7 May 2010; pp. 4815–4821. [[CrossRef](#)]
38. Tsui, C.S.; Gan, J.Q.; Roberts, S.J. A self-paced brain-computer interface for controlling a robot simulator: An online event labelling paradigm and an extended kalman filter based algorithm for online training. *Med. Biol. Eng. Comput.* **2009**, *47*, 257–265. [[CrossRef](#)]
39. Tang, Z.; Sun, S.; Zhang, S.; Chen, Y.; Li, C.; Chen, S. A brain-machine interface based on erd/ers for an upper-limb exoskeleton control. *Sensors* **2016**, *16*, 2050. [[CrossRef](#)]
40. Bahri, Z.; Abdulaal, S.; Buallay, M. Sub-band-power-based efficient brain computer interface for wheelchair control. In Proceedings of the 2014 World Symposium on Computer Applications & Research (WSCAR), Sousse, Tunisia, 18–20 January 2014; pp. 1–7. [[CrossRef](#)]
41. Li, M.; Zhang, Y.; Zhang, H.; Hu, S. An eeg based control system for intelligent wheelchair. In *Applied Mechanics and Materials*; Trans Tech Publications: Bäch, Switzerland, 2013; Volume 300, pp. 1540–1545. [[CrossRef](#)]
42. Dharmasena, S.; Lalitharathne, K.; Dissanayake, K.; Sampath, A.; Pasqual, A. Online classification of imagined hand movement using a consumer grade eeg device. In Proceedings of the 2013 8th IEEE International Conference on Industrial and Information Systems (ICIIS), Peradeniya, Sri Lanka, 17–20 December 2013; pp. 537–541. [[CrossRef](#)]
43. Batres-Mendoza, P.; Ibarra-Manzano, M.; Guerra-Hernandez, E.; Almanza-Ojeda, D.; Montoro-Sanjose, C.; Romero-Troncoso, R.; Rostro-Gonzalez, H. Improving eeg-based motor imagery classification for real-time applications using the qsa method. *Comput. Intell. Neurosci.* **2017**, *2017*, 9817305. [[CrossRef](#)]
44. Cao, L.; Li, J.; Ji, H.; Jiang, C. A hybrid brain computer interface system based on the neurophysiological protocol and brain-actuated switch for wheelchair control. *J. Neurosci. Methods* **2014**, *229*, 33–43. [[CrossRef](#)]
45. Li, J.; Ji, H.; Cao, L.; Zang, D.; Gu, R.; Xia, B.; Wu, Q. Evaluation and application of a hybrid brain computer interface for real wheelchair parallel control with multi-degree of freedom. *Int. J. Neural Syst.* **2014**, *24*, 1450014. [[CrossRef](#)]
46. Allison, B.Z.; Brunner, C.; Kaiser, V.; Müller-Putz, G.; Neuper, C.; Pfurtscheller, G. Toward a hybrid brain-computer interface based on imagined movement and visual attention. *J. Neural Eng.* **2010**, *7*, 026007. [[CrossRef](#)]
47. Long, J.; Li, Y.; Wang, H.; Yu, T.; Pan, J. Control of a simulated wheelchair based on a hybrid brain computer interface. In Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), San Diego, CA, USA, 28 August–1 September 2012; pp. 6727–6730. [[CrossRef](#)]
48. Rani, B.J.; Umamakeswari, A. Electroencephalogram-based brain controlled robotic wheelchair. *Indian J. Sci. Technol.* **2015**, *8* (Suppl. 9), 188–197. [[CrossRef](#)]
49. Almeida, P.; Faria, B.M.; Reis, L.P. Brain Waves Classification Using a Single-Channel Dry EEG Headset: An Application for Controlling an Intelligent Wheelchair. In *Advances in Practical Applications of Agents, Multi-Agent Systems, and Cognitive Mimetics*; The PAAMS Collection; Lecture Notes in Computer Science; Mathieu, P., Dignum, F., Novais, P., De la Prieta, F., Eds.; Springer: Cham, Switzerland, 2023; Volume 13955. [[CrossRef](#)]
50. Open Bci-Publicly Available EEG Datasets. Available online: <https://openbci.com/community/publicly-available-eeg-datasets/> (accessed on 5 February 2024).
51. Lotte, F.; Bougrain, L.; Cichocki, A.; Clerc, M.; Congedo, M.; Rakotomamonjy, A.; Yger, F. A review of classification algorithms for eeg-based brain-computer interfaces: A 10 year update. *J. Neural Eng.* **2018**, *15*, 031005. [[CrossRef](#)]
52. Bzdok, D.; Krzywinski, M.; Altman, N. Points of significance: Machine learning: A primer. *Nat. Methods* **2017**, *14*, 1119–1120. [[CrossRef](#)]
53. Johansson, M. Novel cluster-based svm to reduce classification error in noisy eeg data: Towards real-time brain-robot interfaces. *Comput. Biol. Med.* **2018**, *148*, 105931. [[CrossRef](#)]

54. Singh, A.; Lal, S.; Guesgen, H.W. Architectural review of co-adaptive brain computer interface. In Proceedings of the 2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), Mana Island, Fiji, 11–13 December 2017; pp. 200–207. [CrossRef]
55. Pejas, J.; El Fray, I.; Hyla, T.; Kacprzyk, J. (Eds.) *Advances in Soft and Hard Computing*; Springer: Berlin/Heidelberg, Germany, 2018. [CrossRef]
56. Fernández-Rodríguez, A.; Álvarez, F.; Angevin, R. Review of real brain-controlled wheelchairs. *J. Neural Eng.* **2016**, *13*, 061001. [CrossRef] [PubMed]
57. Abiyev, A.H.; Akkaya, N.; Aytac, E.; Günsel, I.; Çağman, A. Brain based control of wheelchair. In Proceedings of the International Conference on Artificial Intelligence (ICAI), Las Vegas, NV, USA, 27–30 July 2015; p. 542. Available online: <https://api.semanticscholar.org/CorpusID:13468342> (accessed on 9 April 2024).
58. Carrino, F.; Dumoulin, J.; Mugellini, E.; Khaled, O.A.; Ingold, R. A self-paced bci system to control an electric wheelchair: Evaluation of a commercial, lowcost eeg device. In Proceedings of the 2012 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC), Manaus, Brazil, 9–11 January 2012; pp. 1–6. [CrossRef]
59. Rincon, J.; Jaramillo, S.; Cespedes, Y.; Meza, A.M.; Domínguez, G. Motor imagery classification using feature relevance analysis: An emotiv-based bci system. In Proceedings of the 2014 XIX Symposium on Image, Signal Processing and Artificial Vision, Armenia, Colombia, 17–19 September 2014; pp. 1–5. [CrossRef]
60. Lin, J.; Lo, C. Mental commands recognition on motor imagery-based brain computer interface. *Int. J. Comput. Consum. Control.* **2016**, *25*, 18–25.
61. Muñoz, J.; Ríos, L.H.; Henao, O. Low cost implementation of a motor imagery experiment with bci system and its use in neurorehabilitation. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 2014. [CrossRef]
62. Ang, K.; Chin, Z.; Wang, C.; Guan, C.; Zhang, H. Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b. *Front. Neurosci.* **2012**, *6*, 39. [CrossRef] [PubMed]
63. Stock, V.; Balbinot, A. Movement imagery classification in emotiv cap based system by naïve bayes. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 4435–4438. [CrossRef]
64. Kim, P.; Kim, K.; Kim, S. Using common spatial pattern algorithm for unsupervised real-time estimation of fingertip forces from semg signals. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 5039–5045. [CrossRef]
65. Siemens PLM Community. Spectral Leakage. Available online: <https://community.plm.automation.siemens.com/t5/Testing-Knowledge-Base/Windows-and-Spectral-Leakage/ta-p/432760> (accessed on 5 December 2023).
66. Fakhruzzaman, M.N.; Riksakomara, E.; Suryotrisongko, H. Eeg wave identification in human brain with Emotiv EPOC for motor imagery. *Procedia Comput. Sci.* **2016**, *72*, 269–276. [CrossRef]
67. Siribunyaphat, N.; Punsawad, Y. Brain–Computer Interface Based on Steady-State Visual Evoked Potential Using Quick-Response Code Pattern for Wheelchair Control. *Sensors* **2023**, *23*, 2069. [CrossRef] [PubMed]
68. Ramírez-Arias, F.J.; García-Guerrero, E.E.; Tlelo-Cuautle, E.; Colores-Vargas, J.M.; García-Canseco, E.; López-Bonilla, O.R.; Galindo-Aldana, G.M.; Inzunza-González, E. Evaluation of Machine Learning Algorithms for Classification of EEG Signals. *Technologies* **2022**, *10*, 79. [CrossRef]
69. Sabio, J.; Williams, N.S.; McArthur, G.M.; Badcock, N.A. A scoping review on the use of consumer-grade EEG devices for research. *PLoS ONE* **2024**, *19*, e0291186. [CrossRef] [PubMed] [PubMed Central]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.