

Review

A Comprehensive Review on Critical Issues and Possible Solutions of Motor Imagery Based Electroencephalography Brain-Computer Interface

Amardeep Singh * , Ali Abdul Hussain , Sunil Lal  and Hans W. Guesgen 

School of Fundamental Sciences, Massey University, 4410 Palmerston North, New Zealand; A.abdulahussain@massey.ac.nz (A.A.H.); S.Lal@massey.ac.nz (S.L.); H.W.Guesgen@massey.ac.nz (H.W.G.)

* Correspondence: A.Singh1@massey.ac.nz

Abstract: Motor imagery (MI) based brain–computer interface (BCI) aims to provide a means of communication through the utilization of neural activity generated due to kinesthetic imagination of limbs. Every year, a significant number of publications that are related to new improvements, challenges, and breakthrough in MI-BCI are made. This paper provides a comprehensive review of the electroencephalogram (EEG) based MI-BCI system. It describes the current state of the art in different stages of the MI-BCI (data acquisition, MI training, preprocessing, feature extraction, channel and feature selection, and classification) pipeline. Although MI-BCI research has been going for many years, this technology is mostly confined to controlled lab environments. We discuss recent developments and critical algorithmic issues in MI-based BCI for commercial deployment.

Keywords: motor imagery; brain–computer interface (BCI); BCI illiteracy; adaptive BCI; online BCI; asynchronous BCI; BCI calibration; BCI training; electroencephalography (EEG)



Citation: Singh, A.; Hussain, A.A.; Lal, S.; Guesgen, H.W. A Comprehensive Review on Critical Issues and Possible Solutions of Motor Imagery Based Electroencephalography Brain-Computer Interface. *Sensors* **2021**, *21*, 2173. <https://doi.org/10.3390/s21062173>

Academic Editor: Yvonne Tran

Received: 28 December 2020

Accepted: 16 March 2021

Published: 20 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

Numerous people with serious motor disorders are unable to communicate properly if at all. This significantly impacts their quality of life and ability to live independently. In this respect, brain–computer interface (BCI) aims to provide a means of communication. BCIs translate the acquired neural activity into control commands for external devices [1]. Primarily, BCI systems can be cast into various categories that are based on interactions with a user interface and neuroimaging technique applied to capture neural activity. Based on users' interaction with brain–computer interface, the EEG-BCI system is categorized into synchronous and asynchronous BCI. In the synchronous BCI system, brain activity is generated by the user, which is based on some cue or event taking place in the system at a certain time. This cue helps in differentiating between intentional neural activity for a control signal from unintentional neural activity in the brain [2]. On the other hand, asynchronous BCI works independently of a cue. The asynchronous BCI system also needs to differentiate between neural activity that a user intentionally generates from the unintentional neural activity [3].

Based on neuroimaging techniques, BCI systems fall into invasive and non-invasive categories. In an invasive BCI, neural activity is captured under the skull, thus requiring the surgery to plant the sensors in different parts of the brain. This results in a high-quality signal, is but prone to scar tissue build-up over time, resulting in a loss of signal [4].

Additionally, once the implanted sensors cannot be moved to measures the other parts of the brain [5]. In contrast to this, non-invasive BCI captures brain activity from the surface of the skull. A signal that is acquired through non-invasive technologies has a low signal to noise ratio. Electrocorticography (ECoG) and micro electrodes are some examples of invasive neuroimaging techniques. Electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and

functional near infrared (fNIR) are examples of non-invasive neuroimaging techniques [6]. All of these methods work on different principles and they provide different levels of portability, spatial, and temporal resolution [7]. Among these brain imaging methods, an EEG is widely employed because of its ease of use, safety, high portability, relatively low cost, and, most importantly, high temporal resolution.

Electroencephalography (EEG) is one of the non-invasive and portable neuroimaging techniques that records electrical activity generated due to the synchronized activity of cerebral neurons. Primarily, pyramidal neurons' activity contributes more to EEG recordings because of their very stable orientated electric field to the cortical surface [6]. This is due to the perpendicular orientation of pyramidal cells with respect to the cortical surface. As a result, the electrical field is projected stably towards the scalp in contrast to the other brain cells whose electrical field is very dispersed and cancels out [7]. The measured EEG signal is due to the complex firing pattern of billions of neurons in the brain. Owing to this pattern, the EEG signal is a combination of various rhythms that reflect certain cognitive states of the individual [7]. These rhythms have different properties, like frequency, amplitude, and shape etcetera. These properties depend upon individual, external stimulus, and the internal state of the individual. Broadly, these rhythms are classified into various categories that are based on their frequency, amplitude, shape, and spatial localization [6]. Furthermore, these rhythms are broadly categorized under six frequency bands: delta band (1–4 Hz), theta band (4–8 Hz), alpha band (8–12 Hz), mu band (8–12 Hz), beta band (13–25 Hz), and gamma band (>25 Hz). EEG control signals can be categorized as evoked and spontaneous. An evoked signal corresponds to neural activity that is generated due to external stimuli. Examples of evoked control signals are steady-state visual-evoked potentials (SSVEP), visual-evoked potentials (VEP), and P300 [4]. On the other hand, a spontaneous control signal is due to voluntarily neural activity without any aid of external stimulus. Slow cortical potentials (SCPs) and sensorimotor rhythms (SMRs) are such control signals [4]. As mentioned above, an evoked control signal requires external stimulus, thus the user needs to focus on presentation to generate neural activity. This continuous focus causes fatigue in users. Nevertheless, much less training is required to generate evoked control signals. Spontaneous control signals offer natural control over neural activity, but they require long training to master self regulation of brain rhythms. To do so, different cognitive tasks are employed to generate spontaneous control signals.

Motor imagery (MI) is one of the most widely used cognitive tasks, which corresponds to sensorimotor rhythms (SMRs) as a control signal. Motor imagery has advantages for the brain–computer interface in both synchronous and asynchronous mode. MI can be defined as the user sending a command to a system through the imagination of a kinesthetic movement of his/her limbs. For example, a user moving a prosthetic arm by imagining his/her left/right hand moving. The imagination of movement creates a similar brain activity to that of an actual movement, which decreases the percentage of power relative to a reference baseline in both the mu and beta frequencies over the sensorimotor cortex; this is known as event related desynchronization (ERD) [8]. Immediately after the user's imagination task, the user's brain activity can experience an event-related synchronization (ERS), which is the increase to the percentage of power relative to the reference baseline [8]. Because ERD/ERS are mixed with other brain activity created unintentionally by the user, such as involuntarily muscle movements and eye blinks, the signal to noise ratio (SNR) is low. The algorithm that is designed for MI-BCI must be able to differentiate between MI activity for control signal from other involuntary activity. In doing so, the MI-BCI pipeline consists of many stages, like data acquisition, preprocessing, feature extraction, and classification. Therefore, the objective of this manuscript is to review the MI based BCI system with regards to algorithms that were utilized at different stages of MI-BCI pipeline. This brief survey is structured under an architectural framework that helps in mapping the literature to each component of the MI-BCI pipeline. In doing so, this article identifies critical research gaps that warrant further exploration along with current developments to mitigate these issues.

Figure 1 breaks down the contents of the entire article. This review article is divided into two parts. The first part of this article introduces the Architecture of MI based BCI. More specifically, how the EEG signal is captured from the brain is described under Section 2.1. In Section 2.2, we discuss how during the calibration phase the user acquires skills to modulate brain waves into control commands. The signal pre-processing subsection explains how unwanted artifacts are removed from the EEG signal to improve the signal to noise ratio. Section 2.4 discusses different approaches to extract information that is related to a motor imagery event in terms of features that are finally classified into control commands. Subsection on Sections 2.5 and 2.6 deals with issues related to finding optimal channels or features and reducing dimensionality of feature space in order to improve BCI performance. Section 2.7 provides details of how features are classified into control commands. Lastly, Section 2.8 covers how to evaluate the performance of BCI. The last part of this article discusses the key issues that need further exploration along with the current state of the art that address these research challenges.

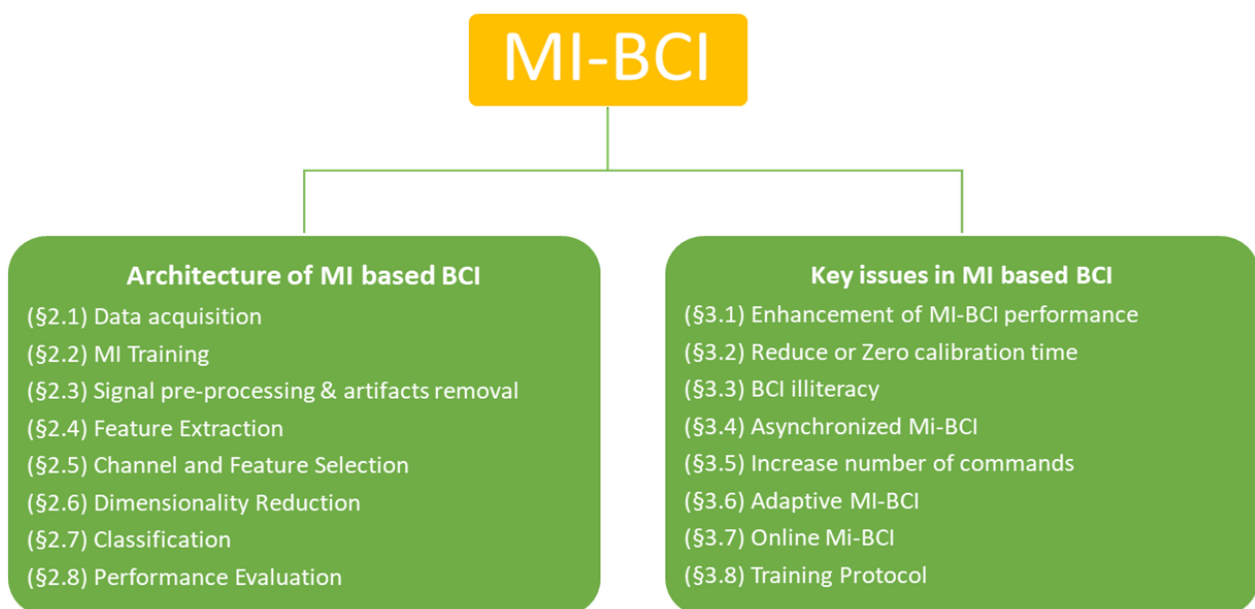


Figure 1. Breakdown of the article.

2. Architecture of MI Based BCI

We present a framework of MI-BCI pipeline encompassing all of the components that are responsible for its working in Figure 2. In short, MI-BCI works in calibration and online mode, respectively. During calibration mode, the user learns voluntary ERD/ERS regulation in the EEG signal and BCI learns ERD/ERS mapping through temporal, spectral, and spatial characteristics of the user's EEG signal. In online mode, the user's characteristics are translated into a control signal for external application and feedback is given to the user. In framework, optional steps that are enclosed in yellow box, such as channel selection, feature selection, and dimensionality reduction. This framework is also helpful in mapping the literature to different components of the MI-BCI pipeline in order to understand the current research gaps.

2.1. Data Acquisition

The signal acquisition unit is represented by electrodes whether they are invasive or non-invasive. In the non-invasive approach, electrodes are usually connected with the skin via conductive gel to create a stable electrical connection for a good signal. The combination of conductive gel and electrode attenuate the transmission of low frequencies, but take a very long time to setup. Another alternative is dry electrodes, which make direct contact with skin without conductive gel. Dry electrodes are easy and faster to apply, but are more

prone to motion artifacts [5]. EEG signals are usually acquired under unipolar and bipolar modes. In unipolar mode, a potential difference between all the electrodes with respect to one reference are acquired. Each electrode-reference pair form one EEG channel. On the contrary, in bipolar mode, the potential difference between two specified electrodes are acquired and each pair make a EEG channel [9]. To standardize positions and naming, electrodes are placed on the scalp under international 10–20 standard. This helps in reliable data collection and consistency among different BCI sessions [10].

Figure 3 shows the international 10–20 electrodes' placement scheme from the side and top view of the head. Once the potential difference has been identified by the EEG electrodes, it is amplified and digitized in order to store it in a computer. This process can be expressed as taking one sample (discrete snapshots) of the continuous cognitive activity. This discrete snapshot (sample) depends on the sampling rate of the acquisition device. For example, an EEG acquisition device with a sampling rate of 256 Hz can take 256 samples per second. High sampling rates and more EEG channels are used to increase the temporal and spatial resolutions of an EEG acquisition device.

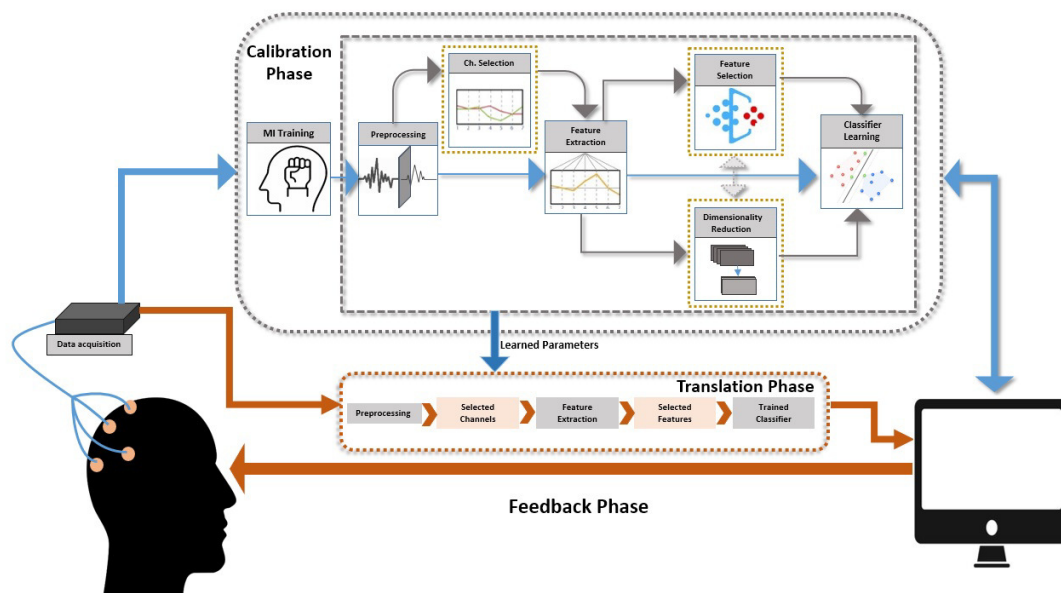


Figure 2. Block diagram showing the typical structure of MI-based brain–computer interface (BCI).

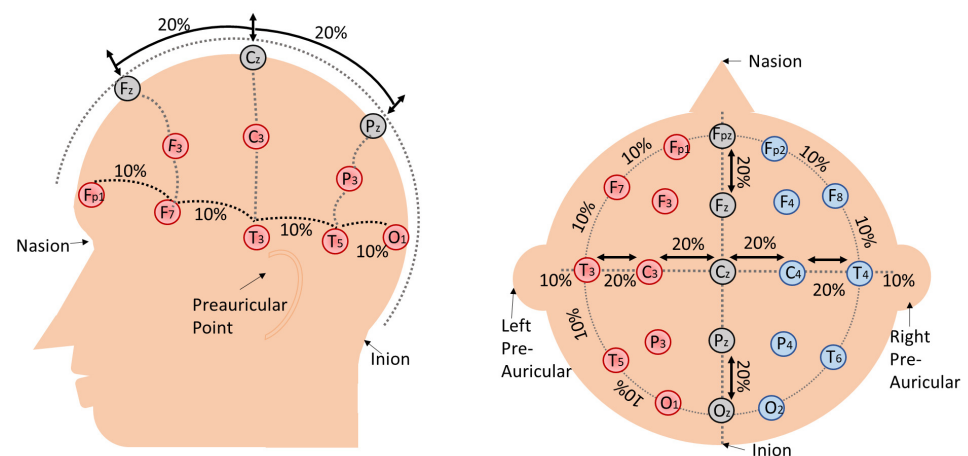


Figure 3. The international 10–20 standard electrode position system. The left image presents a left side view of the head with electrode positions, and the right image shows a top view of the head.

2.2. MI Training

During calibration phase, the user learns how to modulate EEG signals with MI task pattern. Just as with any skill, MI training helps in acquiring the ability to produce a distinct and stable EEG pattern while performing the different MI tasks [11]. The Graz training paradigm is the standard training approach for motor imagery [8,11]. The Graz approach is based on machine learning, where the system adapts with the user's EEG pattern. During this training paradigm, the user is instructed through a cue to perform a motor imagery task, such as left and right-hand imagination. EEG signals that are collected during different imagination tasks are used to train the system differentiate between the MI-tasks from the EEG pattern. Once the system is trained, users are instructed to perform MI tasks, but this time feedback is provided to the user. This process is repeated multiple times over different sessions. Each session has further multiple runs of the Graz training protocol.

The trial time vary depending on scenario. Typically, one trial of graz training protocol lasts eight seconds, as illustrated in Figure 4. At the outset of each MI trail, which is $t = 0$ s, a fixation cross is displayed to instruct the user that the trial has started. After a two-second break ($t = 2$ s), a beep is used to prepare the user for the upcoming MI task. This 2 s break acts as a baseline period to see the MI task pattern in the EEG signal in the upcoming MI task at $t = 3$ s. After three seconds, an arrow appears on the screen indicating the MI task. For example, the arrow in the right direction means right hand motor imagery. No feedback is provided during the initial training phase. After the system is calibrated, feedback is provided for four seconds. The direction of the feedback bar shows recognition of the MI pattern by system and the length of the bar represents confidence of the system in its recognition of the MI class pattern.

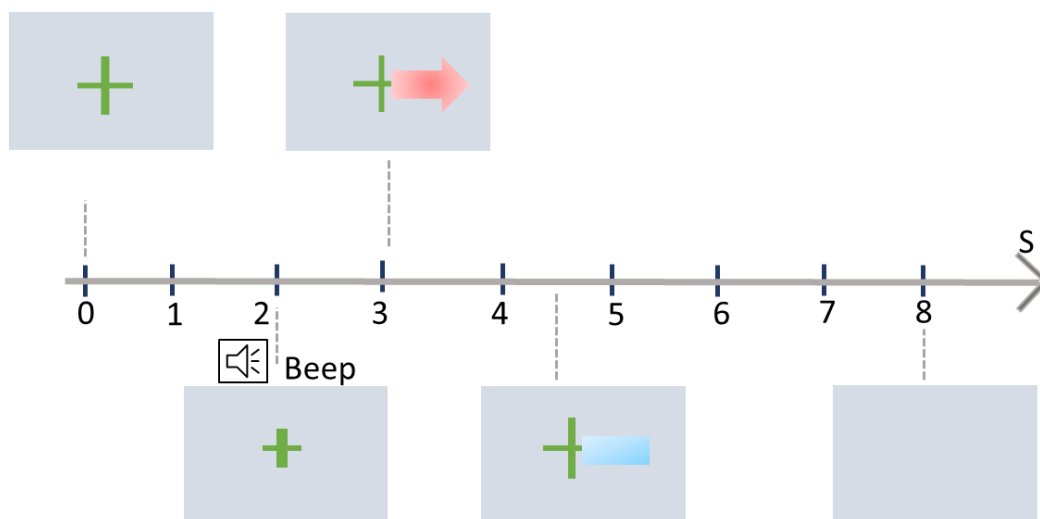


Figure 4. An illustration of one trial's timing in the Graz protocol [11].

Various other extensions of the Graz paradigm is proposed in the literature, mostly focusing on providing alternative MI instructions and feedback from the system. For example, the bar feedback is replaced by auditory [12] and tactile [13] feedback to reduce the workload on the visual channel. Similarly, virtual reality based games and environments are explored to provide MI instructions and feedback for training [14,15].

2.3. Signal Pre-Processing and Artifacts Removal

Artifacts are nothing but unwanted activities during signal acquisition. They are comprised of an incorrect collection of signal or signals acquired from areas other than the cerebral origin of the scalp area. Generally, artifacts are classified into two major categories, termed as endogenous and exogenous artifacts. Endogenous artifacts are generated from the human body excluding the brain, and extra-physiologic artifacts are generated from external sources (i.e., sources from outside the human body) [7]. Some of

the common endogenous and exogenous artifacts that accrue during EEG signal acquisition are bad electrode position, poor ground electrode, obstructions to electrode path (e.g., hair), eye blinks, electrode impedance, electromagnetic noise, equipment problem, power line interference, ocular artifacts, cardiac artifacts, and muscle disturbances [16]. The signal pre-processing block is responsible for the removal of such exogenous and endogenous artifacts from the EEG signal. MI-BCI systems mainly rely on a temporal and spatial filtering approach.

Temporal filtering is the most commonly used pre-processing approach for EEG signals. Temporal filters are usually low pass or band pass filters that are used to restrict signals in the frequency band where neurophysiological information relevant to the cognitive task lies. For MI, this usually means a Butterworth or Chebyshev bandpass filter of 8–30 Hz frequency. This bandpass filter keeps both the mu and beta frequency bands as they are known to be associated with motor-related tasks [8]. However, MI task-related information is also present in the spatial domain. Similar to temporal filters, spatial filters extract the necessary spatial information associated with a motor-related task embedded in EEG signals. A common average reference (CAR) is a spatial filter that removes the common components from all channels, leaving channels with only channel specific signals [17]. This is done by removing the mean of all k channels from each channel x_i :

$$x_i^{\text{CAR}} = x_i - \text{mean}(x_i). \quad (1)$$

CAR benefits from being a very computationally cheap approach. An updated version of CAR is the Laplacian spatial filter. The Laplacian spatial filter aims to remove the common components of neighboring signals, which increases the difference between channels [18]. The Laplacian spatial filter is calculated through the following equation:

$$V_i^{\text{LAP}} = V_i^{\text{ER}} - \sum_{j \in S_i} g_{ij} V_j^{\text{ER}} \quad (2)$$

$$g_{ij} = \frac{d_{ij}}{\sum_{j \in S_i} d_{ij}} \quad (3)$$

where V_i^{LAP} is the i th channel that is filtered by the Laplacian method, V_i^{ER} is the potential difference between i th electrode and reference electrode, S_i is the set of neighboring electrodes to the i th electrode, and d_{ij} is the Euclidean distance between i th and j th electrode [18].

2.4. Feature Extraction

Measuring motor imagery through an EEG leads to a large amount of data due to high sampling rate and electrodes. In order to achieve the best possible performance, it is necessary to work with small values that are capable of discriminating MI task activity from unintentional neural activity. These small values are called “features” and the process to achieve these values is called “feature extraction”. Formally, feature extraction is the mapping of preprocessed large EEG data into a feature space. This feature space should contain all of the necessary discrimination information for a classifier to do its job. For MI-BCI, the feature extraction methods can be divided into six categories: (a) time domain methods that exploit temporal information embedded in the EEG signal; (b) spectral methods extract information that is embedded in the frequency domain of EEG signals; (c) time-frequency methods works together on information in the time and frequency domain; (d) spatial methods extract spatial information from EEG signals coming from multiple electrodes; (e) spatio-temporal methods works together with spatial and temporal information to extract features; (f) spatio-spectral methods use spatial and spectral information of the multivariate EEG signals for feature extraction; and, (e) Riemannian Manifold methods, which are essentially a sub category of spatio-temporal methods that exploits manifold properties of EEG data for feature extraction. Table 1 summarizes all of the feature extraction methods discussed in the following subsections.

Table 1. This table provides a summary of the feature extraction methods.

A Summary of Feature Extraction Methods		
Temporal methods	Statistical Features [19,20]	$Mean(\bar{x}) = \frac{1}{T} \sum_{t=1}^T x_t $ $Std.Dev(\sigma) = \sqrt{\frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})^2}$ $Variance(x) = \frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})^2$ $skewness = \frac{\frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})^3}{\sigma^3}$ $kurtosis = \frac{\frac{1}{T} \sum_{t=1}^T (x_t - \bar{x})^4}{\sigma^4}$
	Hjorth features [21]	$Activity = \frac{\sum_{t=1}^T (x_t - \bar{x})^2}{T}$ $Mobility = \frac{Var(\dot{x}_t)}{Var(x_t)}$ $Complexity = \frac{Mobility(\dot{x}_t)}{Mobility(x_t)}$
	RMS [20]	$RMS_t = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
	IEEG [20]	$IEEG_t = \sum_{i=1}^N x_i $
	Fractal Dimension [22]	$D = \frac{\log(L/a)}{\log(d/a)}$
	Autoregressive modeling [21]	$x_t = \sum_{i=1}^p a_i x_{t-i} + \epsilon_t$ where {a for i = 1, ..., p} are AR model coefficients and p is the model order
	Peak-Valley modeling [23,24]	Cosine angles, Euclidean distance between neighbouring peak and valley points
	Entropy [25,26]	$S = -\sum_{i=1}^N p_i \ln p_i$
	Quaternion modeling [27]	$Mean(\mu) = \frac{\sum(q_{mod})}{N}$ $Variance(\sigma) = \frac{(\sum(q_{mod})^2 - \mu)^2 + \sum(q_{mod})^2}{2N}$ $Contrast(con) = \frac{\sum(q_{mod})^2}{N}$ $Homogeneity(H) = \sum \frac{(1)}{1 + (q_{mod})^2}$ $ClusterShade(cs) = \sum (q_{mod} - \mu)^3$ $Clusterprominence(cp) = \sum (q_{mod} - \mu)^4$
	Spectral methods	Band power [19]
Spectral Entropy [26]		$SH = -\sum_{f_1}^{f_2} \hat{P}(f) \log(\hat{P}(f))$ $\hat{P}(f) = P(f) / \sum_{f_1}^{f_2} P(f)$, P(f) is PSD of signal
Spectral statistical Features[19]		Mean Peak Frequency, Mean Power, Variance of Central Frequency etc.
Time-frequency Methods	STFT [28]	$S(m, k) = \sum_{n=0}^{N-1} s(n + mN)w(n)e^{-j\frac{2\pi}{N} nk}$
	Wavelet transform [29]	$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right)$
	EMD [30]	$x(t) = \sum_{i=1}^n c_i(t) + r_n(t)$
Spatial Methods	CSP [31]	$J(w) = \frac{w^T C_1 w}{w^T C_2 w}$
	BSS [32,33]	$x(t) = As(t)$ $s'(t) = Bx(t)$ Approaches like ICA, CCD estimate $s'(t)$
Spatio-temporal methods	Sample covariance matrices [34]	$C_i = \frac{X_i X_i^T}{tr(X_i X_i^T)}$ Where C_i is covariance matrix of single trial

2.4.1. Time Domain Methods

An EEG is a non-stationary signal whose amplitude, phase, and frequency changes with SMR modulations. Time domain methods investigate how the SMR modulation changes as a function of time [35]. Time domain methods work on each channel individually and extract temporal information related to the task. The extracted features from different channels are fused together to make a feature set for a single MI trial. In MI-BCI literature, statistical features, like mean, root mean square(RMS),integrated EEG (power of signal), standard deviation, variance, skewness, and kurtosis, are vastly employed to classify MI tasks [19,20]. Other alternative time domain methods that are based on variance of signal are Hjorth parameters. A Hjorth parameter measures power (activity), mean frequency (mobility) and change in frequency (complexity) of EEG signal [21]. Similarly, fractal dimension (FD) is non-linear method that measures EEG signals complexity [22]. Autoregressive (AR) modeling of the EEG signal is another typical time domain approach. The

AR models signal from each channel as a weighted combination of its previous samples and AR coefficients are used as features. An extension of AR modelling is adaptive auto regressive modelling (AAR) and it is also used for MI-BCI studies. Unlike AR, the coefficients in AAR are not constant and, in fact, varies with time [21]. Information theory-based features, like entropy, are also used in time domain to quantify complexity of the EEG signal [25]. Temporal domain entropy works with amplitude of EEG signal [26].

Another way of extracting temporal information is to represent the signal in terms of peaks (local maximum) and valley (local minimum) [23]. In this peak-valley representation, various features points are extracted between neighbouring peak and valley points. Using the peak-valley model, Yilmaz et al. [24] approximated EEG signal in 2D vector that contains cosine angle between transition points (peak or valley) and normalized the ratio of Euclidean distance in a left/right transition (peak or valley) points. In the same vein, Mendoza et al. [27] proposed a quaternion based signal analysis that represents a multi-channel EEG signal in terms of their orientation and rotation then obtained statistical features for classification. Recently, EEG signal analysis based on graph theory and functional connectivity (FC) is employed in MI-BCI [36]. These methods take advantage of the functional communication between the brain regions during cognitive task like MI. In graph based methods, the EEG data are represented through graph adjacency matrices that correspond to temporal correlations (correlation approaches used like Pearson or Correntropy) between different brain regions (electrodes). Features are extracted from this graph in terms of the graph node's importance, such as centrality measure [17].

The advent of data driven approaches, like deep learning, has largely alleviated the need for hand crafted features. In these approaches, a raw or preprocessed EEG signal is passed through different convolution and pooling layers to extract temporal information [37]. In the same vein, Lawhern et al. [38] proposed EEGNET deep learning architecture that works with raw EEG signals. It starts with a temporal convolution layer to learn the frequency filters (equivalent to preprocessing), another depth-wise convolution layer is used to learn frequency-specific spatial filters. Lastly, a combination of a depth-wise convolution along with point wise convolution are used to fuse features coming from previous layers for classification. Instead of using a raw or preprocessed signal, another approach is for the signal to be approximated and then passed to a deep neural network model. A one dimension-aggregate approximation (1d-AX) is one way of achieving this [39]. 1d-AX takes a signal from each channel in a single trial, normalizes it, and applies linear regression. These regression results are passed as features to the neural network.

2.4.2. Spectral Domain Methods

Spectral methods extract information from EEG signals in the frequency domain. Similar to the temporal method, statistical methods are also applied in the frequency domain. Samuel et al. [19] used statistical methods in both time and frequency domain to decode motor imagery. The most used spectral method is the power (energy) of EEG signals in specific frequency band. Usually, spectral power is calculated in mu (μ), beta (β), theta (θ), and delta (δ) frequency bands. This is done by decomposing the EEG signal into its frequency components at the chosen frequency band while using Fast Fourier Transform (FFT) [28,40]. The other frequency domain based method is Power Spectral Density (PSD). PSD is the measure of how the power of a signal is distributed over frequency. There are multiple methods of estimating it, such as Welch's averaged modified periodogram [41], Yule-Walker equation [42], or Lomb-Scargle periodogram [43]. Spectral entropy is another spectral feature that relies on PSD to quantify information in the signal [44].

2.4.3. Time-Frequency Methods

Time-frequency (t-f) methods works simultaneously in both temporal and spectral domains to extract information in signal. One of the approaches used in t-f domain is the short Term Fourier Transform (STFT), which segments the signal into overlapping time frames on which FFT is applied by the fixed window function [28]. Another way

to generate t-f spectra is through a wavelet transform [29], which decomposes the signal into wavelets (finite harmonic functions (sin/cos)). This captures the characteristics in the joint time-frequency domain. Another similar method in the t-f domain is empirical mode decomposition (EMD) [30]. However, instead of decomposing the signal into wavelets, it decomposes a signal $x(t)$ into simple oscillatory functions, called Intrinsic Mode Functions (IMFs) [45]. IMFs are a orthogonal representation of signals, such that first IMF captures a higher frequency and subsequent IMFs capture lower frequencies in EEG signals. Table 1 sums up all the t-f methods.

2.4.4. Spatial Domain Methods

Unlike temporal methods that work with only one channel at a time, spatial domain methods work with multiple channels. Spatial methods try to extract features by finding a combination of channels. This can be achieved while using blind source separation (BSS) [46]. BSS assumes that every single channel is the sum of clean EEG signals and several artifacts. Mathematically, this looks like the following:

$$x(t) = As(t)$$

where $x(t)$ is the channels, $s(t)$ is the sources, and A is mixing matrix. They aim to find a matrix B that reverse the channels back into their original sources:

$$s'(t) = Bx(t).$$

Examples of a BSS algorithms are Cortical current density (CCD) [32] and independent component analysis (ICA) [33]. BSS methods are unsupervised; thus, relations between the classes and features are unknown. However, there exist a supervised method that extract features based on class information, and one of such method is Common Spatial Pattern (CSP). CSP is based on the simultaneous diagonalization of two classes of EEG of their two respective estimated covariance matrices. CSP aims at learning a projection matrix W (spatial filters) that maximizes the variance of signal from one class while minimizing the variance from the other class [31]. This is mathematically represented as:

$$J(w) = \frac{w^T C_1 w}{w^T C_2 w}$$

where C_1, C_2 represent the estimated co-variance matrix of each MI class. The above equation can be solved while using the Lagrange multiplier method. CSP is known to be highly sensitive to noise and performs poorly under small sample settings, thus a regularized version has been developed [31]. There are two ways to regularize the CSP algorithm (also known as regularized CSP), either by penalizing its objective function $J(w)$, or regularizing its inputs (covariance matrices) [31]. One can regularize the objective function by adding a penalty term to the denominator:

$$J(w) = \frac{w^T C_1 w}{w^T C_2 w + \alpha P(w)}$$

where $P(\cdot)$ is a penalty function, and α is a constant that is determined by the user ($\alpha = 0$ for CSP) [31]. While CSP inputs can be regularized by:

$$\begin{aligned} \tilde{C}_c &= (1 - \gamma)\bar{C}_c + \gamma I \\ \bar{C}_c &= (1 - \beta)s_t C_c + \beta G_c \end{aligned}$$

where s_t is a scalar and G_c is a "generic" covariance matrix [31]. CSP performance becomes limited when the EEG signal is not filtered in the frequency range appropriate to subject. To address this issue, the filter bank CSP (FBCSP) algorithm is proposed that passes the signal through multiple temporal filters and CSP energy features are computed from each band [47].

Finally, CSP features from sub-bands are fused together for classification. This results in a large number of features, which limits the performance. To address this alternative method, sub-band common spatial pattern (SBCSP) is proposed, which employs linear discriminant analysis (LDA) to reduce the dimensionality. Finding multiple sub-bands to compute CSP energy features increases the computational cost. To solve this, discriminant filter bank CSP (DFBCSP) is proposed, which utilizes the fisher ratio (FR) to select most discriminant sub-bands from multiple overlapping sub-bands [48].

2.4.5. Spatio-Temporal and Spatio-Spectral Methods

Spatio-temporal methods are algorithms that manipulate both time and space (channels) domains. The main spatio-temporal methods that are used in past MI-BCI studies are Riemannian Manifold-based methods (discussed in the next section). Other spatio-temporal methods are usually based on deep learning. Echeverri et al. [46] proposed one such approach, which uses the BSS algorithm to separate the input signal $x(t)$ from a single channel into an equal number of estimated source signals $\hat{s}(t)$. These source signals are sorted, based on a correlation between their spectral components. Finally, continuous wavelet transform is applied to sorted source signals to achieve t-f spectra images that are further subjected to a convolution neural network (CNN) for classification. In the same vein, Li et al. [49] proposed an end-to-end EEG decoding framework that extracts the spatial and temporal features from raw EEG signals. In a similar manner, Yang et al. [50] proposed a combination long short-term memory network (LSTM) and convolutional neural network that concurrently learns temporal and spectral correlations from a raw EEG signal. In addition, they used discrete wavelet transformation decomposition to extract information in the spectral domain for classification of the MI task.

Like spatio-temporal methods, spatio-spectral methods extract information from spectral and spatial domains. Temporal and spatial filters are usually learned in sequential (linear) order, whereas, if they are learned simultaneously, a unified framework will be able to extract information from spatial and spectral domains. For instance, Wu et al. [51] employed a statistical learning theory to learn most discriminating temporal and spectral filters simultaneously. In the same vein, Suk and Lee [52] used a particle-filter algorithm and mutual information between feature vectors and class labels to learn spatio-spectral filters in a unified framework. Similarly, Zhang et al. [53] proposed a deep 3-D CNN network that was based on AlexNet that learns spatial and spectral EEG representation. Likewise, Bang et al. [54] proposed a method that generates 3D input feature matrix for the 3-D CNN network by stacking multiple-band spatio-spectral feature maps from multivariate EEG signal.

2.4.6. Riemannian Geometry Based Methods

Sample covariance matrices (SCM) calculated from EEG signals are widely used in BCI algorithms. SCM lie in the space of symmetric positive definite (SPD) matrices $P(n) = \{\mathbf{P} = \mathbf{P}^T, u^T \mathbf{P} u > 0, \forall u \in \mathbb{R}^n\}$ which forms a Riemannian Manifold [34]. Unlike the Euclidean space, the distance in the Riemannian manifold are curves, as shown in Figure 5. These curves can be measured while using Affine invariant Riemannian metric (AIRM) [55]: Let $X, Y \in \mathcal{S}_+^n$ be two SPD matrices. Then, the AIRM is given as

$$\delta_r^2(X, Y) = \|\text{Log}(X^{-1/2} Y X^{-1/2})\|_{\mathcal{F}}^2 = \|\text{Log}(X^{-1} Y)\|_{\mathcal{F}}^2.$$

Thus, methods in the Euclidean space can not be directly applied to SCMs. One way of using Euclidean methods to deal with SCMs is to project the SCM into a tangent space (see Figure 5). Because the Riemannian manifold (in fact any manifold) locally looks Euclidean, a reference point \mathbf{P}_{ref} for the mapping that is as close as possible to all data points must be chosen. This reference point is usually a Riemannian mean $\mathbf{P}_{ref} = \sigma(\mathbf{P}_i)$.

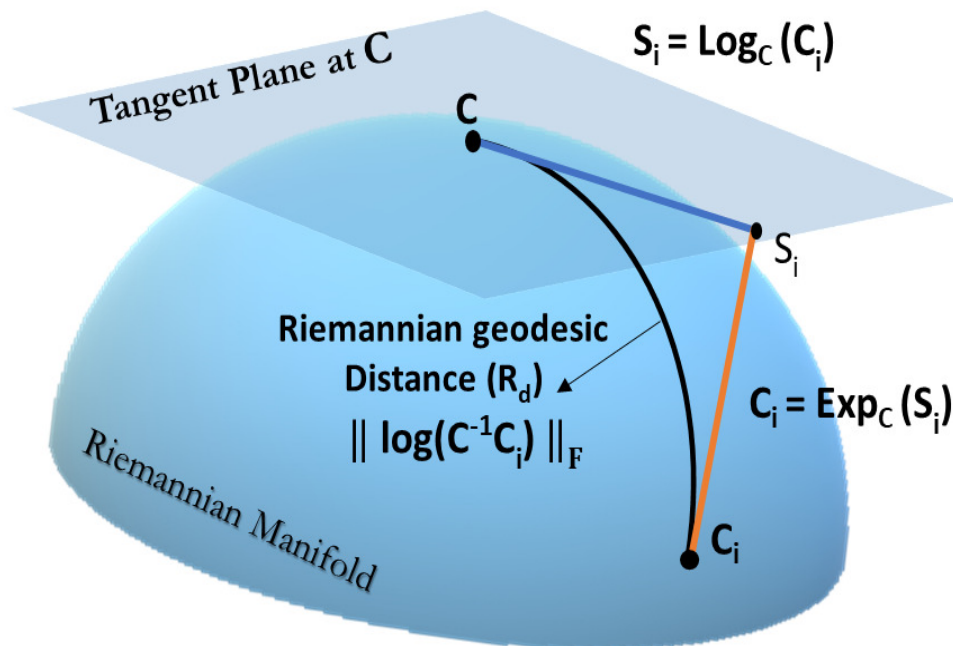


Figure 5. An image of the Riemannian Manifold displaying an example of a geodesic (the shortest distance between two Riemannian points), tangent space, and tangent mapping.

2.5. Channel and Feature Selection

EEG data are usually recorded through a large number of locations across the scalp. This provides a higher spatial resolution and benefits in identifying optimal locations (channels) that are relevant to BCI application or task. Here channel selection techniques significantly contribute to identify optimal channels for particular BCI application. Finding optimal channels not only reduces the computational cost of the system, but also reduces the subject's inconvenience due to the large number of channels. Thus, the main objective of channel selection methods is to identify optimal channels for the BCI task for improving the classification accuracy and reducing computation time in BCI. The channels' selection problem is similar to that of feature selection, where a subset of important features are selected from a vast number of features. Therefore, channel selection techniques are derived from feature selection algorithm. Once the channels are selected, we still need to extract features for classification of the BCI task. We are sometimes even required to use the feature selection algorithm on selected channels to improve the performance of the system. Feature or channel selection algorithms have many stages. Firstly, a candidate subset of features or channels are generated from an original set for evaluation purposes. This candidate subset is evaluated with respect to some selection criterion. This process is repeated for each candidate subset until a stopping criterion is reached. The selection criteria are what differentiates feature selection approaches. There are two stand-alone feature selection approaches filter approach and wrapper approach. A combination of both is sometimes used to make hybrid approaches also known as embedded approach. The embedded method exploits the strengths of both filter and wrapper approaches by combining them in feature selection process. Figure 6 shows a flow diagram of the above-mentioned feature selection techniques.

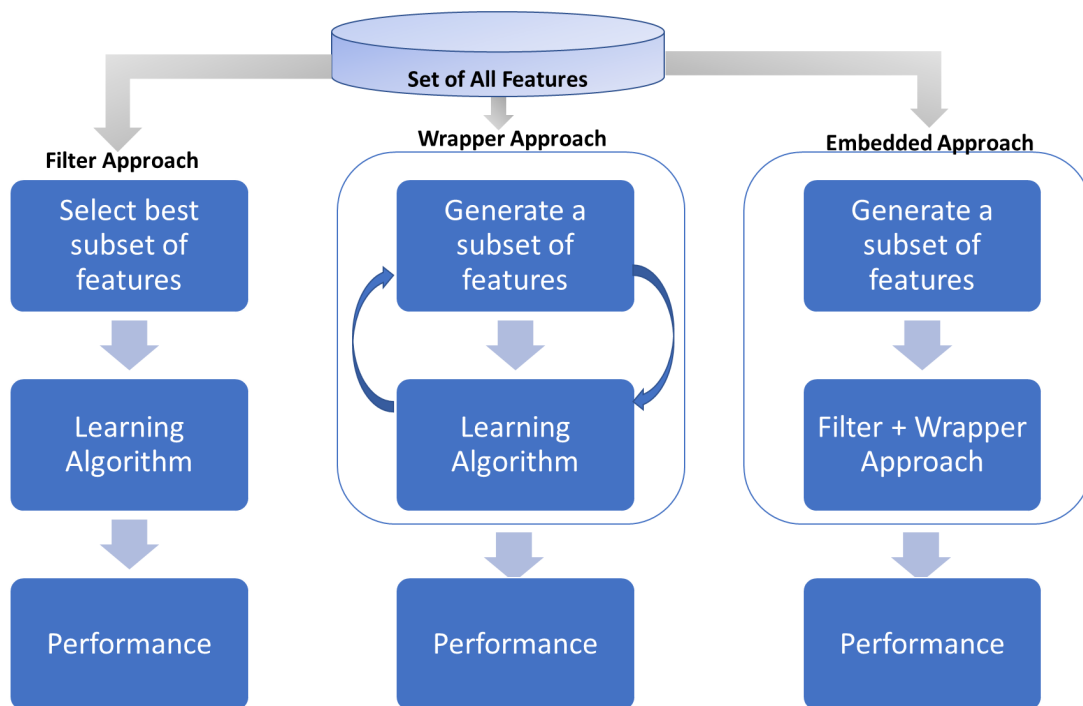


Figure 6. Flow diagram of different feature selection approaches.

2.5.1. Filter Approach

Filter methods start with all of the features and select the best subset of features based on some selection criteria. These selection criteria are usually based on characteristics, such as information gain, consistency, dependency, correlation, and distance measures [56]. The advantage of filter methods is their low computational cost and selection of features is independent of the learning algorithm (classifier). Some of the most widely employed filter methods are correlation criteria and mutual information. Correlation detects the linear dependence between variables x_i (features) and target Y (MI task classes). It is defined as:

$$R(i) = \frac{\text{cov}(x_i, Y)}{\sqrt{\text{var}(x_i)\text{var}(Y)}}$$

where $\text{cov}()$ is the covariance and $\text{var}()$ the variance. Mutual information (I) and its variant are widely used feature selection filter approaches in the MI-BCI literature. Mutual Information [57] $I(c_i; f)$ is the measure of the mutual dependence and uncertainty between two random variables: the features f and the classes c_i . This is measured by subtracting the uncertainty of the class $H(c_i)$ (that is also called initial uncertainty) from the uncertainty of the class given the features $H(c_i|f)$:

$$I(c_i; f) = H(c_i) - H(c_i|f)$$

Class uncertainty $H(c_i)$ and class uncertainty, both given the features $H(c_i|f)$, can be measured using Shannon's information theory entropy:

$$H(c_i) = - \sum_{i=1}^2 P(c_i) \log P(c_i)$$

$$H(c_i|f) = - \sum_{f=1}^{N_f} P(f) \left(\sum_{i=1}^2 P(c_i|f) \log P(c_i|f) \right)$$

where $P(c_i)$ is the probability density function of class c_i , and $P(c_i|f)$ is the conditional probability density function. When mutual Information is equal to zero $I(c_i|f) = 0$, the class c_i and the feature f are independent, and, as MI gets higher, the more relevant feature f to class c_i . Thus, MI can be used to select the features by relevance.

Similarly, t -test [58] measures the relevance of a feature to a class. It achieves this by examining mean $\mu_{i,j}$ and $\sigma_{i,j}$ variance of a feature f_j between class $i = \{1, 2\}$ through the following equation:

$$T(f_j) = \frac{|\mu_{1,j} - \mu_{2,j}|}{\sqrt{\frac{\sigma_{1,j}^2}{n_1} + \frac{\sigma_{2,j}^2}{n_2}}}$$

where n_i (n_1 and n_2) is the number of trials in class $i = \{1, 2\}$. This is then used to select a subset using the highest scoring features. Correlation based feature selection (CFS) [59] evaluates subsets of features based on the hypothesis that a good subset is the one that contains features that are highly correlated with the output classes and not correlated between them. This is computed using heuristic metric $Metric_S$ that divides the productiveness of k feature subset S by the redundancy that exists in the k features that compose the subset S :

$$Metric_S = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}}$$

where \bar{r}_{cf} is the mean of the class-feature correlation, \bar{r}_{ff} is the mean of the inter-feature correlation.

F -score [60] is another feature selection approach that quantifies the discriminative ability of variables (features) based on the following equation:

$$F - score_i = \frac{\sum_{k=1}^c (x_i^k - \bar{x}_i)^2}{\sum_{k=1}^c \left[\frac{1}{N_i^k} - 1 \sum_{j=1}^{N_i^k} (x_{ij}^k - \bar{x}_i^k) \right]} \quad (i = 1, 2, \dots, n)$$

where c is the number of classes, n is the number of features, N_i^k number of samples of feature i in class k , and x_{ij}^k is the j th training sample for feature i in class k . Features are ranked based on F -score, such that a higher F -score value corresponds to most discriminative feature.

2.5.2. Wrapper Approach

Wrapper approaches select a subset of features, present them as input to a classifier for training, observe the resulting performance, and stop the search according to a stopping criterion or propose a new subset if the criterion is not satisfied [56]. Algorithms that fall under the wrapper approach are mainly searching and evolutionary algorithms. Searching algorithms start with an empty set and add features (remove features) until a maximum possible performance from the learning algorithm is reached. Usually, a searching algorithm's stopping criteria is until the number of features reaches a maximum size that is specified for the subset. On the other hand, evolutionary algorithms, such as particle swarm optimization (PSO) [61], differential evolution (DE) [62,63], and artificial bee colony (ABC) [64,65], find an optimal feature subset by maximizing fitness function's performance. Wrapper methods find a more optimal feature subset when compared to filter methods, but the computational cost is very high, thus not being suitable for very large data-sets.

2.6. Dimensionality Reduction

In contrast to feature selection techniques, dimensionality reduction methods tends to reduce the number of features in data, but they do so by creating new combinations

(transformation) of features, whereas feature selection methods, achieve this by including and excluding features from the original feature set. Mathematically, dimensionality reduction can be defined as the transformation of high dimensional data ($\mathbf{X} \in \mathbf{R}^{\mathbf{D}}$) into a lower dimensional data ($\mathbf{Z} \in \mathbf{R}^{\mathbf{d}}$), where $\mathbf{d} \ll \mathbf{D}$. The dimensional reduction techniques can be categorized based on their objective function [66]. Those that are based on optimizing an convex (no local optima) objective function are convex techniques whereas techniques whose optimization function may have local optima are non-convex techniques. Furthermore, these techniques can be linear or non-linear based on the transform function used to map high dimensional to low dimension. The most used linear-convex technique is the Principal Component Analysis (PCA), which transforms data in a direction that maximizes the variance in the data set [67,68]. In a similar vein, Linear Discriminant Analysis (LDA) [69] is a linear dimensional reduction technique that finds a subspace that maximizes the distance between multiple classes. To do so, it uses class labels whereas PCA is an unsupervised technique. Independent Component Analysis (ICA) is another linear method that is found in EEG-BCI literature for dimensionality reduction, which works on the principle that the EEG signal is a linear mixture of various sources and all sources are independent of each other [70]. To address the non-linearity in a data-points structure, PCA can be extended by embedding it with a kernel function (KPCA) [70]. KPCA first transforms the data from the original space into kernel space using non-linear kernel transformation function, and then PCA is applied in kernel space. Likewise, Multilayer Autoencoders (AE) is an unsupervised, non-convex, and non-linear technique for reducing the dimensionality of data [66]. AE [71] takes the original data and reconstructs into lower dimensional output using the neural network. The drawback of the above discussed methods is that they do not consider the geometry of data prior to transformation. Thus, manifold learning for dimensionality reduction has recently gained more attention in MI-BCI research.

Manifold learning-based methods recover the original domain structure in reduced dimensional structure of data. Generally, these methods are non-linear and divided into global and local categories based on data matrix used for mapping high-dimensional to low-dimensional. Global methods used full EEG data covariance matrix and aim to retain global structure, and do not take the distribution of neighbouring points into account [72]. Isometric feature mapping (Isomap) [73,74] and diffusion maps [73,75] are some of these global methods. In order to preserve global structure of manifold, isomap and diffusion maps aim to preserve pairwise geodesic distance and diffusion distance between data-points, respectively. In contrast, local methods use a sparse matrix to solve eigenproblems, and their goal is to retain the local structure of the data. Locally, Linear Embedding [76,77], Laplacian eigenmaps [74], and local tangent space alignment (LTSA) [78] are some of these local methods. LLE assume manifold is linear locally and thus reconstruct data point from linear combination of its neighbouring points. Similar to LLE, Laplacian Eigenmaps [74] preserve the local structure by computing low-dimensional subspace, in which the pairwise distance between a datapoint and its neighbours is minimal. Similarly, the LTSA [78] maps datapoints in high dimensional manifold to its local tangent space and there reconstruct the low dimensional representation of the manifold. All of the above methods are designed for a general manifold, thus approximating the geodesic distance without information of the specific manifold. The EEG covariance matrix lies in Riemannian manifold; therefore, more methods focused on dimensionality reduction are developed.

When considering the space of EEG matrices in Riemannian manifold, Xie et al. [78] proposed bilinear sub-manifold learning (BSML) that preserve the pairwise Riemannian geodesic distance between the data points instead of approximating the geodesic distance. Likewise, Horev et al. [55] extended PCA in Riemannian manifold by finding a matrix $W \in \mathbb{R}^{n \times p}$ that maps the data from the current Riemannian space to a lower dimension Riemannian space while maximising variance. Along the same context, Davoudi et al. [79] proposed a nonlinear dimensionality reduction methods that preserves the distances to the local mean (DPLM) and takes the geometry of the symmetric positive definite manifold into account. Tanaka et al. [80] proposed creating a graph that contains the location electrodes

and their respective signals, and later applies the graph Fourier transform (GFT) to reduce the dimensions.

2.7. Classification

Classification is the mapping of the feature space ($\mathbf{Z} \in \mathbb{R}^d$) into the target space ($y \in TargetSpace$). This mapping is usually created by three things: a mapping function $f \in FunctionSpace$, an objective function $J(\mathbf{w})$, and a minimization/maximization algorithm (iterative or by direct calculation). Each of these has a role in the classifications process. The mapping function f determines both the space at that is being worked on and the approximation abilities of the classifier, whereas the objective function $J(\mathbf{w})$ describes the problem that the classifier aims to solve. Finally, the minimization/maximization algorithm aims at finding the best (optimal) mapping function $f : FeatureSpace \rightarrow TargetSpace$ that maps the data to its targets based on the objective function $J(\mathbf{w})$. The classification algorithms fall into Euclidean and Riemannian manifold based on how they interpret EEG feature space.

2.7.1. Euclidean Space Methods

Euclidean space \mathbb{R}^n is the space of all n -dimensional real number vectors. Most of the classification algorithms work in this space. One of such algorithms is Decision Trees (DT) [81]. DTs creates a tree structure where each node $f(x)$ (shown in Table 2) is a piecewise function that outputs a *child* based on a feature x_i and threshold c . Both the feature x_i and the threshold c are determined by maximising (i.e., greedy algorithm) an objective function (e.g., gain impurity or information gain). This process is then repeated for each *child* output. If an output *child* does not improve the objective function, the node $f(x)$ outputs a class $\in \{1, -1\}$ instead.

Linear discriminant analysis (LDA) [82] is an algorithm that creates a projection vector \mathbf{w} that maximises the distance between classes S_B and minimizes the variance within a class S_W ($J_{LDA}(\mathbf{w}) = \max_{\mathbf{w} \in \mathbb{R}^n} \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}$). This is done by finding a generalized eigenvector of $S_B \mathbf{w} = \lambda S_W \mathbf{w}$. The classification is achieved by finding a threshold c that separates both classes, such as, if the dot product is below the threshold c , it belongs to class 1; otherwise it belongs to class 2. Duda et al. [83] described extension of LDA for multi-class problem.

The support vector machine (SVM) is another classification algorithm that works in the Euclidean space [82]. We later discuss the extension of this algorithm into the Riemannian manifold. SVM works by projecting the data points $\{x_i\}_{i=1}^M$ onto a hyperplane $\mathcal{H} \{\phi(x_i)\}_{i=1}^M$. A plane in the hyperplane \mathcal{H} is then created by solving the objective function (shown in Table 2) subject to $\alpha_i \geq 0$ and $\sum_i \alpha_i y_i = 0$ using quadratic programming where $\langle, \rangle_{\mathcal{H}}$ is the dot product in hyperplane \mathcal{H} . This plane is then used to distinguish between classes $f_{SVM}(x) = \text{sgn}(b + \sum_i y_i \alpha_i k(x, x_i))$, where $k(x, x_i) = \langle \phi(x_i), \phi(x) \rangle_{\mathcal{H}}$ is the hyperplane kernel. Different kernels exist for hyperplanes, such as the linear kernel $k(x, x_i) = x^T x_i$, the polynomial kernel $k(x, x_i) = (x^T x_i + c)^d$, where c is a constant, and the exponential kernel $k(x, x_i) = \exp(-\gamma \|x - x_i\|^2)$.

While DT, LDA and SVM have limited approximation abilities, multilayer perceptron (MLP) has no limits, as it is a universal approximate function. MLP $f(x) = \sum_i w_i^{(2)} \psi_i^{(1)}(\sum_j w_j^{(1)} x_j + b)$, as the name suggests, is a multilayer algorithms with each layer containing perceptrons that can fire $\psi(\cdot)$. The layers are connected by weights w that are trained using a minimization algorithm, such as stochastic gradient descent (SGD) or Adam algorithm. A convolutional neural network (CNN) is an extension to MLP. It extends the MLP algorithm by adding a convolution and pooling layers. In the convolution layer, the high-level information is extracted by using a matrix kernel that is applied to each part of the data matrix. While in the pooling layer, it extracts dominant features and decreases the computational power that is required to process the data by finding the maximum or average of the sub-matrices.

2.7.2. Riemannian Space Methods

A Riemannian manifold is created when the EEG data are taken and converted into sample covariance matrices (SCM). This Riemannian manifold differs for the Euclidean space. For example, a metric for measuring distances between two points in the Riemannian manifold is not equivalent to its Euclidean counterpart. The minimum distance to Riemannian mean (MDRM) is the most popular classification algorithm in the Riemannian manifold [34]. MDRM is the extension of the Euclidean classification algorithm in the Riemannian manifold. This algorithm take in the data in the form of sample covariance matrices (SCM) and then calculates the Riemannian mean $\sigma(\mathbf{P}_1, \dots, \mathbf{P}_m) = \arg \min_{\mathbf{P} \in P(n)} \sum_{i=1}^m \delta_R^2(\mathbf{P}, \mathbf{P}_i)$ for each class using it to label data where $\delta_R(\mathbf{P}_1, \mathbf{P}_2) = \|\text{Log}(\mathbf{P}_1^{-1} \mathbf{P}_2)\|_F = \left[\sum_{i=1}^n \log^2 \lambda_i \right]^{1/2}$ is the Riemannian distance. The Riemannian mean equation could be thought of as its objective functions $J(\mathbf{P})$, while the algorithm that is used to find it could be conceptualised as a minimisation algorithm. MDRM has the following mapping function:

$$f_{MDRM}(\mathbf{P}_{m+1}) = \arg \min_{j \in \{1, 2, \dots\}} \delta_R(\mathbf{P}_{m+1}, \mathbf{P}_{\Omega_j})$$

where \mathbf{P}_{Ω_j} is the mean of class j . Similarly, Riemannian SVM (R-SVM) [34], is the natural extension of SVM algorithm into the Riemannian manifold. It uses the tangent space of a reference matrix \mathbf{C}_{ref} as its hyper plane. This results in the following kernel:

$$k_R(\text{vect}(\mathbf{C}_i), \text{vect}(\mathbf{C}_j); \mathbf{C}_{\text{ref}}) = \langle \phi(\mathbf{C}_i), \phi(\mathbf{C}_j) \rangle_{\mathbf{C}_{\text{ref}}}$$

where $\text{vect}(\mathbf{C}) = [C_{1,1}; \sqrt{2}C_{1,2}; C_{2,2}; \sqrt{2}C_{1,3}; \sqrt{2}C_{2,3}; C_{3,3}; \dots; C_{E,E}]$ is the vectorized form of a symmetric matrix, $\phi(\mathbf{C}) = \text{Log}_{\mathbf{C}_{\text{ref}}}(\mathbf{C})$ is the map from the Riemannian manifold to the tangent space of \mathbf{C}_{ref} , and $\langle A, B \rangle_{\mathbf{C}} = \text{tr}(A\mathbf{C}^{-1}B\mathbf{C}^{-1})$ is the scalar product in the tangent space of \mathbf{C}_{ref} .

Table 2. This table provides a summary of the classification methods described in the Section 2.7.

	Mapping Function	Objective Function	Min/Max Algorithm
DT	$f(x) = \begin{cases} \text{child}_1 & x_i \leq c \\ \text{child}_2 & \text{otherwise} \end{cases}$	Gain impurity, information gain	greedy algorithm
LDA	$f(x) = \begin{cases} 1 & w^T x < c \\ -1 & \text{otherwise} \end{cases}$	$J(\mathbf{w}) = \max_{\mathbf{w} \in \mathbb{R}^n} \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}$	Eigen value solver
SVM R-SVM	$f(x) = \text{sgn}(b + \sum_i y_i \alpha_i k(x, x_i))$	$\max_{\alpha \in \mathbb{R}^m} \left(\sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \right)$	Quadratic Programming
MLP CNN	$f(x) = \sum_i w_i \psi_i(\cdot)$ $f(x) = \text{conv} + \text{pool} + \text{MLP}$	MSE, Cross entropy, Hinge	SGD, Adam
MDRM	$f(\mathbf{P}) = \arg \min_{j \in \{1, 2, \dots\}} \delta_R(\mathbf{P}, \mathbf{P}_{\Omega_j})$	$J(\mathbf{P}_{\Omega}) = \arg \min_{\mathbf{P}_{\Omega} \in P(n)} \sum_i \delta_R^2(\mathbf{P}_{\Omega}, \mathbf{P}_i)$	Averaging approaches

2.8. Performance Evaluation

The general architecture of motor-imagery based brain-computer interface is well understood, yet numerous novel MI based interfaces and strategies are proposed to enhance the performance of MI-BCI. Thus, performance evaluation metrics play an important role in quantifying diverse MI strategies. Accuracy is the most widely used performance evaluation, which measures the performance of algorithm in terms of correctly predicting target class trials. Accuracy metrics are mostly employed where the number of trials for all classes are equal and there is no bias towards a particular target class [84]. In the case of unbalanced (unequal number of trials) classes, Cohen's kappa coefficient is employed [85]. The Kappa coefficient equates an observed accuracy with respect to an expected accuracy (random chance). If kappa coefficient is 0, it means that there is no correlation with the

target class and predicted class, where, as kappa coefficient, 1 denotes perfect classification. If the MI classification is biased towards one class, then the confusion matrix (CM) is an important tool to quantify the performance of the system. Table 3 illustrates the confusion matrix for a multi-class problem. Metrics, like sensitivity and specificity, can be obtained from CM to identify the percentages of correctly classified trials from each MI class.

Table 3. Multi class Confusion matrix.

		Prediction			
		Class ₁	Class ₂	Class _k	Class _n
Target	Class ₁	D ₁₁	D ₁₂	D _{1k}	D _{1n}
	Class ₂	D ₂₁	D ₂₂	D _{2k}	D _{2n}
	Class _k	D _{k1}	D _{k2}	D _{kk}	D _{kn}
	Class _n	D _{n1}	D _{n2}	D _{n3}	D _{nn}

MI-BCI can be interpreted as a communication channel between user and machine, thus the information transfer rate (ITR) of each trial can be calculated in order to measure the bit-rate of the system. ITR can be obtained through CM (based on Accuracy) according to Wolpaw et al.'s [86] method as well as based on the performance and distribution of each MI classes [87]. The metrics discussed above are summarized in Table 4 and applicable for both synchronized and self-placed (asynchronous) as well as multi-class MI-BCIs. As a BCI can be defined as an encoder-decoder system where the user encodes information in EEG signals and the BCI decodes it into commands. The above metrics evaluate how well the BCI decode user's MI task into commands, but it does not quantify how well the user modulates EEG patterns with MI tasks [88]. Therefore, there is room for improving performance metrics that measure user MI skills or a user's encoding capability.

Lotte and Jeunet [88] have proposed stability and distinctiveness metrics to address some of the limitations mentioned above. Stability metrics measure how a stable MI EEG pattern is produced by a user. It is done by measuring the average distance between each MI task trial covariance matrix and mean covariance matrix for this MI task (left/right etc.). Distinct metrics measure the distinctiveness between MI task EEG patterns. Mathematically, distinct metrics is defined as the ratio of the between class variance to the within class variance. Stability and distinct metrics are both defined in the Riemannian manifold, as described in Table 4.

Table 4. Summary of all the Metrics.

Metrics	Two Class	Multi Class (N-Class)
Accuracy	$\frac{D_{11}+D_{22}}{N_{all}}$	$\frac{\sum_{i=1}^N D_{ii}}{N_{all}}$, where $N_{all} = \sum_{i,j=1}^N D_{ij}$
Kappa	$\frac{Accuracy - expected\ accuracy}{1 - expected\ accuracy}$	$expected\ accuracy(A_e) = \frac{\sum_{i=1}^N N}{D_i D_i}$
BCI decoding capability	sensitivity $\frac{D_{22}}{D_{21}+D_{22}}$	$\frac{D_{kk}}{N_k}$, where $N_k = \sum_{i=1}^N D_{k,i}$
ITR _{Wolpaw}	$ITR_{Wolpaw} = \log_2 N + Acc. \log_2(Acc) + (1 - Acc) \log_2(\frac{1-Acc}{N-1})$	
Stability	$\frac{1}{1+\sigma_{CA}}$	
User encoding capability	Distinct	$\frac{\sum_i \delta_r(C_i^A, \bar{C}^A)}{\sum_i \sigma_{CA_i}}$

3. Key Issues in MI Based BCI

MI based BCI still face multiple issues for it to be commercially usable. A usable MI based BCI should be plug and play, self paced, highly responsive, and consistent, so that that everybody can use it. This could be achieved by solving the following challenges:

3.1. Enhancement of MI-BCI Performance

A high performance MI-based BCI is important, as it increases the responsiveness of the device and prevents user frustration, hence improving the users experience. Improving

the performance could be achieved by improving its pre-processing stage, channel selection stage, feature selection stage, dimensionality reduction stage, or a combination of them.

3.1.1. Enhancement of MI-BCI Performance Using Preprocessing

Recent enhancements in the pre-processing step have revolved around two aspects: enhancing the incoming signal or enhancing the filtering of the signal. The former can be achieved by reconstructing the signal [89,90], enhancing the spatial resolution [91], or adding artificial noise [92]. In Casals et al. [89], they reconstructed corrupted EEG channels by using a tensor completion algorithm. The tensor completion algorithm applied a mask to this corrupted data in order to estimate it from observed EEG data. They found that this reconstructed the data of the corrupted channels and improved the classification performance in MI-BCI, whereas Gaur et al. [90] used multivariate empirical mode decomposition (MEMD) to decompose the EEG signal into a set of intrinsic mode functions (IMFs). Based on a median frequency measure, a set of IMFs is selected to reconstruct EEG signals. The CSP features are extracted from the reconstructed EEG signal for classification. One can enhance the spatial resolution of the EEG signal by using local activities estimation (LAE) method [91]. The LAE method estimates the recorded value of an EEG channel based on the weighted sum of local values of all EEG channels. The weights that are assigned to each channel for a weighted sum are based on the distance between channels. Similarly, enhancing the filtering of the signal can be achieved by automated filter (subject specific) tuning based on optimization algorithm like particle swarm optimization (PSO), artificial bee colony (ABC), and genetic algorithm (GA) [93]. Kim et al. [94] and Sun et al. [95] both proposed filters that are aimed to remove artifacts. Kim et al. [94] removed ocular artifacts by using an adaptive filtering algorithm that was based on ICA. Sun et al. [95] removed EOG artifacts by a contralateral channel normalization model that aims at extracting EOG artifacts from the EEG signal while retaining MI-related neural potential through finding the weights of EOG artifact interference with the EEG recordings. The Hijorth parameter was then extracted from the enhanced EEG signal for classification. In contrast to the above methods, Sampanna and Mitaim [92] have used the PSO algorithm to search for the optimal Gaussian noise intensity to be added in signals. This helps in achieving higher accuracy when compared to a conventionally filtered EEG signal. The Signal that is reliable at run time is very important for online evaluation of MI-BCI. To address this, Sagha et al. [96] proposed a method that quantifies electrode reliability at run time. They proposed two metrics that are based on Mahalanobis distance and information theory to detect anomalous behaviour of EEG electrodes.

3.1.2. Enhancement of MI-BCI Performance Using Channel Selection

Channel selection can both remove redundant and non-task relevant channels [97] and reduce power consumption of the device [98]. Removing channels can improve performance by reducing the search space [97], while reducing the power consumption can increase the longevity of a battery-based device [98]. Yang et al. [99] selected an optimal number of channels and time segments to extract features based on Fisher's discriminant analysis. They used the F score to measure discrimination power of time domain features obtained from different channels and different time segments. Jing et al. [100] selected high quality trials (free from artifacts) to find an optimal channel for a subject based on the "maximum projection on primary electrodes". These channels are used to calculate ICA filters for MI-BCI classification pipeline. This method has shown good improvement in classification accuracy even in session to session and subject to subject transfer MI-BCI scenarios. Park et al. [101] applied particle swarm optimization algorithm to find subject specific optimal number of electrodes. These electrodes' EEG data is further used for classification. Jin et al. [102] selected electrodes that contain more correlated information. To do this, they applied Z-score normalization to EEG signals from different channels, and then computed pearson's coefficients to measure the similarity between every pair of electrodes. From selected channels, RCSP features are extracted for SVM model based

classification. This significantly improves the accuracy compared to traditional methods. Yu et al. [103] used Fly optimization algorithm (FOA) to select the best channel for subject and then extracted CSP features from these channels for the classification. They also compared FOA performance with GA and PSO. Ramakrishnan and Satyanarayana [98] used a large (64) and small (19) number of channels in data acquisition for training and testing phase, respectively. They calculated inverse Karhunen Loeve Transform (KLT) matrix from training trials. This inverse KLT matrix is used to reconstruct all the missing channels in the testing phase. Masood et al. [104] employed various flavors of the CSP algorithm [31] to obtain the spatial filter weights of each electrode. Based on maximal values of spatial pattern coefficients, electrodes are selected to compute features for MI-CSP classification.

3.1.3. Enhancement of MI-BCI Performance Using Feature Selection

Similar to channel selection, feature selection improves the performance by finding the most optimal features. Similarly, Yang et al. [105], in their study, decomposed EEG signals from C3, Cz, and C4 channels into a series of overlapping time-frequency areas. They achieved this by cutting the filtered signals from filter bank of width 4 Hz and step 1 Hz (e.g., 8–12, 9–13, ..., 26–30) into multiple overlapping time segments. They used an F-score to select optimal time-frequency areas to extract features for MI-BCI classification. Rajan and Devassy [106] used a boosting approach that improved the classification by a combination of feature vectors. Baboukani et al. [107] used an Ant Colony Optimization technique to select a subset of features for SVM based classification of MI-BCI. Wang et al. [108] divided all of the electrodes in several sensor groups. From these sensor groups, CSP features are extracted to calculate EDRs. These EDRs are fused together, based on information fusion to obtain discriminate features for ensemble classification. Liu et al. [109] proposed a feature selection method that is based on the firefly algorithm and learning automata. These selected features are used to classify by a spectral regression discriminant analysis (SRDA) classifier. Kumar et al. [110] used the mutual information technique to extract suitable features from CSP features from filter banks. Samanta et al. [111] used an auto encoder-based deep feature extraction technique to extract meaningful features from images of a brain connectivity matrix. The brain connectivity matrix is constructed based on mutual correlation between different electrodes.

3.1.4. Enhancement of MI-BCI Performance Using Dimensionality Reduction

Xie et al. [112] learned low dimensional embedding on the Riemannian manifold based on prior information of EEG channels. Where, as, She et al. [113] extracted IMFs from EEG signals, and then employed Kernel spectral regression to reduce the dimension of IMFs. In doing so, they constructed a nearest neighbour graph to model the IMFs intrinsic structure. Özdenizci and Erdoğan [114] proposed the information theory based linear and non-linear feature transformation approach to select optimal feature for multi-class MI-EEG BCI system. Pei et al. [71] used stacked auto-encoders on spectral features to reduce the dimension and achieve high accuracy in a multi class asynchronous MI-BCI system. Razzak et al. [115] applied sparse PCA to reduce the dimensionality of features for SVM based classification. Horev et al. [55] extended the PCA to SPD manifold space, such that it preserved more variance in data while mapping SPD matrices to a lower dimension. Harandi et al. [116] proposed an algorithm that maintains the SPD matrices geometry while mapping it in a lower dimension. This is done by preserving the local structure's distance with respect to the local mean. In addition to it, this mapping minimizes the geodesic distance in samples that belong to the same class as well as maximizes the geodesic distance between samples belonging to a different class. Davoudi et al. [79] adapted Harandi's geometry preserving the dimensionality reduction technique in an unsupervised manner. Similarly, Tanaka et al. [80] proposed graph fourier transform for reducing dimensionality of SPD matrices through Tangent space mapping. This method has shown improvement in the performance for a small training dataset.

3.1.5. Enhancement of MI-BCI Performance with Combination of All

Li et al. [117] used the TPCT imaging method to fix the electrode positions and assigned time-frequency feature values to each pixel in the MI-EEG image. This way promotes feature fusion from the time, space, and frequency domains, respectively. These high dimensional images are fed to the modified VGG16 network [118]. Wang et al. [119] extracted a subset of channels from the motor imagery region. From these extracted channels, a subject-specific time window and frequency band is obtained to extract CSP features for classification. Sadiq et al. [120] manually selected the channels from the sensory motor cortex area of the brain. The EEG signal from these selected channels is decomposed into ten IMFs using adaptive empirical wavelet transform. The most sensitive mode out of ten is selected based on PSD and the Hilbert transform (HT) method extracts the instantaneous amplitude (IA) and instantaneous frequency (IF) from each channel. The statistical features are extracted from IF and IA components for classification. Selim et al. [121] used the bio-inspired algorithm (attractor metagene (AM)) to select the optimal time interval and CSP features for classification. Furthermore, they used the Bat optimization algorithm (BA) to optimize SVM parameters to enhance the classifier's performance. Athif and Ren [122] proposed the wave CSP technique that used wavelet transform and CSP filtering technique to enhance the signal to noise ratio of the EEG signal and to obtain key features for classification. Li et al. [123] optimized the spatial filter by employing Fisher's ration in objective function. This not only avoids using regularization parameters but also selects optimal features for classification. Li et al. [124] designed a spectral component CSP algorithm that utilized ICA to extract relevant motor information from EEG amplitude features that were obtained from CSP. Liu et al. [125] proposed an adaptive boosting algorithm that selects the most suitable EEG channels and frequency band for the CSP algorithm.

3.2. Reduce or Zero Calibration Time

Every day, a BCI user is required to go through a calibration phase for him/her to use BCI. This can be inconvenient, annoying, and frustrating. This section describes an on-going research solution to reduce the calibration phase or completely remove it. There are three categories of solutions: subject specific methods, transfer learning methods, and subject independent methods.

3.2.1. Subject-Specific Methods

Subject-specific methods for the reduction of calibration time mostly aim at more efficiently extracting features (i.e., with a small amount of training data). This can be achieved by the particle swarm optimization based learning strategy to find optimal parameters for spiking neural model (SNM) (deep learning model) [126]. This method automatically adjusts the parameters, removes the need for manual tuning, and increases the efficiency of SNM. However, this requires very subject-specific optimization of the parameters for best results [127]. Whereas, Zhao et al. [128] proposed the use of a framework that transforms EEG signals into three-dimensional space to preserve the temporal and spatial distribution of EEG signal and uses multi-branch 3D convolutional neural network to take advantage of temporal and spatial features in EEG signal. They showed that this approach significantly improves the accuracy under a small training dataset. Another approach of reducing calibration time is by a subject specific modification of the CSP algorithm. For example, Park and Chung [129] improved CSP by electing the CSP features from good local channels, rather than all channels. They selected good local channels that are based on the variance ratio dispersion score (VRDS) and inter-class feature distance (ICFD). Furthermore, they extended this approach in Filter Bank CSP by selecting good local channels for each frequency band, whereas Ma et al. [130] optimized SVM classifier's kernel and penalty parameters through a particle swarm optimization algorithm to obtain optimal CSP features. Furthermore, Costa et al. [131] proposed an adaptive CSP algorithm to overcome the limitation of CSP in short calibration sessions. They iteratively update the coefficients of

the CSP filters while using a recursive least squares (RLS) approach. This algorithm can be enhanced based on right channel selection and training free BCI system by modifying the algorithm with unsupervised techniques. Kee et al. [25] proposed Renyi entropy as a new alternative feature extraction method for small sample setting MI-BCI. Their method outperforms conventional CSP and regularized CSP design in small training datasets. Lotte and Guan [31] proposed Weighted Tikhonov Regularization for the CSP objective function that gives different penalties for different channels based on their degree of usefulness to classify a given mental state. They also extended the conventional CSP method for a small sample setting in [132] by penalizing the CSP objective function through prior information of EEG channels. Prior information of EEG channels was also used by Singh et al. [133] to obtain a smooth spatial filter in order to reduce the dimension of covariance matrices of trials under a small training set. They used MDRM for the classification of covariance matrices. This approach has shown higher performance under a high dimensional small sample setting.

3.2.2. Transfer Learning Methods

An investigation on inter-session and inter-subject variabilities in multi-class MI-based BCI revealed the feasibility of developing calibration-free BCIs in subjects sharing common sensorimotor dynamics [134]. Transfer learning methods have been developed based on this concept of using other subjects/sessions. Transfer learning methods aim to use other subjects data either to increase the amount of data that the classifier can be trained on or to regularize (prevent overfitting) the algorithm. The former can be seen in He and Wu [135], Hossain et al. [136], and Dai et al. [137]. He and Wu [135] used Euclidean-space alignment (EA) on the top of CSP to enable transfer learning from other subjects. EA projects all subjects into a similar distribution while using the Euclidean mean. Hossain et al. [136] extended FBCSP by adding selective informative instance transfer learning (SIITAL). The SIITAL trains the FBCSP with both source and target subjects by iteratively training the model and selecting the most relevant samples of the source subjects based on that model. Dai et al. [137] proposed unified cross-domain learning framework that uses the FBCSP method [138] to extract the features from source and target subjects. This is achieved by ensemble classifiers that are trained on misclassified samples and contribute to the overall model based on their classification accuracy, while the latter can be seen in Azab et al. [139], Singh et al. [140,141], Park and Lee [138], and Jiao et al. [142]. Azab et al. [139] proposed a logistic regression-based transfer learning approach that assigns different weights to a previously recorded session or source subject in order to represent similarities between these sessions/subjects features distribution and the new subject features distribution. Based on Kullback–Leibler divergence (KL) metrics, similar source/session feature space to target subject is chosen to obtain subject-specific common spatial patterns features for classification. Singh et al. [140,141] proposed a framework that takes advantage of both Euclidean and Riemannian approaches. They used a Euclidean subject to subject transfer approach to obtain optimized spatial filter for the target subject and employed Riemannian geometry-based classification to take advantage of the geometry of covariance matrices. Park and Lee [138] extended the FBCSP with regularization. They obtained an optimized spatial filter for each frequency band using information from other subjects' trials. The CSP features from each frequency band are obtained and, finally, based on mutual information most discriminate CSP features are selected for classification. Jiao et al. [142] proposed Sparse Group Representation Model for reducing the calibration time. In their work, they constructed a composite dictionary matrix with training samples from source subjects and target subject. A sparse representation-based model is then used to estimate the most compact representation of a target subject samples for classification by explicitly exploiting within-group sparse and group-wise sparse constraints in the dictionary matrix. The former has the advantage of being applicable to all the trained subjects over the latter.

3.2.3. Subject Independent Methods

Subject-independent methods aim to eliminate the calibration stage, allowing for the user to plug and play the BCI device. One way of achieving this is by projecting all different subjects/sessions' data into a unified space. Rodrigues et al. [143] proposed the Riemannian Procrustes Analysis as a projection based method. It transforms subject-specific data into a unified space by applying a sequence of geometrical transformations on their SCMs. These geometrical transformations aim to match the distribution of all subjects in high-dimensional space. These geometrically transformed SCMs are then fed to the MDRM classification model to discriminate the MI tasks. However, this method still requires the creation of the geometrical transformations that are based on the targets' session; thus, it is not entirely calibration-free, but it paves the way for fully subject independent MI-BCIs. Another way of achieving subject-independence is to create a universal map that can take in any subject data and output the command. Zhu et al. [144] proposed a deep learning framework for creating a universal neural network, called separate channel CNN (SCCN). SCCN contains three blocks: the CSP block, Encoder block, and recognition block. The CSP block was used to extract the temporal features from each channel. The encoder block then encodes those extracted features, followed by a concatenation of the encoded features and feeding them into the recognition block for classification. Joadder et al. [145] also proposed a universal MI-BCI map that extracts sub-band energy, fractal dimension, Log Variance, and Root Mean Square (RMS) features from spatial filtered EEG signal (CSP) for Linear Discriminant Analysis (LDA) classification model. They evaluated their design on a different time window after cue, different frequency band and different number of EEG channels and obtained good performance as compared to existing subject-dependent methods. Although both Zhu et al. [144] and Joadder et al. [145] classifiers are subject-independent, the CSP extracted features are not. Zhao et al. [146] hypothesized that there exists a universal CSP that is subject-independent. They used a multi-subject multi-subset approach where they took each subject in the training data and randomly picked samples to create multiple subsets and calculated a CSP on each subset. This was followed by a fitness evaluation-based distance between these CSP vectors (density and distance between highly dense vectors). They also proposed a semi-supervised approach as a classifier; however, unlike the universal CSP, it required unlabelled target data. In the same vein, Kwon et al. [147] followed the same universal CSP concept. Unlike Zhao et al. [146], they only trained one CSP on all of the available source subject's data and, since they had a larger dataset, they assumed that it would find the universal CSP. Mutual information and CNN was then used for a complete subject-independent algorithm.

3.3. BCI Illiteracy

BCI illiteracy subject is defined as the subject who cannot achieve a classification accuracy higher than 70% [11,148–153]. BCI illiteracy indicates that the user is unable to generate required oscillatory pattern during MI task. This leads to poor performance of MI-BCI. Some of the researchers focus on predicting whether a user falls under BCI illiterate category or not. This can help us to design a better algorithm for decoding MI or designing better training protocol to improve user skills. For instance, Ahn et al. [154] demonstrated that self-assessed motor imagery accuracy prediction has a positive correlation with actual performance. This can be valuable information to find BCI inefficiency in the user. While, Shu et al. [149], in their work, proposed two physiological variables, which is, laterality index (LI) and cortical activation strength (CAS), to predict MI-BCI performance prior to clinical BCI usage. Their proposed predictors exhibited a linear correlation with BCI performance, whereas Darvishi et al. [155] proposed a simple reaction time (SRT) as the BCI performance predictor. SRT is a metric that reflects the time that is required for a subject to respond to a defined stimulus. Their results indicate that SRT is correlated with BCI performance and BCI performance can be enhanced if the feedback interval is updated in accordance with the subject's SRT. In the same vein, Müller et al. [156] has theoretically shown that adaptation that is too fast may confuse the user, while an adaptation that is

too slow might not be able to track EEG variabilities due to learning. They created an online co-adaptation BCI system by ever-changing feedback according to the user and the system's learning. In the same vein, the co-adaptive approach to address BCI illiteracy has also been proposed by Acqualagna et al. [150]. Their paradigm was composed of two algorithms: a pre-trained subject independent classifier based on simple features, and a supervised subject optimized algorithm that can be modified to run in an unsupervised setting based manner. The approach of Acqualagna et al. is based on the classification of users put forth by Vidaurre et al. [157]. Vidaurre et al. [157], in their study, classified users in three categories: for category I users (Cat I), the classifier can be successfully trained and they gain good BCI control in the online feedback session. For Category II users (Cat II), the classifier can be successfully trained; however, good performance cannot be achieved in the feedback phase. For Category III users (Cat III), successful training of the classifier is not achieved. In the same vein, Lee et al. [158] found that that a universal BCI illiterate user does not exist (i.e., all of the participants were able to control at least one type of BCI system). Their study paves way to design a BCI system based on user's skill.

Another way of addressing BCI illiteracy problem is to design novel solutions that can improve performance, even in the case of BCI illiterate user. Similarly, Zhang et al. [153] addressed BCI illiteracy through a combination of CSP and brain network features. They constructed a task-related brain network by calculating the coherence between EEG channels, the graph-based analysis showed that the node degree and clustering coefficient have intensity differences between left and right-hand motor imagery. Their work suggests that there is a need to explore more feature extraction methods to address the BCI illiteracy problem. Furthermore, Yao et al. [148] proposed a hybrid BCI system to address the BCI inefficiency that is based on somatosensory attentional (SA) and motor imagery (MI) modalities. SA and MI are generated by attentional concentration intention (at some focused body part) and mentally simulating the kinesthetic movement, respectively. SA and MI are reflected through EEG signals at the somatosensory and motor cortices, respectively. In their work, they demonstrate that the combination of SA and MI would provide distinctive features to improve performance and increase the number of commands in a BCI system. In the same vein, Sannelli et al. [159] created an ensemble of adaptive spatial filters to increase BCI performance for BCI inefficient users. External factors can also improve BCI accuracy. For instance, Vidaurre et al. [160] proposed assistive peripheral electrical stimulation to modulate activity in the sensorimotor cortex. It is proposed that this will elicit short-term and long-term improvements in sensorimotor function, thus improving BCI illiteracy among users.

3.4. Asynchronised MI-BCI

MI-based BCI is usually trained in a synchronous manner, which is, there exists a sequence of instructions (or cue) that a user follows to produce an ERD/ERS phenomenon. However, in a real-world application, user want to execute control signal at his own will rather than waiting for cue. Therefore, there has been an increasing interest in creating an asynchronous MI. That is, MI-based BCI can detect that the user has an intention to undertake motor imagery, and then classifies MI task. This is done by splitting the incoming data into segments with overlapping periods. Each segment represents a potential MI command. One way of determining whether this potential MI command is an actual MI command is to build a classifier for that purpose. For example, the study of Yu et al. [161] presents the self-paced operation of a brain-computer interface (BCI), which can be voluntarily used to control a movement of a car (starting the engine, turning right, turning left, moving forward, moving backward, and stopping the engine). The system involved two classifiers: control intention classifier (CIC) and left/right classifier (LRC). The CIC is implemented in the first phase to identify the user intention being "idle" or "MI task-related". If an MI task-related is identified, a second phase follows the first phase by classifying it. Similarly, both Cheng et al. [162] and Antelis et al. [163] proposed a deep learning method that is trained to distinguish between resting state, transition state, and

execution state. However, Cheng proposed a convolutional neural network, followed by a fully connected network (CNN-FC), while Antelis proposed Dendrite morphological neural networks (DMNN). Another approach is to let the subject achieve a set number of consistent right/left classification within a set period for an action to be taken, thus confirming the command and avoiding randomness [164], both adding a classifier and classifying multiple times, adds computational time and complexity to the system, with the latter also adding the time required for classification. Sun et al. [165] suggested a method that avoids these constraints by using a threshold on an existing classifier that separates idle from MI task-related. He et al. [166] proposed a similar approach for continuous application, such as mouse movement. This is achieved through moving the object (in this case a mouse) by the confidence level of the classifier. The threshold-based method of addressing this challenge requires defining a threshold that could be difficult and user-dependent. This brings us to the last methodology of addressing this challenge, which is, by adding an idle class into the classifier [167–170]. All of the above-motivated methods, except the method proposed by Yousefi et al. [170], use a target-oriented paradigm where the user is asked to perform a task and the algorithm is evaluated based on the user's ability to achieve that task. However, Yousefi et al. [170] tested their algorithm by giving the user a specified time interval to perform any task the user desired and after the time has passed, the user provides feedback as to whether the algorithm responded to his commands. In conclusion, all of the algorithms can run asynchronously, given that they have a reasonable run time.

3.5. Increase Number of Commands

More diverse and complex applications, like spellers etcetera, can be developed with high ITR and increased number of classes in MI-BCI. Traditionally, MI-BCI was designed as binary class (left and right) problem. The first way to extend MI-BCI into multi-class is by employing a hybrid approach during which the MI paradigm is complemented with other mental strategies. For example, Yu et al. [171] proposed a hybrid asynchronous brain-computer interface that is based on sequential motor imagery (MI) and P300 potential to execute eleven control functions for wheelchairs. The second way to achieve multi-class MI-BCI is algorithmically. For example, the traditional CSP algorithm is extended to recognize four MI tasks [172]. In the similar manner, Wentrup and Buss [173] proposed information theoretic feature extraction frameworks for CSP algorithm by extending it for multiclass MI-BCI system. In the same vein, Christensen et al. [174] extended FBCSPs for five class MI-BCI system. Similarly, Razzak et al. [175] proposed a novel multiclass support matrix machine to handle multiclass MI imagery tasks. Likewise, Barachant et al. [176] presented a new classification method based on Riemannian geometry that uses covariance matrices to classify multi-class BCI. Faiz and Hamadani [177] controlled humanoid robotic hand gestures through five class online MI BCI while using a commercial EEG headset. They use AR and CSP feature extractions and PCA to reduce the dimension of AR features. Finally, CSP and AR features are concatenated and trained by a SVM classifier to achieve multi-class recognition.

3.6. Adaptive BCI

The consistency of the accuracy of the classifier during long sessions is one of the issues still being worked in EEG based MI-BCI. This is because EEG is a non-stationary signal that get impacted over time as well as when there is change in recording environment and state of mind (e.g., fatigue, attention, motivation, emotion, etc). Adaptive methods have been proposed to address this challenge. For instance, Aliakbaryhosseinabadi et al. [178] demonstrated that it is possible to detect a user's attention diversion during a MI task, whereas Dagaev et al. [179] extracted the target state (LH, RH) from background state (environment, emotional, and cognitive condition, etc.). This was achieved by asking subjects in the training stage to open and close their eyes during the trials. These instructions act as the two different background conditions. The methods that detect cause of change in user signals other than the MI task could pave the way for adaptive MI-BCI by giving both the

user real-time neurofeedback and giving the adaptive algorithm additional information to work with while decoding MI task.

Another way to address this challenge is to modify the training protocol or extracting more information during it. Mondini et al. [180] and Schwarz et al. [181] both modified the training protocol. By creating an adaptive training protocol, Mondini et al. [180] fulfilled three tasks: (a) adapt the training session based on the subject's ability, which is, make the training short and restart the training from the beginning with different motor imagery strategy if the system performance is lower than a certain threshold; (b) present training cue (left/right) in a biased manner that is present left cue more often if the left imagery performance is low when compared to the right; and, (c) keep challenging the performance of the user by only giving feedback if it exceeds an adaptive threshold. Schwarz et al. [181] proposed a co-adaptive online learning BCI model that uses the concept of semi-supervised retraining. The Schwarz model uses a few initial supervised calibration trials per MI tasks and then performs recurrent retraining by using artificially generated labels. This ensures feedback to the user after a very short training and engages the user in mutual learning with the system. Information gathered during training protocol, such as command delivery time (CDT) and the probability of the next action, could be used to address this challenge. Saeedi et al. used CDT [182] to provide a system that delivers adaptive assistance, which is, if the current trial is long, then the system will slow down to give enough time to the user to execute the MI tasks. Their study suggests that the brain pattern is different for short, long and time-out commands. They were able to differentiate between command type using only one second before the trial started, while Perdikis et al. [183] proposed using the probability of next action to adapt the classifier. Specifically, they implemented an online speller based on the BrainTree MI text-entry system that uses probabilistic contextual information to adapt an LDA classifier. The final method observed in the literature to address this challenge was to create an adaptive classifier. Faller et al. [184] proposed an online adaptive MI-BCI that auto-calibrates. Their system in regular interval not only discriminates features for classifier retraining, but also learns to reject outliers. Their system starts to provide feedback after minimal training and keeps improving by learning subject-specific parameters on the run. Raza et al. [185] proposed an unsupervised adaptive ensemble learning algorithm that tackles non-stationary based co-variate shifts between two BCI sessions. This algorithm paves the way for online adaption to variabilities between BCI sessions. In the same vein, Rong et al. [186] proposed an online method that handles the statistical difference between sessions. They used an adaptive fuzzy inference system.

3.7. Online MI-BCI

After an adaptive BCI, the BCI mode is one key factor that determines MI based system's usability and efficacy. MI-BCI systems are operated in offline or online mode through cue-based paradigms, where self-placed (asynchronous) are mostly online systems. Mostly, the literature proposed improvements in offline mode of MI-BCI systems; very few test their proposed algorithms in the online environment. In online BCI studies, Sharghian et al. [187] proposed MI-EEG, which uses sparse representation-based classification (SRC). Their approach obtains an online dictionary learning scheme from the extracted band power from a spatial-filtered signal. This dictionary leads to reconstruction of sparse signal for classification. In the same vein, Zhang et al. [188] proposed an incremental linear discriminant analysis algorithm that extract AR features from preferable incoming data. Their method paved way for fully auto-calibrating an online MI-BCI system. Similarly, Yu et al. [167] proposed an asynchronous MI BCI system to control wheelchair navigation. Perez [189] extended the fuzzy logic framework for adaptive online MI-BCI system and evaluated it through the realistic navigation of a bipedal robot. Ang and Guan [190] introduced an adaptive strategy that continuously computes the subject-specific model during an online phase. Abdalsalam et al. [191] controlled the screen cursor through a four class MI-BCI system. Their results suggest that online feedback increases ERDs over the mu (8–10 Hz) and upper beta (18–24 Hz) band, which results in a higher cursor control success

rate. Many studies have demonstrated the efficiency of virtual reality (VR) and gaming environment in a online BCI [192]. Achancaray et al. [193] in the same vein, verified that virtual reality based online feedback has positive effects on the subject. It has been observed that motor cortex increases its activation level (in alpha and beta band) due to an immersive VR experience. This is very helpful in supporting upper limb rehabilitation of post-stroke patients. Similarly, Alchalabi and Faubert [194] used VR based neurofeedback in the Online MI-BCI session. Cubero et al. [195] proposed an online system that is based on an endless running game that runs on three class MI-BCI. They used graphic representation of EEG signals for multi-resolution analysis to take advantage of spatial dimension, along with temporal and spectral dimensions.

3.8. Training Protocol

Similar to other normal user skills, BCI control is also a skill that can be learned and improved with proper training. A typical BCI training protocol is a combination of user instructions, cues on screen to modulate the user's neural activity in a specific manner, and, lastly, a feedback mechanism that represents confidence of the classifier in recognition of the mental task to user. Unfortunately, standard training protocol does not satisfy the psychology of human learning; usually being boring and very long. Meng and He [196] studied the effect of MI training on users. They found out that, with a few hours of MI training, there is change in electrophysiological properties. Their study suggested design engaging training protocol and multiple training sessions, rather than a long training session for low BCI performers. In the same vein, Kim et al. [197] proposed a self placed training protocol, in which the user performs MI task continuously without an inter-stimulus-interval. During each trial, the user has to imagine a single MI task (e.g., RH for 60 s). The results of this protocol showed that it reduces the calibration time when compared to conventional MI training protocol. Jeunet et al. [198] surveyed the cognitive and psychological factors that are related to MI-BCI and categorized these factors into three categories (a) user-technology relationship, (b) attention, and (c) spatial abilities. Their work is very useful for designing a new training protocol that takes advantage of these factors. Furthermore, in another study, Jeunet et al. [11] found that spatial ability plays an important role in BCI performance of a subject. They suggested having pre-training sessions to explore spatial ability for BCI training.

Many studies proposed new training strategies that use other mental strategies to compliment MI training (kinesthetic imagination of limbs). For instance, Zhang et al. [199] proposed a new BCI training paradigm that combines conventional MI training protocol with covert verb reading. This improves the performance of MI-BCI and paves the way for utilizing semantic processing with motor imagery. Along the same lines, Wang et al. [200] proposed a hybrid MI-paradigm that uses speech imagery with motor imagery. In this paradigm, the user repeatedly and silently reads move (left/right) cues during imagination. Standard training protocols are fixed that are not tailored made to user's need and experience. With respect to this, Wang et al. [201] proposed MI training with visual-haptic neurofeedback. Their findings validate that their approach improves cortical activations at the sensorimotor area, thus leading to an improvement in BCI performance. Liburkina et al. [202] proposed a MI training protocol that gives cue to perform and feedback to the user through vibration. Along the same lines, Pillette et al. [203] designed an intelligent tutoring system that provides support during MI training and enhance user experience/performance on MI-BCI system. Skola et al. [204] proposed a virtual reality-based MI-BCI training that uses a virtual avatar to provide feedback. Their training helps in maintaining high levels of attention and motivation. Furthermore, their proposed method improves the BCI skills of first time users.

4. Conclusions

In this paper, we have provided an extensive review of methodologies for designing an MI-BCI system. In doing so, we have created a generic framework and mapped literature

related to different components (data acquisition, MI training, preprocessing, feature extraction, channel and feature selection, classification, and performance metrics) in it. This will help in visualizing gaps to be filled by future studies in order to further improve BCI usability.

Despite many outstanding developments in MI-BCI research, some critical issues still need to be resolved. Mostly, studies are on synchronized MI-BCI in offline mode. There is a need to have more studies on online BCI. Typically, researchers use performance evaluation metrics, as per their convenience. It would be better to have general BCI standards that can be widely adhered by researchers. Our literature survey found that enhancing the performance is still a critical issue even after two decades of research. Due to availability of high computational resources, present studies employ methods based on deep learning and Riemannian geometry more than traditional machine learning methods. With current advancement in algorithms, future research should concentrate more on eliminating or reducing long calibration in MI-BCI. Future studies should focus on more diverse BCI applications that can be developed with increased number of commands. Our review shows that BCI illiteracy is a critical issue that can be addressed either by using better training protocol that suit users' requirements or through smart algorithms. Finally, EEG is a non-stationary signal that changes over time as user's state of mind changes. This causes inconsistency in the BCI classifier's performance; thus, it is important to make progress in development of adaptive methods in order to address this challenge in an online settings.

Author Contributions: Conceptualization, A.S. and A.A.H.; methodology, A.S. and A.A.H.; validation, A.S., A.A.H., S.L. and H.W.G.; formal analysis, A.S. and A.A.H.; investigation, A.S. and A.A.H.; resources, A.A.H.; data curation, A.S. and A.A.H.; writing—original draft preparation, A.S. and A.A.H.; writing—review and editing, A.S., A.A.H., S.L. and H.W.G.; visualization, A.S.; supervision, S.L. and H.W.G.; project administration, S.L. and H.W.G.; funding acquisition, S.L. and H.W.G.; First and second author has contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: School of Fundamental Sciences, Massey University provided funding to cover the publication cost.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 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. [\[CrossRef\]](#)
2. Bashashati, H.; Ward, R.K.; Bashashati, A.; Mohamed, A. Neural Network Conditional Random Fields for Self-Paced Brain Computer Interfaces. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; pp. 939–943. [\[CrossRef\]](#)
3. Rashid, M.; Sulaiman, N.; Abdul Majeed, A.P.P.; Musa, R.M.; Ab. Nasir, A.F.; Bari, B.S.; Khatun, S. Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review. *Front. Neurobot.* **2020**, *14*, 25. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Ramadan, R.A.; Vasilakos, A.V. Brain computer interface: Control signals review. *Neurocomputing* **2017**, *223*, 26–44. [\[CrossRef\]](#)
5. Martini, M.L.; Oermann, E.K.; Opie, N.L.; Panov, F.; Oxley, T.; Yaeger, K. Sensor Modalities for Brain-Computer Interface Technology: A Comprehensive Literature Review. *Neurosurgery* **2020**, *86*, E108–E117. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Bucci, P.; Galderisi, S. Physiologic Basis of the EEG Signal. In *Standard Electroencephalography in Clinical Psychiatry*; John Wiley and Sons, Ltd.: Hoboken, NJ, USA, 2011; Chapter 2, pp. 7–12. [\[CrossRef\]](#)
7. Farnsworth, B. *EEG (Electroencephalography): The Complete Pocket Guide*; IMotions, Global HQ: Copenhagen, Denmark, 2019.
8. Pfurtscheller, G.; Neuper, C. Motor imagery and direct brain-computer communication. *Proc. IEEE* **2001**, *89*, 1123–1134. [\[CrossRef\]](#)
9. Otaiby, T.; Abd El-Samie, F.; Alshebeili, S.; Ahmad, I. A review of channel selection algorithms for EEG signal processing. *EURASIP J. Adv. Signal Process.* **2015**, *2015*, 1–21. [\[CrossRef\]](#)

10. Wan, X.; Zhang, K.; Ramkumar, S.; Deny, J.; Emayavaramban, G.; Siva Ramkumar, M.; Hussein, A.F. A Review on Electroencephalogram Based Brain Computer Interface for Elderly Disabled. *IEEE Access* **2019**, *7*, 36380–36387. [[CrossRef](#)]
11. Jeunet, C.; Jahanpour, E.; Lotte, F. Why standard brain-computer interface (BCI) training protocols should be changed: An experimental study. *J. Neural Eng.* **2016**, *13*. [[CrossRef](#)]
12. McCreadie, K.A.; Coyle, D.H.; Prasad, G. Is Sensorimotor BCI Performance Influenced Differently by Mono, Stereo, or 3-D Auditory Feedback? *IEEE Trans. Neural Syst. Rehabil. Eng.* **2014**, *22*, 431–440. [[CrossRef](#)]
13. Cincotti, F.; Kauhanen, L.; Aloise, F.; Palomäki, T.; Caporusso, N.; Jylänki, P.; Mattia, D.; Babiloni, F.; Vanacker, G.; Nuttin, M.; et al. Vibrotactile Feedback for Brain-Computer Interface Operation. *Comput. Intell. Neurosci.* **2007**, *2007*, 48937. [[CrossRef](#)]
14. Lotte, F.; Faller, J.; Guger, C.; Renard, Y.; Pfurtscheller, G.; Lécuyer, A.; Leeb, R. Combining BCI with Virtual Reality: Towards New Applications and Improved BCI. In *Towards Practical Brain-Computer Interfaces*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 197–220. [[CrossRef](#)]
15. Leeb, R.; Lee, F.; Keinrath, C.; Scherer, R.; Bischof, H.; Pfurtscheller, G. Brain-Computer Communication: Motivation, Aim, and Impact of Exploring a Virtual Apartment. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2007**, *15*, 473–482. [[CrossRef](#)]
16. Islam, M.K.; Rastegarnia, A.; Yang, Z. Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiol. Clin. Neurophysiol.* **2016**, *46*, 287–305. [[CrossRef](#)]
17. Uribe, L.F.S.; Filho, C.A.S.; de Oliveira, V.A.; da Silva Costa, T.B.; Rodrigues, P.G.; Soriano, D.C.; Bocato, L.; Castellano, G.; Attux, R. A correntropy-based classifier for motor imagery brain-computer interfaces. *Biomed. Phys. Eng. Express* **2019**, *5*, 065026. [[CrossRef](#)]
18. Xu, B.; Zhang, L.; Song, A.; Wu, C.; Li, W.; Zhang, D.; Xu, G.; Li, H.; Zeng, H. Wavelet Transform Time-Frequency Image and Convolutional Network-Based Motor Imagery EEG Classification. *IEEE Access* **2019**, *7*, 6084–6093. [[CrossRef](#)]
19. Samuel, O.W.; Geng, Y.; Li, X.; Li, G. Towards Efficient Decoding of Multiple Classes of Motor Imagery Limb Movements Based on EEG Spectral and Time Domain Descriptors. *J. Med. Syst.* **2017**, *41*, 194. [[CrossRef](#)] [[PubMed](#)]
20. Hamedi, M.; Salleh, S.; Noor, A.M.; Mohammad-Rezazadeh, I. Neural network-based three-class motor imagery classification using time-domain features for BCI applications. In *Proceedings of the 2014 IEEE REGION 10 SYMPOSIUM*, Kuala Lumpur, Malaysia, 14–16 April 2014; pp. 204–207. [[CrossRef](#)]
21. Rodríguez-Bermúdez, G.; García-Laencina, P.J. Automatic and Adaptive Classification of Electroencephalographic Signals for Brain Computer Interfaces. *J. Med. Syst.* **2012**, *36*, 51–63. [[CrossRef](#)] [[PubMed](#)]
22. Güçlü, U.; Güçlütürk, Y.; Loo, C.K. Evaluation of fractal dimension estimation methods for feature extraction in motor imagery based brain computer interface. *Procedia Comput. Sci.* **2011**, *3*, 589–594. [[CrossRef](#)]
23. Adam, A.; Ibrahim, Z.; Mokhtar, N.; Shapiai, M.I.; Cumming, P.; Mubin, M. Evaluation of different time domain peak models using extreme learning machine-based peak detection for EEG signal. *SpringerPlus* **2016**, *5*, 1–14. [[CrossRef](#)]
24. Yilmaz, C.M.; Kose, C.; Hatipoglu, B. A Quasi-probabilistic distribution model for EEG Signal classification by using 2-D signal representation. *Comput. Methods Programs Biomed.* **2018**, *162*, 187–196. [[CrossRef](#)]
25. Kee, C.Y.; Ponnambalam, S.G.; Loo, C.K. Binary and multi-class motor imagery using Renyi entropy for feature extraction. *Neural Comput. Appl.* **2016**, *28*, 2051–2062. [[CrossRef](#)]
26. Chen, S.; Luo, Z.; Gan, H. An entropy fusion method for feature extraction of EEG. *Neural Comput. Appl.* **2016**, *29*, 857–863. [[CrossRef](#)]
27. Batres-Mendoza, P.; Montoro-Sanjose, C.R.; Guerra-Hernandez, E.I.; Almanza-Ojeda, D.L.; Rostro-Gonzalez, H.; Romero-Troncoso, R.J.; Ibarra-Manzano, M.A. Quaternion-Based Signal Analysis for Motor Imagery Classification from Electroencephalographic Signals. *Sensors* **2016**, *16*, 336. [[CrossRef](#)] [[PubMed](#)]
28. Aggarwal, S.; Chugh, N. Signal processing techniques for motor imagery brain computer interface: A review. *Array* **2019**, *1*, 100003. [[CrossRef](#)]
29. Gao, Z.; Wang, Z.; Ma, C.; Dang, W.; Zhang, K. A Wavelet Time-Frequency Representation Based Complex Network Method for Characterizing Brain Activities Underlying Motor Imagery Signals. *IEEE Access* **2018**, *6*, 65796–65802. [[CrossRef](#)]
30. Ortiz, M.; Iáñez, E.; Contreras-Vidal, J.L.; Azorín, J.M. Analysis of the EEG Rhythms Based on the Empirical Mode Decomposition During Motor Imagery When Using a Lower-Limb Exoskeleton. A Case Study. *Front. Neurobot.* **2020**, *14*, 48. [[CrossRef](#)]
31. Lotte, F.; Guan, C. Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 355–362. [[CrossRef](#)]
32. Li, M.; Zhang, C.; Jia, S.; Sun, Y. Classification of Motor Imagery Tasks in Source Domain. In *Proceedings of the 2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, Changchun, China, 5–8 August 2018; pp. 83–88. [[CrossRef](#)]
33. Rejer, I.; Górski, P. EEG Classification for MI-BCI with Independent Component Analysis. In *Proceedings of the 10th International Conference on Computer Recognition Systems CORES 2017*; Kurzynski, M., Wozniak, M., Burduk, R., Eds.; Springer: Cham, Switzerland, 2018; pp. 393–402.
34. Barachant, A.; Bonnet, S.; Congedo, M.; Jutten, C. Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing* **2013**, *112*, 172–178. [[CrossRef](#)]
35. Suma, D.; Meng, J.; Edelman, B.J.; He, B. Spatial-temporal aspects of continuous EEG-based neurobotic control. *J. Neural Eng.* **2020**, *17*, 066006. [[CrossRef](#)]
36. Stefano Filho, C.A.; Attux, R.; Castellano, G. Can graph metrics be used for EEG-BCIs based on hand motor imagery? *Biomed. Signal Process. Control* **2018**, *40*, 359–365. [[CrossRef](#)]

37. Schirrmester, R.T.; Springenberg, J.T.; Fiederer, L.D.J.; Glasstetter, M.; Eggensperger, K.; Tangermann, M.; Hutter, F.; Burgard, W.; Ball, T. Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* **2017**, *38*, 5391–5420. [[CrossRef](#)] [[PubMed](#)]
38. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng.* **2018**. [[CrossRef](#)]
39. Wang, P.; Jiang, A.; Liu, X.; Shang, J.; Zhang, L. LSTM-Based EEG Classification in Motor Imagery Tasks. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 2086–2095. [[CrossRef](#)] [[PubMed](#)]
40. Rashkov, G.; Bobe, A.; Fastovets, D.; Komarova, M. Natural image reconstruction from brain waves: A novel visual BCI system with native feedback. *bioRxiv* **2019**, 1–15. [[CrossRef](#)]
41. Virgilio G., C.D.; Sossa A., J.H.; Antelis, J.M.; Falcón, L.E. Spiking Neural Networks applied to the classification of motor tasks in EEG signals. *Neural Netw.* **2020**, *122*, 130–143. [[CrossRef](#)]
42. Lee, S.B.; Kim, H.J.; Kim, H.; Jeong, J.H.; Lee, S.W.; Kim, D.J. Comparative analysis of features extracted from EEG spatial, spectral and temporal domains for binary and multiclass motor imagery classification. *Inf. Sci.* **2019**, *502*, 190–200. [[CrossRef](#)]
43. Chu, Y.; Zhao, X.; Zou, Y.; Xu, W.; Han, J.; Zhao, Y. A Decoding Scheme for Incomplete Motor Imagery EEG With Deep Belief Network. *Front. Neurosci.* **2018**, *12*, 680. [[CrossRef](#)]
44. Zhang, R.; Xu, P.; Chen, R.; Li, F.; Guo, L.; Li, P.; Zhang, T.; Yao, D. Predicting Inter-session Performance of SMR-Based Brain-Computer Interface Using the Spectral Entropy of Resting-State EEG. *Brain Topogr.* **2015**, *28*, 680–690. [[CrossRef](#)]
45. Guo, X.; Wu, X.; Zhang, D. Motor imagery EEG detection by empirical mode decomposition. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–8 June 2008; pp. 2619–2622. [[CrossRef](#)]
46. Ortiz-Echeverri, C.J.; Salazar-Colores, S.; Rodríguez-Reséndiz, J.; Gómez-Loenzo, R.A. A New Approach for Motor Imagery Classification Based on Sorted Blind Source Separation, Continuous Wavelet Transform, and Convolutional Neural Network. *Sensors* **2019**, *19*, 4541. [[CrossRef](#)]
47. Ang, K.K.; Chin, Z.Y.; Zhang, H.; Guan, C. Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–8 June 2008; pp. 2390–2397. [[CrossRef](#)]
48. Thomas, K.P.; Guan, C.; Lau, C.T.; Vinod, A.; Ang, K. A New Discriminative Common Spatial Pattern Method for Motor Imagery Brain-Computer Interface. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 2730–2733. [[CrossRef](#)]
49. Li, Y.; Zhang, X.; Zhang, B.; Lei, M.; Cui, W.; Guo, Y. A Channel-Projection Mixed-Scale Convolutional Neural Network for Motor Imagery EEG Decoding. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1170–1180. [[CrossRef](#)]
50. Yang, J.; Yao, S.; Wang, J. Deep Fusion Feature Learning Network for MI-EEG Classification. *IEEE Access* **2018**, *6*, 79050–79059. [[CrossRef](#)]
51. Wu, W.; Gao, X.; Hong, B.; Gao, S. Classifying Single-Trial EEG During Motor Imagery by Iterative Spatio-Spectral Patterns Learning (ISSPL). *IEEE Trans. Biomed. Eng.* **2008**, *55*, 1733–1743. [[CrossRef](#)] [[PubMed](#)]
52. Suk, H.; Lee, S. A probabilistic approach to spatio-spectral filters optimization in Brain-Computer Interface. In Proceedings of the 2011 IEEE International Conference on Systems, Man, and Cybernetics, Anchorage, AK, USA, 9–12 October 2011; pp. 19–24. [[CrossRef](#)]
53. Zhang, P.; Wang, X.; Zhang, W.; Chen, J. Learning Spatial-Spectral-Temporal EEG Features with Recurrent 3D Convolutional Neural Networks for Cross-Task Mental Workload Assessment. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 31–42. [[CrossRef](#)] [[PubMed](#)]
54. Bang, J.S.; Lee, M.H.; Fazli, S.; Guan, C.; Lee, S.W. Spatio-Spectral Feature Representation for Motor Imagery Classification Using Convolutional Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–12. [[CrossRef](#)] [[PubMed](#)]
55. Horev, I.; Yger, F.; Sugiyama, M. Geometry-aware principal component analysis for symmetric positive definite matrices. *Mach. Learn.* **2017**, *106*, 493–522. [[CrossRef](#)]
56. Venkatesh, B.; Anuradha, J. A Review of Feature Selection and Its Methods. *Cybern. Inf. Technol.* **2019**, *19*, 3–26. [[CrossRef](#)]
57. Battiti, R. Using Mutual Information for Selecting Features in Supervised Neural Net Learning. *IEEE Trans. Neural Netw.* **1994**, *5*, 537–550. [[CrossRef](#)]
58. Homri, I.; Yacoub, S. A hybrid cascade method for EEG classification. *Pattern Anal. Appl.* **2019**, *22*, 1505–1516. [[CrossRef](#)]
59. Ramos, A.C.; Hernandex, R.G.; Vellasco, M. Feature Selection Methods Applied to Motor Imagery Task Classification. In Proceedings of the 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Cartagena, Colombia, 2–4 November 2016. [[CrossRef](#)]
60. Kashef, S.; Nezamabadi-pour, H.; Nikpour, B. Multilabel feature selection: A comprehensive review and guiding experiments. *WIREs Data Min. Knowl. Discov.* **2018**, *8*, e1240. [[CrossRef](#)]
61. Atyabi, A.; Shic, F.; Naples, A. Mixture of autoregressive modeling orders and its implication on single trial EEG classification. *Expert Syst. Appl.* **2016**, *65*, 164–180. [[CrossRef](#)]
62. Baig, M.Z.; Aslam, N.; Shum, H.P.; Zhang, L. Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG. *Expert Syst. Appl.* **2017**, *90*, 184–195. [[CrossRef](#)]
63. Das, S.; Suganthan, P.N. Differential evolution: A survey of the state-of-the-art. *IEEE Trans. Evol. Comput.* **2011**, *15*, 4–31. [[CrossRef](#)]

64. Karaboga, D. *An Idea Based on Honey Bee Swarm for Numerical Optimization*; Technical Report-tr06; Erciyes University, Engineering Faculty, Computer Engineering Department: Kayseri, Turkey, 2005; Volume 200, pp. 1–10.
65. Rakshit, P.; Bhattacharyya, S.; Konar, A.; Khasnobish, A.; Tibarewala, D.; Janarthanan, R. Artificial Bee Colony Based Feature Selection for Motor Imagery EEG Data. In *Proceedings of the Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)*; Springer: New Delhi, India, 2012; pp. 127–138. [[CrossRef](#)]
66. van der Maaten, L.; Postma, E.; Herik, H. Dimensionality Reduction: A Comparative Review. *J. Mach. Learn. Res. JMLR* **2007**, *10*, 13.
67. Gupta, A.; Agrawal, R.K.; Kaur, B. Performance enhancement of mental task classification using EEG signal: A study of multivariate feature selection methods. *Soft Comput.* **2015**, *19*, 2799–2812. [[CrossRef](#)]
68. Jusas, V.; Samuvel, S.G. Classification of Motor Imagery Using a Combination of User-Specific Band and Subject-Specific Band for Brain-Computer Interface. *Appl. Sci.* **2019**, *9*, 4990. [[CrossRef](#)]
69. Ayesha, S.; Hanif, M.K.; Talib, R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Inf. Fusion* **2020**, *59*, 44–58. [[CrossRef](#)]
70. Scholkopf, B.; Smola, A.; Muller, K.R. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Comput.* **1996**, *10*, 1299–1319. [[CrossRef](#)]
71. Pei, D.; Burns, M.; Chandramouli, R.; Vinjamuri, R. Decoding Asynchronous Reaching in Electroencephalography Using Stacked Autoencoders. *IEEE Access* **2018**, *6*, 52889–52898. [[CrossRef](#)]
72. Tenenbaum, J.; Silva, V.; Langford, J. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* **2000**, *290*, 2319–2323. [[CrossRef](#)]
73. Iturralde, P.; Patrone, M.; Lecumberry, F.; Fernández, A. Motor Intention Recognition in EEG: In Pursuit of a Relevant Feature Set. In *Proceedings of the Pattern Recognition, Image Analysis, Computer Vision, and Applications, Buenos Aires, Argentina, 3–6 September 2012*; pp. 551–558. [[CrossRef](#)]
74. Gramfort, A.; Clerc, M. Low Dimensional Representations of MEG/EEG Data Using Laplacian Eigenmaps. In *Proceedings of the 2007 Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging, Hangzhou, China, 12–14 October 2007*; pp. 169–172. [[CrossRef](#)]
75. Lafon, S.; Lee, A.B. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1393–1403. [[CrossRef](#)]
76. Lee, F.; Scherer, R.; Leeb, R.; Schlögl, A.; Bischof, H.; Pfurtscheller, G. Feature Mapping using PCA, Locally Linear Embedding and Isometric Feature Mapping for EEG-based Brain Computer Interface. In *Proceedings of the 28th Workshop of the Austrian Association for Pattern Recognition, Hagenberg, Austria, 17–18 June 2004*; pp. 189–196.
77. Li, M.; Luo, X.; Yang, J.; Sun, Y. Applying a Locally Linear Embedding Algorithm for Feature Extraction and Visualization of MI-EEG. *J. Sens.* **2016**, *2016*, 1–9. [[CrossRef](#)]
78. Xie, X.; Yu, Z.L.; Lu, H.; Gu, Z.; Li, Y. Motor Imagery Classification Based on Bilinear Sub-Manifold Learning of Symmetric Positive-Definite Matrices. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 504–516. [[CrossRef](#)] [[PubMed](#)]
79. Davoudi, A.; Ghidary, S.S.; Sadatnejad, K. Dimensionality reduction based on distance preservation to local mean for symmetric positive definite matrices and its application in brain-computer interfaces. *J. Neural Eng.* **2017**, *14*, 036019. [[CrossRef](#)] [[PubMed](#)]
80. Tanaka, T.; Uehara, T.; Tanaka, Y. Dimensionality reduction of sample covariance matrices by graph fourier transform for motor imagery brain-machine interface. In *Proceedings of the 2016 IEEE Statistical Signal Processing Workshop (SSP), Palma de Mallorca, Spain, 26–29 June 2016*; pp. 1–5. [[CrossRef](#)]
81. Roy, S.; Rathee, D.; Chowdhury, A.; Prasad, G. Assessing impact of channel selection on decoding of motor and cognitive imagery from MEG data. *J. Neural Eng.* **2020**, *17*, 1–15 [[CrossRef](#)] [[PubMed](#)]
82. 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](#)] [[PubMed](#)]
83. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*, 2nd ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2001. [[CrossRef](#)]
84. Thomas, E.; Dyson, M.; Clerc, M. An analysis of performance evaluation for motor-imagery based BCI. *J. Neural Eng.* **2013**, *10*, 031001. [[CrossRef](#)]
85. Schlögl, A.; Kronegg, J.; Huggins, J.; Mason, S. Evaluation Criteria for BCI Research. In *Toward Brain-Computer Interfacing*; MIT Press: Cambridge, UK, 2007; Volume 1, pp. 327–342..
86. Wolpaw, J.R.; Ramoser, H.; McFarland, D.J.; Pfurtscheller, G. EEG-based communication: Improved accuracy by response verification. *IEEE Trans. Rehabil. Eng.* **1998**, *6*, 326–333. [[CrossRef](#)]
87. Nykopp, T. Statistical Modelling Issues for the Adaptive Brain Interface. Ph.D. Thesis, Helsinki University of Technology, Espoo, Finland, 2001.
88. Lotte, F.; Jeunet, C. Defining and quantifying users' mental imagery-based BCI skills: A first step. *J. Neural Eng.* **2018**, *15*, 046030. [[CrossRef](#)]
89. Solé-Casals, J.; Caiafa, C.; Zhao, Q.; Cichocki, A. Brain-Computer Interface with Corrupted EEG Data: A Tensor Completion Approach. *Cognit. Comput.* **2018**. [[CrossRef](#)]
90. Gaur, P.; Pachori, R.B.; Wang, H.; Prasad, G. An Automatic Subject Specific Intrinsic Mode Function Selection for Enhancing Two-Class EEG-Based Motor Imagery-Brain Computer Interface. *IEEE Sens. J.* **2019**, *19*, 6938–6947. [[CrossRef](#)]

91. Togha, M.M.; Salehi, M.R.; Abiri, E. Improving the performance of the motor imagery-based brain-computer interfaces using local activities estimation. *Biomed. Signal Process. Control* **2019**, *50*, 52–61. [[CrossRef](#)]
92. Sampanna, R.; Mitaim, S. Noise benefits in the array of brain-computer interface classification systems. *Inform. Med. Unlocked* **2018**, *12*, 88–97. [[CrossRef](#)]
93. Kumar, S.; Sharma, A. A new parameter tuning approach for enhanced motor imagery EEG signal classification. *Med. Biol. Eng. Comput.* **2018**, *56*. [[CrossRef](#)] [[PubMed](#)]
94. Kim, C.S.; Sun, J.; Liu, D.; Wang, Q.; Paek, S.G. Removal of ocular artifacts using ICA and adaptive filter for motor imagery-based BCI. *IEEE/CAA J. Autom. Sin.* **2017**, 1–8. [[CrossRef](#)]
95. Sun, L.; Feng, Z.; Chen, B.; Lu, N. A contralateral channel guided model for EEG based motor imagery classification. *Biomed. Signal Process. Control* **2018**, *41*, 1–9. [[CrossRef](#)]
96. Sagha, H.; Perdakis, S.; Millán, J.d.R.; Chavarriaga, R. Quantifying Electrode Reliability During Brain–Computer Interface Operation. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 858–864. [[CrossRef](#)] [[PubMed](#)]
97. Feng, J.K.; Jin, J.; Daly, I.; Zhou, J.; Niu, Y.; Wang, X.; Cichocki, A. An Optimized Channel Selection Method Based on Multifrequency CSP-Rank for Motor Imagery-Based BCI System. *Comput. Intell. Neurosci.* **2019**, 2019. [[CrossRef](#)] [[PubMed](#)]
98. Ramakrishnan, A.; Satyanarayana, J. Reconstruction of EEG from limited channel acquisition using estimated signal correlation. *Biomed. Signal Process. Control* **2016**, *27*, 164–173. [[CrossRef](#)]
99. Yang, Y.; Bloch, I.; Chevallier, S.; Wiart, J. Subject-Specific Channel Selection Using Time Information for Motor Imagery Brain–Computer Interfaces. *Cognit. Comput.* **2016**, *8*, 505–518. [[CrossRef](#)]
100. Ruan, J.; Wu, X.; Zhou, B.; Guo, X.; Lv, Z. An Automatic Channel Selection Approach for ICA-Based Motor Imagery Brain Computer Interface. *J. Med. Syst.* **2018**, *42*, 253. [[CrossRef](#)]
101. Park, S.M.; Kim, J.Y.; Sim, K.B. EEG electrode selection method based on BPSO with channel impact factor for acquisition of significant brain signal. *Optik* **2018**, *155*, 89–96. [[CrossRef](#)]
102. Jin, J.; Miao, Y.; Daly, I.; Zuo, C.; Hu, D.; Cichocki, A. Correlation-based channel selection and regularized feature optimization for MI-based BCI. *Neural Netw.* **2019**, *118*, 262–270. [[CrossRef](#)]
103. Yu, X.Y.; Yu, J.H.; Sim, K.B. Fruit Fly Optimization based EEG Channel Selection Method for BCI. *J. Inst. Control Robot. Syst.* **2016**, *22*, 199–203. [[CrossRef](#)]
104. Masood, N.; Farooq, H.; Mustafa, I. Selection of EEG channels based on Spatial filter weights. In Proceedings of the 2017 International Conference on Communication, Computing and Digital Systems (C-CODE 2017), Islamabad, Pakistan, 8–9 March 2017; pp. 341–345. [[CrossRef](#)]
105. Yang, Y.; Chevallier, S.; Wiart, J.; Bloch, I. Subject-specific time-frequency selection for multi-class motor imagery-based BCIs using few Laplacian EEG channels. *Biomed. Signal Process. Control* **2017**, *38*, 302–311. [[CrossRef](#)]
106. Rajan, R.; Thekkan Devassy, S. Improving Classification Performance by Combining Feature Vectors with a Boosting Approach for Brain Computer Interface (BCI). In *Intelligent Human Computer Interaction*; Horain, P., Achard, C., Mallem, M., Eds.; Springer: Cham, Switzerland, 2017; pp. 73–85.
107. Shahsavari Baboukani, P.; Mohammadi, S.; Azemi, G. Classifying Single-Trial EEG During Motor Imagery Using a Multivariate Mutual Information Based Phase Synchrony Measure. In Proceedings of the 2017 24th National and 2nd International Iranian Conference on Biomedical Engineering (ICBME), Tehran, Iran, 30 November–1 December 2017; pp. 1–4. [[CrossRef](#)]
108. Wang, J.; Feng, Z.; Lu, N.; Sun, L.; Luo, J. An information fusion scheme based common spatial pattern method for classification of motor imagery tasks. *Biomed. Signal Process. Control* **2018**, *46*, 10–17. [[CrossRef](#)]
109. Liu, A.; Chen, K.; Liu, Q.; Ai, Q.; Xie, Y.; Chen, A. Feature Selection for Motor Imagery EEG Classification Based on Firefly Algorithm and Learning Automata. *Sensors* **2017**, *17*, 2576. [[CrossRef](#)]
110. Kumar, S.; Sharma, A.; Tsunoda, T. An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information. *BMC Bioinform.* **2017**, *18*, 125–137. [[CrossRef](#)] [[PubMed](#)]
111. Samanta, K.; Chatterjee, S.; Bose, R. Cross Subject Motor Imagery Tasks EEG Signal Classification Employing Multiplex Weighted Visibility Graph and Deep Feature Extraction. *IEEE Sens. Lett.* **2019**, *4*, 1–4. [[CrossRef](#)]
112. Xie, X.; Yu, Z.L.; Gu, Z.; Zhang, J.; Cen, L.; Li, Y. Bilinear Regularized Locality Preserving Learning on Riemannian Graph for Motor Imagery BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 698–708. [[CrossRef](#)] [[PubMed](#)]
113. She, Q.; Gan, H.; Ma, Y.; Luo, Z.; Potter, T.; Zhang, Y. Scale-Dependent Signal Identification in Low-Dimensional Subspace: Motor Imagery Task Classification. *Neural Plast.* **2016**, *2016*, 1–15. [[CrossRef](#)]
114. Özdenizci, O.; Erdoğmuş, D. Information Theoretic Feature Transformation Learning for Brain Interfaces. *IEEE Trans. Biomed. Eng.* **2020**, *67*, 69–78. [[CrossRef](#)]
115. Razzak, I.; Hameed, I.A.; Xu, G. Robust Sparse Representation and Multiclass Support Matrix Machines for the Classification of Motor Imagery EEG Signals. *IEEE J. Transl. Eng. Health Med.* **2019**, *7*, 1–8. [[CrossRef](#)]
116. Harandi, M.T.; Salzmann, M.; Hartley, R. From Manifold to Manifold: Geometry-Aware Dimensionality Reduction for SPD Matrices. In *Computer Vision—ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Cham, Switzerland, 2014; pp. 17–32.
117. Li, M.; Han, J.; Duan, L. A novel MI-EEG imaging with the location information of electrodes. *IEEE Access* **2019**, *8*, 3197–3211. [[CrossRef](#)]
118. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556.

119. Wang, H.T.; Li, T.; Huang, H.; He, Y.B.; Liu, X.C. A motor imagery analysis algorithm based on spatio-temporal-frequency joint selection and relevance vector machine. *Kongzhi Lilun Yu Yingyong/Control Theory Appl.* **2017**, *34*, 1403–1408. [[CrossRef](#)]
120. Sadiq, M.T.; Yu, X.; Yuan, Z.; Fan, Z.; Rehman, A.U.; Li, G.; Xiao, G. Motor Imagery EEG Signals Classification Based on Mode Amplitude and Frequency Components Using Empirical Wavelet Transform. *IEEE Access* **2019**, *7*, 127678–127692. [[CrossRef](#)]
121. Selim, S.; Tantawi, M.M.; Shedeed, H.A.; Badr, A. A CSP AM-BA-SVM Approach for Motor Imagery BCI System. *IEEE Access* **2018**, *6*, 49192–49208. [[CrossRef](#)]
122. Athif, M.; Ren, H. WaveCSP: A robust motor imagery classifier for consumer EEG devices. *Australas. Phys. Eng. Sci. Med.* **2019**, *42*, 1–10. [[CrossRef](#)]
123. Li, X.; Guan, C.; Zhang, H.; Ang, K.K. A Unified Fisher's Ratio Learning Method for Spatial Filter Optimization. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2727–2737. [[CrossRef](#)] [[PubMed](#)]
124. Li, L.; Xu, G.; Zhang, F.; Xie, J.; Li, M. Relevant Feature Integration and Extraction for Single-Trial Motor Imagery Classification. *Front. Neurosci.* **2017**, *11*, 371. [[CrossRef](#)]
125. Liu, Y.; Zhang, H.; Chen, M.; Zhang, L. A Boosting-Based Spatial-Spectral Model for Stroke Patients' EEG Analysis in Rehabilitation Training. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2016**, *24*, 169–179. [[CrossRef](#)]
126. Cachón, A.; Vázquez, R.A. Tuning the parameters of an integrate and fire neuron via a genetic algorithm for solving pattern recognition problems. *Neurocomputing* **2015**, *148*, 187–197. [[CrossRef](#)]
127. Salazar-Varas, R.; Vazquez, R.A. Evaluating spiking neural models in the classification of motor imagery EEG signals using short calibration sessions. *Appl. Soft Comput.* **2018**, *67*, 232–244. [[CrossRef](#)]
128. Zhao, X.; Zhang, H.; Zhu, G.; You, F.; Kuang, S.; Sun, L. A Multi-Branch 3D Convolutional Neural Network for EEG-Based Motor Imagery Classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 2164–2177. [[CrossRef](#)] [[PubMed](#)]
129. Park, Y.; Chung, W. Frequency-Optimized Local Region Common Spatial Pattern Approach for Motor Imagery Classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1378–1388. [[CrossRef](#)] [[PubMed](#)]
130. Ma, Y.; Ding, X.; She, Q.; Luo, Z.; Potter, T.; Zhang, Y. Classification of Motor Imagery EEG Signals with Support Vector Machines and Particle Swarm Optimization. *Comput. Math. Methods Med.* **2016**, *2016*, 1–8. [[CrossRef](#)] [[PubMed](#)]
131. Costa, A.; Møller, J.; Iversen, H.; Puthusserypady, S. An adaptive CSP filter to investigate user independence in a 3-class MI-BCI paradigm. *Comput. Biol. Med.* **2018**, *103*, 24–33. [[CrossRef](#)] [[PubMed](#)]
132. Lotte, F.; Guan, C. Spatially Regularized Common Spatial Patterns for EEG Classification. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 3712–3715. [[CrossRef](#)]
133. Singh, A.; Lal, S.; Guesgen, H.W. Reduce Calibration Time in Motor Imagery Using Spatially Regularized Symmetric Positives-Definite Matrices Based Classification. *Sensors* **2019**, *19*, 379. [[CrossRef](#)]
134. Saha, S.; Ahmed, K.I.U.; Mostafa, R.; Hadjileontiadis, L.; Khandoker, A. Evidence of Variabilities in EEG Dynamics During Motor Imagery-Based Multiclass Brain–Computer Interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 371–382. [[CrossRef](#)]
135. He, H.; Wu, D. Transfer Learning for Brain–Computer Interfaces: An Euclidean Space Data Alignment Approach. *IEEE Trans. Biomed. Eng.* **2019**, *67*, 399–410. [[CrossRef](#)] [[PubMed](#)]
136. Hossain, I.; Khosravi, A.; Hettiarachchi, I.T.; Nahavandhi, S. Informative instance transfer learning with subject specific frequency responses for motor imagery brain computer interface. In Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; pp. 252–257. [[CrossRef](#)]
137. Dai, M.; Wang, S.; Zheng, D.; Na, R.; Zhang, S. Domain Transfer Multiple Kernel Boosting for Classification of EEG Motor Imagery Signals. *IEEE Access* **2019**, *7*, 49951–49960. [[CrossRef](#)]
138. Park, S.; Lee, D.; Lee, S. Filter Bank Regularized Common Spatial Pattern Ensemble for Small Sample Motor Imagery Classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 498–505. [[CrossRef](#)]
139. Azab, A.M.; Mihaylova, L.; Ang, K.K.; Arvaneh, M. Weighted Transfer Learning for Improving Motor Imagery-Based Brain–Computer Interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1352–1359. [[CrossRef](#)] [[PubMed](#)]
140. Singh, A.; Lal, S.; Guesgen, H.W. Motor Imagery Classification Based on Subject to Subject Transfer in Riemannian Manifold. In Proceedings of the 2019 7th International Winter Conference on Brain–Computer Interface (BCI), Gangwon, Korea, 18–20 February 2019; pp. 1–6. [[CrossRef](#)]
141. Singh, A.; Lal, S.; Guesgen, H.W. Small Sample Motor Imagery Classification Using Regularized Riemannian Features. *IEEE Access* **2019**, *7*, 46858–46869. [[CrossRef](#)]
142. Jiao, Y.; Zhang, Y.; Chen, X.; Yin, E.; Jin, J.; Wang, X.; Cichocki, A. Sparse Group Representation Model for Motor Imagery EEG Classification. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 631–641. [[CrossRef](#)] [[PubMed](#)]
143. Rodrigues, P.L.C.; Jutten, C.; Congedo, M. Riemannian Procrustes Analysis: Transfer Learning for Brain–Computer Interfaces. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 2390–2401. [[CrossRef](#)]
144. Zhu, X.; Li, P.; Li, C.; Yao, D.; Zhang, R.; Xu, P. Separated channel convolutional neural network to realize the training free motor imagery BCI systems. *Biomed. Signal Process. Control* **2019**, *49*, 396–403. [[CrossRef](#)]
145. Joadder, M.; Siuly, S.; Kabir, E.; Wang, H.; Zhang, Y. A New Design of Mental State Classification for Subject Independent BCI Systems. *IRBM* **2019**, *40*, 297–305. [[CrossRef](#)]
146. Zhao, X.; Zhao, J.; Cai, W.; Wu, S. Transferring Common Spatial Filters With Semi-Supervised Learning for Zero-Training Motor Imagery Brain–Computer Interface. *IEEE Access* **2019**, *7*, 58120–58130. [[CrossRef](#)]

147. Kwon, O.; Lee, M.; Guan, C.; Lee, S. Subject-Independent Brain-Computer Interfaces Based on Deep Convolutional Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 3839–3852. [[CrossRef](#)]
148. Yao, L.; Sheng, X.; Zhang, D.; Jiang, N.; Mrachacz-Kersting, N.; Zhu, X.; Farina, D. A Stimulus-Independent Hybrid BCI Based on Motor Imagery and Somatosensory Attentional Orientation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 1674–1682. [[CrossRef](#)]
149. Shu, X.; Chen, S.; Yao, L.; Sheng, X.; Zhang, D.; Jiang, N.; Jia, J.; Zhu, X. Fast Recognition of BCI-Inefficient Users Using Physiological Features from EEG Signals: A Screening Study of Stroke Patients. *Front. Neurosci.* **2018**, *12*, 93. [[CrossRef](#)]
150. Acqualagna, L.; Botrel, L.; Vidaurre, C.; Kübler, A.; Blankertz, B. Large-scale assessment of a fully automatic co-adaptive motor imagery-based brain computer interface. *PLoS ONE* **2016**, *11*, e0148886. [[CrossRef](#)]
151. Shu, X.; Chen, S.; Chai, G.; Sheng, X.; Jia, J.; Zhu, X. Neural Modulation by Repetitive Transcranial Magnetic Stimulation (rTMS) for BCI Enhancement in Stroke Patients. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Honolulu, HI, USA, 17–21 July 2018; pp. 2272–2275. [[CrossRef](#)]
152. Sannelli, C.; Vidaurre, C.; Müller, K.R.; Blankertz, B. A large scale screening study with a SMR-based BCI: Categorization of BCI users and differences in their SMR activity. *PLoS ONE* **2019**, *14*, e0207351. [[CrossRef](#)]
153. Zhang, R.; Li, X.; Wang, Y.; Liu, B.; Shi, L.; Chen, M.; Zhang, L.; Hu, Y. Using Brain Network Features to Increase the Classification Accuracy of MI-BCI Inefficiency Subject. *IEEE Access* **2019**, *7*, 74490–74499. [[CrossRef](#)]
154. Ahn, M.; Cho, H.; Ahn, S.; Jun, S.C. User's Self-Prediction of Performance in Motor Imagery Brain-Computer Interface. *Front. Hum. Neurosci.* **2018**, *12*, 59. [[CrossRef](#)]
155. Darvishi, S.; Gharabaghi, A.; Ridding, M.C.; Abbott, D.; Baumert, M. Reaction Time Predicts Brain-Computer Interface Aptitude. *IEEE J. Transl. Eng. Health Med.* **2018**, *6*, 1–11. [[CrossRef](#)]
156. Müller, J.; Vidaurre, C.; Schreuder, M.; Meinecke, F.; von Büna, P.; Müller, K.R. A mathematical model for the two-learners problem. *J. Neural Eng.* **2017**, *14*. [[CrossRef](#)]
157. Vidaurre, C.; Sannelli, C.; Müller, K.R.; Blankertz, B. Machine-learning-based coadaptive calibration for Brain-computer interfaces. *Neural Comput.* **2011**, *23*, 791–816. [[CrossRef](#)] [[PubMed](#)]
158. Lee, M.H.; Kwon, O.Y.; Kim, Y.J.; Kim, H.K.; Lee, Y.E.; Williamson, J.; Fazli, S.; Lee, S.W. EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy. *GigaScience* **2019**, *8*, giz002. [[CrossRef](#)] [[PubMed](#)]
159. Sannelli, C.; Vidaurre, C.; Müller, K.R.; Blankertz, B. Ensembles of adaptive spatial filters increase BCI performance: An online evaluation. *J. Neural Eng.* **2016**, *13*, 46003. [[CrossRef](#)] [[PubMed](#)]
160. Vidaurre, C.; Murguialday, A.R.; Haufe, S.; Gómez, M.; Müller, K.R.; Nikulin, V. Enhancing sensorimotor BCI performance with assistive afferent activity: An online evaluation. *NeuroImage* **2019**, *199*, 375–386. [[CrossRef](#)]
161. Yu, Y.; Zhou, Z.; Yin, E.; Jiang, J.; Tang, J.; Liu, Y.; Hu, D. Toward brain-actuated car applications: Self-paced control with a motor imagery-based brain-computer interface. *Comput. Biol. Med.* **2016**, *77*, 148–155. [[CrossRef](#)] [[PubMed](#)]
162. Cheng, P.; Autthasan, P.; Pijarana, B.; Chuangsuwanich, E.; Wilairaprasitporn, T. Towards Asynchronous Motor Imagery-Based Brain-Computer Interfaces: A joint training scheme using deep learning. In Proceedings of the 2018 IEEE Region 10 Conference (TENCON 2018), Jeju, Korea, 28–31 October 2018; pp. 1994–1998. [[CrossRef](#)]
163. Sanchez-Ante, G.; Antelis, J.; Gudiño-Mendoza, B.; Falcon, L.; Sossa, H. Dendrite morphological neural networks for motor task recognition from electroencephalographic signals. *Biomed. Signal Process. Control* **2018**, *44*, 12–24. [[CrossRef](#)]
164. Jiang, Y.; Hau, N.T.; Chung, W. Semiasynchronous BCI Using Wearable Two-Channel EEG. *IEEE Trans. Cognit. Dev. Syst.* **2018**, *10*, 681–686. [[CrossRef](#)]
165. Sun, Y.; Feng, Z.; Zhang, J.; Zhou, Q.; Luo, J. Asynchronous motor imagery detection based on a target guided sub-band filter using wavelet packets. In Proceedings of the 2017 29th Chinese Control And Decision Conference (CCDC), Chongqing, China, 28–30 May 2017; pp. 4850–4855. [[CrossRef](#)]
166. He, S.; Zhou, Y.; Yu, T.; Zhang, R.; Huang, Q.; Chuai, L.; Madah-Ui-Mustafa; Gu, Z.; Yu, Z.L.; Tan, H.; Li, Y. EEG- and EOG-based Asynchronous Hybrid BCI: A System Integrating a Speller, a Web Browser, an E-mail Client, and a File Explorer. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *1*. [[CrossRef](#)]
167. Yu, Y.; Liu, Y.; Jiang, J.; Yin, E.; Zhou, Z.; Hu, D. An Asynchronous Control Paradigm Based on Sequential Motor Imagery and Its Application in Wheelchair Navigation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 2367–2375. [[CrossRef](#)]
168. An, H.; Kim, J.; Lee, S. Design of an asynchronous brain-computer interface for control of a virtual Avatar. In Proceedings of the 2016 4th International Winter Conference on Brain-Computer Interface (BCI), Gangwon, Korea, 22–24 February 2016; pp. 1–2. [[CrossRef](#)]
169. Jiang, Y.; He, J.; Li, D.; Jin, J.; Shen, Y. Signal classification algorithm in motor imagery based on asynchronous brain-computer interface. In Proceedings of the 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Auckland, New Zealand, 20–23 May 2019; pp. 1–5. [[CrossRef](#)]
170. Yousefi, R.; Rezazadeh, A.; Chau, T. Development of a robust asynchronous brain-switch using ErrP-based error correction. *J. Neural Eng.* **2019**, *16*. [[CrossRef](#)]
171. Yu, Y.; Zhou, Z.; Liu, Y.; Jiang, J.; Yin, E.; Zhang, N.; Wang, Z.; Liu, Y.; Wu, X.; Hu, D. Self-Paced Operation of a Wheelchair Based on a Hybrid Brain-Computer Interface Combining Motor Imagery and P300 Potential. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 2516–2526. [[CrossRef](#)]

172. Wang, L.; Wu, X. Classification of Four-Class Motor Imagery EEG Data Using Spatial Filtering. In Proceedings of the 2008 2nd International Conference on Bioinformatics and Biomedical Engineering, Shanghai, China, 16–18 May 2008; pp. 2153–2156. [[CrossRef](#)]
173. Grosse-Wentrup, M.; Buss, M. Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction. *IEEE Trans. Biomed. Eng.* **2008**, *55*, 1991–2000. [[CrossRef](#)] [[PubMed](#)]
174. Christensen, S.M.; Holm, N.S.; Puthusserypady, S. An Improved Five Class MI Based BCI Scheme for Drone Control Using Filter Bank CSP. In Proceedings of the 2019 7th International Winter Conference on Brain-Computer Interface (BCI), Gangwon, Korea, 18–20 February 2019; pp. 1–6. [[CrossRef](#)]
175. Razzak, I.; Blumenstein, M.; Xu, G. Multiclass Support Matrix Machines by Maximizing the Inter-Class Margin for Single Trial EEG Classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1117–1127. [[CrossRef](#)]
176. Barachant, A.; Bonnet, S.; Congedo, M.; Jutten, C. Multiclass Brain Computer Interface Classification by Riemannian Geometry. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 920–928. [[CrossRef](#)] [[PubMed](#)]
177. Faiz, M.Z.A.; Al-Hamadani, A.A. Online Brain Computer Interface Based Five Classes EEG To Control Humanoid Robotic Hand. In Proceedings of the 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), Budapest, Hungary, 1–3 July 2019; pp. 406–410. [[CrossRef](#)]
178. Aliakbaryhosseinabadi, S.; Kamavuako, E.N.; Jiang, N.; Farina, D.; Mrachacz-Kersting, N. Classification of Movement Preparation Between Attended and Distracted Self-Paced Motor Tasks. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 3060–3071. [[CrossRef](#)] [[PubMed](#)]
179. Dagaev, N.; Volkova, K.; Ossadtchi, A. Latent variable method for automatic adaptation to background states in motor imagery BCI. *J. Neural Eng.* **2017**, *15*, 016004. [[CrossRef](#)] [[PubMed](#)]
180. Mondini, V.; Mangia, A.; Cappello, A. EEG-Based BCI System Using Adaptive Features Extraction and Classification Procedures. *Comput. Intell. Neurosci.* **2016**, *2016*, 1–14. [[CrossRef](#)]
181. Schwarz, A.; Brandstetter, J.; Pereira, J.; Muller-Putz, G.R. Direct comparison of supervised and semi-supervised retraining approaches for co-adaptive BCIs. *Med. Biol. Eng. Comput.* **2019**, *57*, 2347–2357. [[CrossRef](#)]
182. Saeedi, S.; Chavarriaga, R.; Millán, J.d.R. Long-Term Stable Control of Motor-Imagery BCI by a Locked-In User Through Adaptive Assistance. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 380–391. [[CrossRef](#)]
183. Perdikis, S.; Leeb, R.; Millán, J.d.R. Context-aware adaptive spelling in motor imagery BCI. *J. Neural Eng.* **2016**, *13*, 036018. [[CrossRef](#)]
184. Faller, J.; Vidaurre, C.; Solis-Escalante, T.; Neuper, C.; Scherer, R. Autocalibration and Recurrent Adaptation: Towards a Plug and Play Online ERD-BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2012**, *20*, 313–319. [[CrossRef](#)]
185. Raza, H.; Rathee, D.; Zhou, S.M.; Cecotti, H.; Prasad, G. Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related EEG-based brain-computer interface. *Neurocomputing* **2019**, *343*, 154–166. [[CrossRef](#)]
186. Rong, H.; Li, C.; Bao, R.; Chen, B. Incremental Adaptive EEG Classification of Motor Imagery-based BCI. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–7. [[CrossRef](#)]
187. Sharghian, V.; Rezaii, T.Y.; Farzamnia, A.; Tinati, M.A. Online Dictionary Learning for Sparse Representation-Based Classification of Motor Imagery EEG. In Proceedings of the 2019 27th Iranian Conference on Electrical Engineering (ICEE), Yazd, Iran, 30 April–2 May 2019; pp. 1793–1797. [[CrossRef](#)]
188. Zhang, Z.; Foong, R.; Phua, K.S.; Wang, C.; Ang, K.K. Modeling EEG-based Motor Imagery with Session to Session Online Adaptation. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 1988–1991. [[CrossRef](#)]
189. Andreu-Perez, J.; Cao, F.; Hagra, H.; Yang, G. A Self-Adaptive Online Brain-Machine Interface of a Humanoid Robot Through a General Type-2 Fuzzy Inference System. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 101–116. [[CrossRef](#)]
190. Ang, K.K.; Guan, C. EEG-Based Strategies to Detect Motor Imagery for Control and Rehabilitation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 392–401. [[CrossRef](#)]
191. Abdalsalam, E.; Yusoff, M.Z.; Malik, A.; Kamel, N.; Mahmoud, D. Modulation of sensorimotor rhythms for brain-computer interface using motor imagery with online feedback. *Signal Image Video Process.* **2017**. [[CrossRef](#)]
192. Ron-Angevin, R.; Díaz-Estrella, A. Brain-computer interface: Changes in performance using virtual reality techniques. *Neurosci. Lett.* **2009**, *449*, 123–127. [[CrossRef](#)]
193. Achancaray, D.; Pacheco, K.; Carranza, E.; Hayashibe, M. Immersive Virtual Reality Feedback in a Brain Computer Interface for Upper Limb Rehabilitation. In Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 7–10 October 2018; pp. 1006–1010. [[CrossRef](#)]
194. Alchalabi, B.; Faubert, J. A Comparison between BCI Simulation and Neurofeedback for Forward/Backward Navigation in Virtual Reality. *Comput. Intell. Neurosci.* **2019**, *2019*, 1–12. [[CrossRef](#)]
195. Asensio-Cubero, J.; Gan, J.; Palaniappan, R. Multiresolution Analysis over Graphs for a Motor Imagery Based Online BCI Game. *Comput. Biol. Med.* **2015**, *68*. [[CrossRef](#)] [[PubMed](#)]
196. Jianjun, M.; He, B. Exploring Training Effect in 42 Human Subjects Using a Non-invasive Sensorimotor Rhythm Based Online BCI. *Front. Hum. Neurosci.* **2019**, *13*, 128. [[CrossRef](#)]
197. Kim, S.; Lee, M.; Lee, S. Self-paced training on motor imagery-based BCI for minimal calibration time. In Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; pp. 2297–2301. [[CrossRef](#)]

198. Jeunet, C.; N’Kaoua, B.; Lotte, F. Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates. In *Brain-Computer Interfaces: Lab Experiments to Real-World Applications*; Progress in Brain Research; Coyle, D., Ed.; Elsevier: Amsterdam, The Netherlands, 2016; Chapter 1, Volume 228, pp. 3–35. [[CrossRef](#)]
199. Zhang, H.; Sun, Y.; Li, J.; Wang, F.; Wang, Z. Covert Verb Reading Contributes to Signal Classification of Motor Imagery in BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 45–50. [[CrossRef](#)]
200. Wang, L.; Liu, X.; Liang, Z.; Yang, Z.; Hu, X. Analysis and classification of hybrid BCI based on motor imagery and speech imagery. *Measurement* **2019**, *147*, 106842. [[CrossRef](#)]
201. Wang, Z.; Zhou, Y.; Chen, L.; Gu, B.; Liu, S.; Xu, M.; Qi, H.; He, F.; Ming, D. A BCI based visual-haptic neurofeedback training improves cortical activations and classification performance during motor imagery. *J. Neural Eng.* **2019**, *16*, 066012. [[CrossRef](#)]
202. Liburkina, S.; Vasilyev, A.; Yakovlev, L.; Gordleeva, S.; Kaplan, A. Motor imagery based brain computer interface with vibrotactile interaction. *Zhurnal Vysshei Nervnoi Deyatelnosti Imeni I.P. Pavlova* **2017**, *67*, 414–429. [[CrossRef](#)]
203. Pillette, L.; Jeunet, C.; Mansencal, B.; N’Kambou, R.; N’Kaoua, B.; Lotte, F. A physical learning companion for Mental-Imagery BCI User Training. *Int. J. Hum.-Comput. Stud.* **2020**, *136*, 102380. [[CrossRef](#)]
204. Škola, F.; Tinková, S.; Liarokapis, F. Progressive Training for Motor Imagery Brain-Computer Interfaces Using Gamification and Virtual Reality Embodiment. *Front. Hum. Neurosci.* **2019**, *13*, 329. [[CrossRef](#)] [[PubMed](#)]