Journal of Neural Engineering

TOPICAL REVIEW

How to successfully classify EEG in motor imagery BCI: a metrological analysis of the state of the art

To cite this article: Pasquale Arpaia *et al* 2022 *J. Neural Eng.* **19** 031002

Journal of Neural Engineering

CrossMark

RECEIVED
20 November 2021

REVISED
13 May 2022

ACCEPTED FOR PUBLICATION
31 May 2022

PUBLISHED
17 June 2022

TOPICAL REVIEW

How to successfully classify EEG in motor imagery BCI: a metrological analysis of the state of the art

Pasquale Arpaia[1,2,3,*], Antonio Esposito[1,2,4], Angela Natalizio[1,4] and Marco Parvis[4]

[1] Augmented Reality for Health Monitoring Laboratory (ARHeMLab), Università degli Studi di Napoli Federico II, Naples, Italy
[2] Department of Electrical Engineering and Information Technology (DIETI), Università degli Studi di Napoli Federico II, Naples, Italy
[3] Centro Interdipartimentale di Ricerca in Management Sanitario e Innovazione in Sanità (CIRMIS), Università degli Studi di Napoli Federico II, Naples, Italy
[4] Department of Electronics and Telecommunications (DET), Polytechnic of Turin, Turin, Italy
[*] Author to whom any correspondence should be addressed.

E-mail: pasquale.arpaia@unina.it

Keywords: electroencephalogram (EEG), motor imagery (MI), brain-computer interface (BCI), classification, machine learning, deep learning, brain inspired network

Abstract

Objective. Processing strategies are analyzed with respect to the classification of electroencephalographic signals related to brain-computer interfaces (BCIs) based on motor imagery (MI). A review of literature is carried out to understand the achievements in MI classification, the most promising trends, and the challenges in replicating these results. Main focus is placed on performance by means of a rigorous metrological analysis carried out in compliance with the international vocabulary of metrology. Hence, classification accuracy and its uncertainty are considered, as well as repeatability and reproducibility. *Approach.* The paper works included in the review concern the classification of electroencephalographic signals in motor-imagery-based BCIs. Article search was carried out in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses standard and 89 studies were included. *Main results.* Statistically-based analyses show that brain-inspired approaches are increasingly proposed, and that these are particularly successful in discriminating against multiple classes. Notably, many proposals involve convolutional neural networks. Instead, classical machine learning approaches are still effective for binary classifications. Many proposals combine common spatial pattern, least absolute shrinkage and selection operator, and support vector machines. Regarding reported classification accuracies, performance above the upper quartile is in the 85%–100% range for the binary case and in the 83%–93% range for multi-class one. Associated uncertainties are up to 6% while repeatability for a predetermined dataset is up to 8%. Reproducibility assessment was instead prevented by lack of standardization in experiments. *Significance.* By relying on the analyzed studies, the reader is guided towards the development of a successful processing strategy as a crucial part of a BCI. Moreover, it is suggested that future studies should extend these approaches on data from more subjects and with custom experiments, even by investigating online operation. This would also enable the quantification of the results reproducibility.

1. Introduction

A brain-computer interface (BCI) provides a direct communication channel between the user's brain and external devices, thus enabling non-muscular interactions [1]. A BCI involves three main steps: signals acquisition, processing, and translation into a command. The BCI loop is then typically closed with feedback, which can be either naturally or artificially provided. According to the degree of invasiveness of the neuroimaging technique(s) adopted in brain signals acquisition, a BCI system can be invasive, semi-invasive, or non-invasive [2]. In non-invasive BCIs, brain signals are acquired without entering the scalp. Although less risky, the measurement of brain activity is significantly affected with respect to invasive or semi-invasive techniques [3]. Despite that, electroencephalography (EEG) is a widely spread non-invasive technique [4]. In addition to non-invasiveness, its main advantages include high temporal resolution, low cost, wearability, and portability.

A useful distinction between BCIs is carried out in terms of reactive, passive, and active paradigms [5]. Reactive BCIs rely on the detection of brain potential evoked by sensory stimulation, e.g. visual, auditory, or haptic [6, 7]. Passive BCIs are instead based on spontaneous brain potentials whilst not involving voluntary control. These can as an example improve human-computer interaction by considering cognitive mental load [8, 9]. Finally, in active BCIs, the users voluntarily modulate their spontaneous brain potentials as a means for communication and control. As an advantage, active BCIs are independent of external events since they rely on performing intentional mental tasks [10].

Among the active paradigms, motor imagery (MI) is widely exploited in BCI research [11]. It consists of the imagination of a movement without its actual execution [12]. MI generates phenomena known as event-related desynchronization and event-related synchronization, a decrease or an increase in the power of sensorimotor rhythms, respectively [13]. These brain rhythms are localized over sensorimotor areas in the μ band (7–13 Hz) and β band (13–30 Hz) [14]. MI-BCIs have been extensively used as assistive tools for disabled people [15]. The most important achievements include rehabilitation [16–18], wheelchair control [19], cursor control [20, 21], and spelling systems [22]. Moreover, they appear as a promising technology also in virtual reality, gaming, robotic arm control, and navigation in 2D and 3D environments [23–26].

As previously stated, independence from external stimuli is one of the main advantages of active BCIs. This means that no additional hardware is required for stimulation and asynchronous paradigms are theoretically possible. Nevertheless, a relatively long training period is required for the users before being capable to modulate brain rhythms [11, 27, 28]. In addition, because of baseline brain activity overlapped with the neurophysiological phenomena of interest, MI detection is affected by low signal-to-noise ratio [11], and within- and cross-subject variability must be handled [29]. These aspects imply lack of robustness for the system.

Many challenges must be still faced in order to bring MI-BCI systems into everyday life [11, 30]. This is indeed true for signal acquisition, where the intention is to ensure a reliable, affordable, wearable, and portable end-user device [11]. Nevertheless, much research considers the development of EEG processing strategies for MI-BCI with the aim of improving the classification performance [31–33]. A crucial point for a robust and reliable system consists of achieving high classification performance indeed.

In recent years, many reviews resumed the characteristics of an MI-BCI [34], the signals exploited in control applications [2], the most popular applications [35], their usage in everyday life [4, 36], and open challenges along with possible solutions [37, 38]. Most reviews on MI focused on rehabilitation [15, 18, 39, 40]. Other studies, instead, gave an overview of the most popular signal processing steps and strategies in MI classification [11, 41], and deep learning techniques were of particular concern in recent works [42, 43]. However, current reviews do not compare the studies in terms of performance, and best strategies are merely highlighted per singular processing step, without considering the processing strategy as a whole.

In the present work, after a preparatory overview of MI data creation, an analysis of the state of the art is carried out by focusing on strategies proposed for MI processing. Indeed, identifying a suitable signal processing approach is a crucial requirement in building effective BCIs. The aim is to address the following questions:

- What levels of performance have been achieved so far in classifying EEG signals associated with MI?
- Which are the most promising trends in MI-BCI processing? Are there common elements in successful strategies?
- Which challenges should be addressed in terms of replication of the results?

Therefore, the remainder of the paper is organized as follows. Section 2 recalls background knowledge associated with the creation of MI data. Section 3 states the methods adopted in searching, selecting, and analyzing the studies. Then, section 4 reports the analysis carried out for the included studies. Finally, section 5 discusses the analysis results and addresses future challenges.

2. Background

Figure 1 represents a typical BCI system, including the acquisition block, the processing block, and the final application. Before reviewing signal processing approaches, an overview of MI data creation appears essential. This involves either the hardware and the adopted experimental procedures. Hence, EEG signal acquisition is discussed in the following, as well as typical experimental procedures applied in the MI-BCI field.

2.1. EEG signal acquisition

Utmost signal quality can be achieved with wired EEG acquisition systems adopting a large number of electrodes (typically up to 64 [44]) placed all over the scalp. A wired system is little constrained in terms of transferable amount of data and latency, and it is

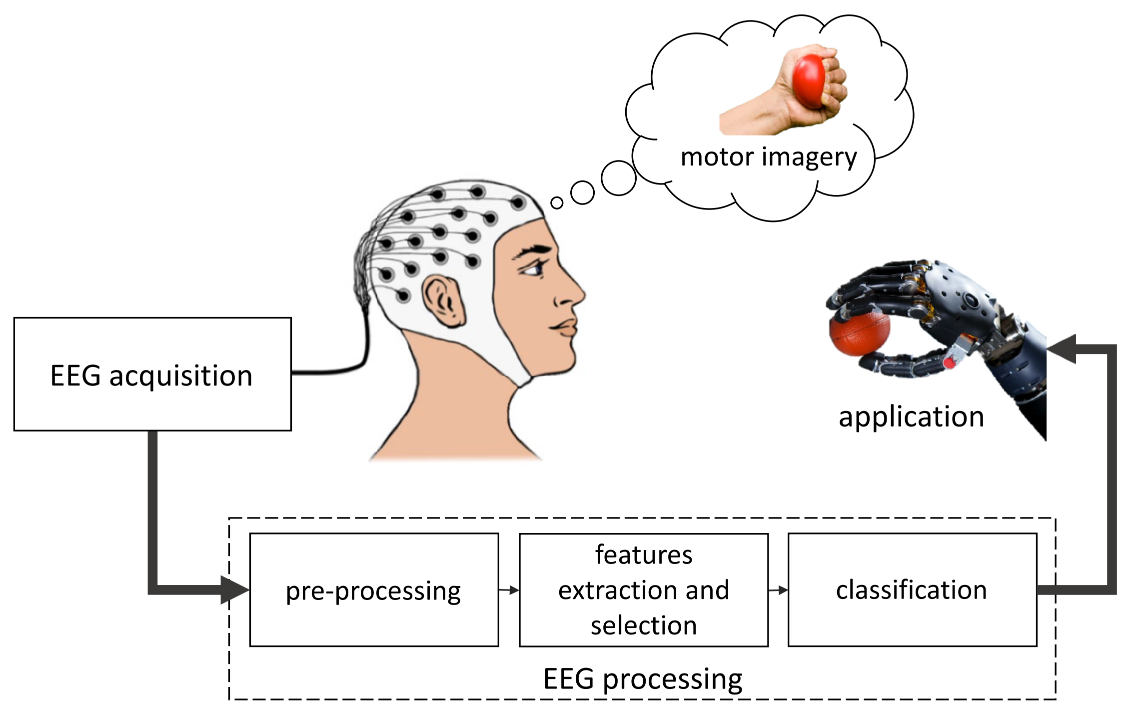

Figure 1. Functional components of a MI-BCI system based on EEG.

robust against packet loss and data corruption. However, the biggest drawback of a cable connection is limited user's mobility. The placement of electrodes on the scalp is regulated by the international standards, such as the 10–20 system, or the 10–10 system [45, 46]. The actual number of electrodes do depend on the application, and it could vary from as low as 1 to more than 200 [47]. The gold standard for clinical recordings are wet electrodes. These typically have silver/silver chloride (Ag/AgCl) coating, they use conductive gels or paste, and have a 1–3 mm diameter. The gel ensures a proper skin-electrode contact, namely a low impedance at the skin-electrode interface associated with good signal-to-noise ratio and high signal reliability. Thus, voltages in the μV range can be accurately detected with down to 10 ms time resolution [47]. Setups with the mentioned characteristics are reasonably considered a reference for EEG acquisition and they are commonly employed in research or even clinical applications.

In recent times, an increasing number of attempts has been made to develop EEG systems for 'out-of-the-lab' acquisitions [48]. Many of them are wireless devices (commonly WiFi or Bluetooth), thus allowing freedom of movement at the expense of greater latency and less robustness in data transmission if compared to wired devices. In addition, they have a reduced number of electrodes with respect to reference EEG setups. This clearly enhances usability and comfort for the end user, but it also implies that processing techniques must rely on a limited number of signals. In such a scenario, conductive gels are undesirable. Thus, semi-dry or dry electrodes are used. The former rely on electrolyte liquids, such as saline solution [49], while the latter do not need any conductive gel or liquid, but they are usually made of conductive materials like foams, rubbers, or carbon nanotubes [50–52].

Signal conditioning and digitization introduce reasonably negligible uncertainty during acquisition, even in commercial setups [53]. Therefore, the bottleneck to signal quality is in sensing with the electrodes.

EEG systems with wet and dry electrodes were compared in [54, 55]. While confirming shorter preparation times for dry electrodes, lowering contact impedance was time-consuming and these electrodes became uncomfortable after a few hours.

In [56], electrodes were also evaluated in a MI experiment by comparing them in terms of mean accuracy on 20 subjects, and no significant difference was revealed in distinguishing between left and right hand. Although dry and semi-dry electrodes are suggested as suitable alternatives to wet ones by different studies [54–57], only a few explicitly considered active BCI paradigms like MI. Instead, experiments were mostly considering evoked potentials, and further studies would be needed to better evaluate differences between electrodes for MI measurement. Hence, as it will be clear from the following, many available data for MI were acquired with setups relying on wet electrodes.

2.2. MI-BCI experimental procedures

The acquisition of EEG data requires specific procedures in order to investigate the neurophysiological phenomena of interest or validate the design of a BCI

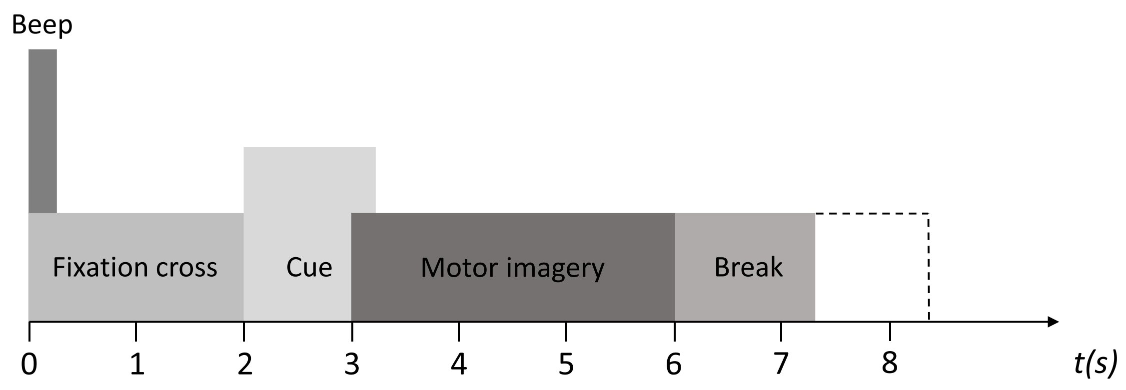

Figure 2. A representative example of synchronous timing scheme for MI experiments [71].

device. MI experiments usually take place over several sessions carried out on different days. Each session is divided into parts referred to as 'runs', and a run consists of repeating different tasks. Typically, the sequence of tasks is randomized while the number of trials per each task is fixed. Two to four MI tasks are currently considered among imagining the movement of a hand, both hands, a foot, both feet, the tongue, a wrist, an elbow, a forearm, or fingers [43]. Furthermore, at the beginning of a session, the EEG might be recorded during guided eye movements as a baseline for artifact removal.

Indeed, the BCI competitions held in 2000s [58–61] had a strong impact on the BCI community. Their goal was to promote signal processing in the field of BCI and challenge new paradigms and complex data [61]. The data generated in those events are still largely used today, and they have been inspiring experimental procedures especially in terms of timing schemes [62–65]. Two main timing schemes can be distinguished: synchronous or asynchronous schemes [66]. In a synchronous BCI system, interactions between the user and the system can only happen in specific time windows preceded by a cue. Conversely, in an asynchronous BCI, the user can perform the mental task at any time, though the distinction between an intentionally generated signal from an involuntary one is non-trivial. Figure 2 reports a representative example of a synchronous MI experiment recalled from the BCI competition IV. This consists of a cue-based paradigm where a single trial is divided into four main parts:

(a) an initial relax phase, often triggered by an acoustic signal, during which the user has to stare at a fixation cross appearing at the center of a screen; the cross helps the user to limit involuntary eye movements;
(b) a cue phase, during which an indication is provided about the imagination task to perform a while after;
(c) an interval for performing the MI task, usually with a 3–5 s duration;
(d) an ending relax phase (break), whose duration is random (few seconds) to avoid user's adaptation to timing.

Different schemes were also proposed in accordance with specific experimental needs by eventually relying on this basic scheme. For instance, some authors took into account distinguishing between resting state and MI [67], or others considered both MI and execution tasks [68]. Some studies also suggest to enlarge the MI window, e.g. up to 10 s, so as to make the system suitable for real-time applications [69]. Finally, in some other paradigms, the cue and the onset of MI overlap [70].

A major limitation in MI-BCI is the user's ability to properly imagine the movement [72], and many training sessions are required before being able to modulate sensorimotor rhythms. In these regards, neurofeedback is sought to improve MI training by engaging the user while performing the mental task. Closing the sensorimotor loop with neurofeedback is claimed to substantially change the way the user imagines a movement [73]. Therefore, several experimental procedures include runs with online EEG processing and sensory feedback [61, 74–76]. A typical timing scheme for MI experiments with neurofeedback is compatible with the one of figure 2, with the exception that the cue indication persists over all the MI period. Moreover, a sensory feedback is given during the MI period thanks to an online classification of the ongoing EEG signals. In classifying MI-related signals, the knowledge about the experimental procedure is surely essential, and the MI period must be usually extracted for processing. In these regards, researchers are currently investigating how often the neurofeedback should be provided [77], if to provide only positive or also negative feedback [69], and whether multi-sensory feedback is better than unimodal one [78]. As a side note, some researchers argue that feedbacks may introduce artifacts in the recorded signals [77], and suitable paradigms should be used to investigate that. These questions remain mostly unanswered due to the large number of involved variables, but sharing common methods and paradigms should help in finding the answers.

3. Methods

The current section describes the methods adopted when reviewing the literature about signal processing approaches for MI-BCIs. The aim was to compare the best processing approaches for EEG signals classification and then guide the reader towards the development of a successful processing strategy. In this regard, the studies were included according to a standardized procedure, and particular attention was given to the qualitative aspects of their proposals. Then, a rigorous analysis of the performance was carried out by exploiting the framework established with the international vocabulary of metrology. In doing that, statistical tools were employed for an objective quantification of the classification performance.

3.1. Articles search and selection

The present study was carried out according to the 'Preferred Reporting Items for Systematic reviews and Meta-Analyses' (PRISMA) statement [79] schematically represented in figure 3. The article search focused on paper works aiming to classify MI signals in BCI systems, while papers primarily focusing on channel reduction or strategies for improving training were not taken into account. To compare their results, studies considering at least one public dataset were selected. This also guarantees the repeatability of the reported results, i.e. the proposed processing can be applied to the same experimental setting to assess the compatibility of the achieved results. Furthermore, the reproducibility of the results can be evaluated by testing the proposed strategies on different datasets, either public or non-public. Article search was carried out by means of the Scopus and PubMed search engines. The following combinations of keywords were exploited: *(BCI AND motor AND imagery AND public AND dataset) OR (BCI AND competition AND dataset) OR (motor AND imagery AND dataset)*. As an alternative to *BCI*, its extended version (*brain computer interface*) was exploited. Only journal papers, written in English, published between 2017 and 2021, and at the final publication stage were included in the present study.

As a preliminary step, the duplicated papers were excluded. Thereafter, the title and abstract of each record were screened. Part of them were then excluded according to the following criteria:

- signal source—only studies exploiting EEG signals were included, while works considering neuroimaging techniques like fNIRS, ECoG, or even hybrid combinations with EEG were excluded;
- dataset—works testing the proposed approaches by removing subjects from the whole dataset without reasonable motivations were excluded, and datasets including only one subject were excluded as well;
- paper scope—works aiming to improve the classification of MI were only included, while articles related to topics like channel reduction or transfer learning for training time improvement were excluded.

After that, the remaining papers were considered for eligibility by means of their full texts. Part of them were furtherly excluded for not matching the aforementioned criteria. Finally, a total of 89 studies were actually considered for the present review. Further details on screening and eligibility are shown in figure 3.

3.2. Data collection

The information extracted from the included studies was enclosed in the eight categories reported below:

(i) dataset;
(ii) MI tasks;
(iii) pre-processing step;
(iv) features extraction step;
(v) feature selection step;
(vi) classification step;
(vii) assessment method;
(viii) performance.

Notably, in some approaches, the processing steps (iii)–(vi) may be indistinguishable. Meanwhile, assessment methods refer to how data were used to validate the proposal, and highest achieved results are reported as performance.

3.3. Metrological analysis

The included studies were analyzed by relying on the results reported in the respective studies. In carrying out a rigorous analysis, a metrological framework was adopted as it provides a set of statistical tools for an objective quantification of results. Aspects of concern were the uncertainty, the repeatability, and the reproducibility of the results. In agreement with the international vocabulary of metrology [80], the uncertainty characterizes the degree of dispersion of the classification performance, the repeatability characterizes the dispersion of results when taking into account the same experimental setups, while the reproducibility refers to the dispersion of results associated with setups with differences under control. By exploiting a type-A statistical assessment [81], the uncertainty of a mean performance value was calculated as the associated standard deviation divided by the square root of the number of values contributing to the mean. In the present analysis, the standard

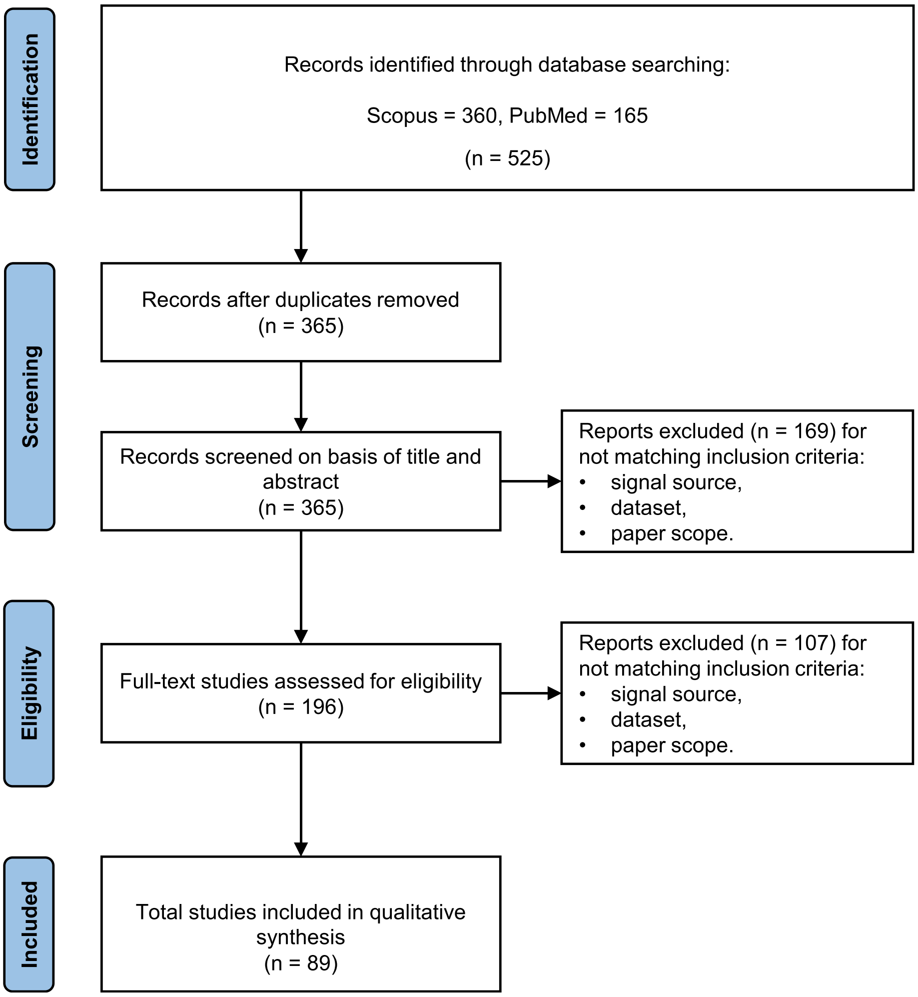

Figure 3. PRISMA flow diagram for identification, screening, eligibility, and inclusion of relevant studies. The number of studies is reported for each step.

deviation was either retrieved from each study, or it was calculated from the reported performance values. In calculating the repeatability, the range of performance values was considered once the dataset was fixed. This range corresponded to the difference between the maximum value and the minimum one. Finally, the same criterion was used for calculating the range associated with the reproducibility, though in this case the performance values were associated with different datasets.

For comparison purposes, studies exploiting at least one public dataset had to be considered in analyzing MI-BCIs achievements. Then, a focus had to be made on metrics for reporting the performance of MI-BCI prior to identifying the best proposals. Overall, statistically-based analyses were carried out by considering the distribution of reported performance, so to retrieve not only median performances, but also the most performing approaches with results above the upper quartile of the distribution. The focus on most promising approaches thus allowed to highlight promising trends in terms of successful processing strategies. The results of collected data analyses are reported and discussed in the following.

4. Results

In this section, analysis results are reported in terms of exploited data, processing approaches, performance assessment methods, and achieved performance. Common trends are highlighted by also considering their chronological evolution. In analyzing proposals and results, some hints are given on how to develop an effective processing approach for MI classification. A focus on metrological aspects is given for a rigorous quantification of current levels of performance and highlighting the possibility to replicate most promising results.

4.1. Data

Twelve public datasets were exploited in total, including those made publicly available by the BCI competitions. A brief description of these is given in table 1 and the respective reference is indicated for further

Table 1. Summary of public datasets employed by the studies included in the comparative analysis. LH: left hand, RH: right hand, F: both feet, LF: left foot, RF: right foot, T: tongue. *The dataset should be public but it is not available on any platform.

Datasets	Subjects	Channels	Tasks	% of use
BCI competition IV-2a [71]	9	22	LH, RH, F, T	58
BCI competition IV-2b [26]	9	3	LH, RH	40
BCI competition III-4a [60]	5	118	RH, RF	30
BCI competition III-3a [60]	3	64	LH, RH, F, T	19
BCI competition IV-1 [61]	7	59	LH, RH, F	15
Dataset from GigaScience [82]	52	64	LH, RH	3
High-Gamma Dataset [83]	14	128	LH, RH, F, rest	3
PhysioNet EEG [84]	109	64	Opening and closing left/right fist or both fists/feet	2
MAMEM Phase I [85]	34	61	LH, RH	2
Upper limb movements [86]	15	61	Elbow flexion/extension, forearm supination/pronation, hand open/close (right upper limb)	1
SMR-BCI [87]	14	15	RH, F	1
Dataset from Shanghai Jiao Tong University (SJTU) [88]*	5	62	LH, RH	1

details. The highlighted information consists of the number of participants, number of EEG channels, number of MI tasks, and a percentage of usage for the dataset estimated from the included studies. Most studies exploited more than one dataset to validate their methods. However, only a limited percentage of the studies (9 out of 89) exploited their own data to furtherly test the proposed methods. A detailed description of these non-public datasets is given in table 2 for the sake of repeatability of the results. In addition to the reported information, specifying the channel positions is also recommended. The considered datasets as a whole, either public or non-public, relied on synchronous acquisition paradigms, with the only exception of the dataset 1 from the BCI competition IV and the online experiments carried out in [70, 89], where an asynchronous paradigm was adopted.

4.2. Processing approaches

Measured EEG signals contain artifacts from external sources (e.g. environmental) and internal ones (e.g. physiological). A pre-processing phase is thus required in order to improve the EEG signal quality, especially during online experiments [36]. Temporal and spatial filtering approaches are mostly used in MI-BCI [11]. Regarding temporal filtering, Butterworth or Chebyshev filters with band from 8 to 30 Hz are typically applied to remove artifacts while preserving the information in μ and β bands. Meanwhile, common average referencing and Laplacian spatial filters [95] are widely used to extract spatial information associated with motor activities.

Time-domain, frequency-domain, and spatial domain are popular feature extraction techniques in EEG-based MI-BCIs. Time-domain methods study sensorimotor rhythms over time by means of statistics like mean, root mean square, standard deviation, variance, skewness, and kurtosis [96], but also power analysis or auto-regressive (AR) modeling [97] were explored, as well as graph-based methods [98]. These methods typically operate channel-by-channel. In extracting frequency-domain information, the most exploited methods are fast Fourier transform [33], power spectral density [20], and band power [99]. Temporal or frequency features alone may not be sufficient for the classifier. Therefore, time-frequency methods are employed too. Common algorithms are the short-term Fourier transform [100] and the wavelet transforms [101]. Finally, in the spatial domain, the most popular approach is the common spatial pattern (CSP) with its variants [38]. In order to retrieve band-specific information, this is often used in combination with filter banks [102]. Independent component analysis (ICA) and the Laplacian filtering are also used extensively for extracting spatial domain features [38]. Once features are extracted, a feature selection process is also required to eliminate redundant information. Correlation criteria and mutual information are widely used ones [11].

Lastly, classification strategies are involved. Common classification strategies are decision trees [103], support vector machines (SVMs) [104, 105], and linear discriminant analysis (LDA) [106], while most recent approaches are deep neural networks [107] or spiking neural networks [108]. Notably, most recent approaches can accomplish feature extraction, selection, and classification as a single pipeline. It is worth noting that, especially with deep approaches, it is easy to incur in overfitting due to the scarcity of available data [43]. To overcome this problem, data augmentation or transfer learning strategies can be used. Data augmentation is a set of techniques aiming to artificially increase the amount of available data by generating new data samples from existing training data. In doing that, better generalization to further unseen data is provided [109]. Meanwhile, transfer learning allows to train a model by relying on the knowledge

Table 2. Details about the non-public datasets employed by the studies included in the comparative analysis. Empty spaces are associated with unavailable information. Channel position refers to the 10-20 or 10-10 standards. BP: band-pass, LP: low-pass. LH: left hand, RH: right hand, F: both feet, LF: left foot, RF: right foot, T: tongue, (n.s.): not specified.

	Subjects	Device	Electrode type	Channels	Reference channel	Ground channel	Sample rate (Sa/s)	Filters	Sessions and trials	Tasks
[90]	12			64			250	8–30 Hz BP	234 training trials, 234 testing trials	LH, RH
[91]	12	Neuroscan		60	2 (n.s.)		1000			LH, RH
[67]	11	eegoTMrt Ant Neuro		64	CPz	AFz	512	0.5–40 Hz BP	160 trials	Dominant hand MI (kinesthetic grasping)
[65]	8	Open BCI	Dry	8	Fz	A2			5 sessions, 120 trials per session	LH, RH, LF, RF
[89]	8	UE-16B EEG amplifier		16	A1, A2	Forehead	1000	100 Hz LP, 50 Hz notch	1 training session (100 trials), 1 online session	LF, RH, F, T
[92]	5	g.tec		2	Fz		256	50 Hz notch	2 Sessions	LH, RH
[70]	5	Emotiv Epoc+	Semi dry	6	P3 and P4		128		180 offline trials, 420 online trials	LH, RH
[93]	5	SynAmps2 system (Neuroscan)		21	Vertex	Forehead	1000	0.5–200 Hz BP	3 sessions, 40 trials per session	RH index finger (extensions/flex-ions), idle state
[93]	4	SynAmps2 system (Neuroscan)		27	Mastoid behind the left ear	Forehead	1000	0.5–200 Hz BP	3 sessions, 40 trials per session	LH, RH
[94]	1	Emotiv Epoc					128		160 trials	LH, RH

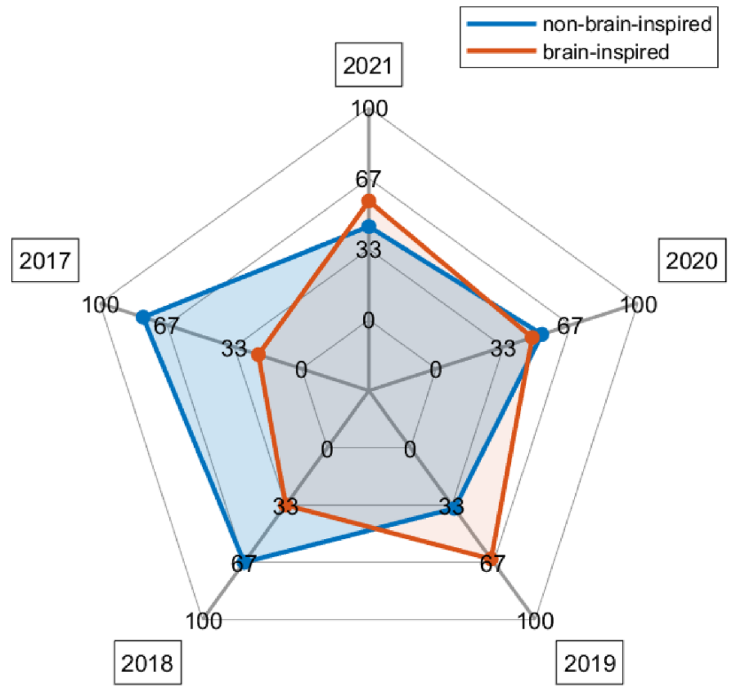

Figure 4. Percentage of studies proposing non-brain-inspired and BI approaches in the last five years.

learned from another pre-trained model [43], with the advantage of reducing training time.

As a first result of the analysis, two main approaches could be distinguished for signal processing in accordance with literature [110–112]: non-brain-inspired machine learning approaches and brain-inspired (BI) ones. Overall, machine learning algorithms learn from data without making data regularities explicit. Within this field, it appears useful to underline the BI approaches due to the increasing interest on them. Therefore, non-brain-inspired approaches (non-BI) consist of learning from data usually by extracting features designed *a-priori*. They simply need to learn, through training, how to handle each new problem. Examples of these algorithms are decision trees or SVM. Meanwhile, BI approaches propose algorithms inspired by human neurons. The neuron, or node, receives the input from external sources and, to produce an output, it has weights, biases parameters, and activation functions to be learnt from data. Among the BI approaches, three main sub-areas can be mentioned: artificial neural networks, deep neural networks, and spiking neural networks. The first two cases take inspiration from the value scaling performed by the synapses. Specifically, these algorithms consist of a set of interconnected nodes (neurons), each one implementing a weighted sum of input values and thresholding through a non-linear function. In designing a neural network, their structure is merely defined, while the effective number of neurons and their connections are identified by training the model. Each network has an input layer, hidden layers, and an output layer. Actually, deep neural networks are a subset of artificial ones in which more than one hidden layer is considered [111]. Spiking neural networks are instead inspired by communication of dendrites and axons [108]. Hence, they are based on spike-like pulses. The information transmitted depends on the amplitude of a spike and the time at which the pulse arrives. In addition, the computation that occurs in the neuron is a function of the pulse amplitude and the temporal relationship between different pulses.

In the current distinction, proposals exploiting a neural network for at least a part of the processing fall into the BI category. As a whole, the included works could be separated into 53% non-BI and 47% BI. However, by considering their chronological evolution (figure 4), in the last 5 years non-BI approaches have been gradually less exploited in favor of BI ones.

4.3. Assessment and performance

The reviewed papers reported results in terms of classification accuracy or equivalent metrics like error rate and Cohen's kappa coefficient. It is worth recalling that classification accuracy measures the percentage of correctly predicted target trials with reference to all available trials, the error rate is simply its complement, while the Cohen's kappa coefficient also takes into account the possibility of an agreement occurring by chance [113]. Note that the concept of classification accuracy should not be confused with the measurement accuracy, defined in [80] as the 'closeness of agreement between a measured quantity value and a true quantity value of a measurand'.

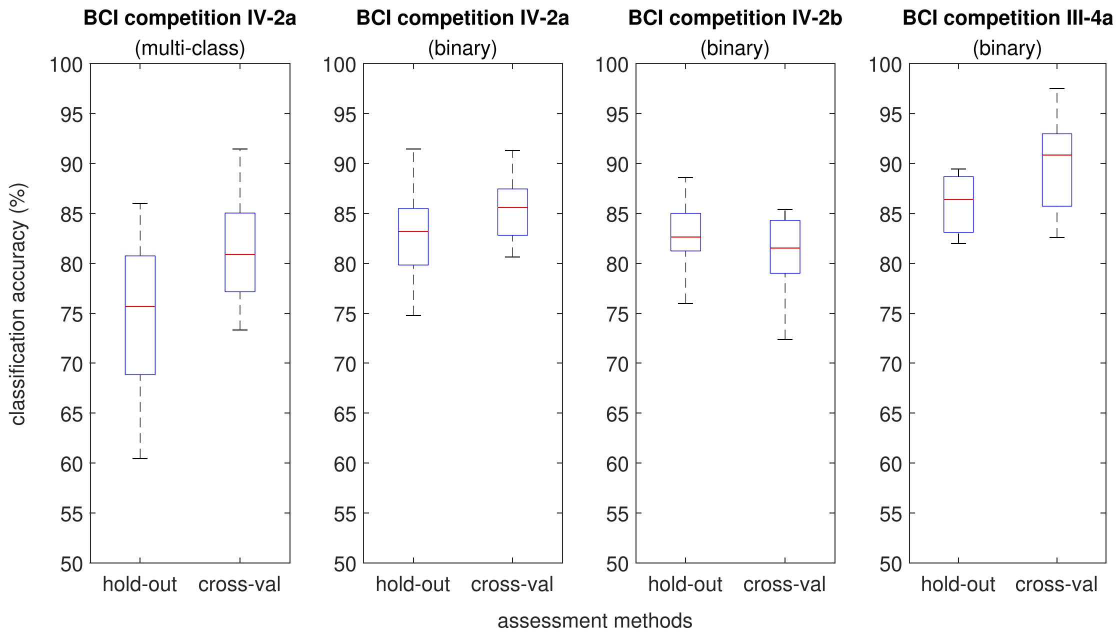

Figure 5. Performances assessed with hold-out versus cross-validation on data from BCI competitions.

In assessing classification accuracy, different methods are adopted to split a dataset in training and evaluation data. The most frequently exploited are cross-validation and hold-out. In cross-validation, the dataset is split into *k* parts (folds), where *k-1* folds are used to train the algorithm and the remaining one is used for evaluation. The splitting is repeated *k* times and the mean accuracy over these iterations is given as the final performance. A standard deviation is also associated with that. In the hold-out method, the dataset is just split once into two parts, a part for training and a part for evaluation. Therefore, accuracy is only calculated once.

In the present review, the 45% of works adopted cross-validation, with a number of folds between 5 and 30, while the 37% of them exploited the hold-out method. The remaining 8% includes unspecified methods or other methods such as cross-validation on segmented trials and averaging over different training/test splits. Finally, the 10% did not specify the adopted assessment method. A comparative analysis of assessment methods was carried out through the box-plots of figure 5. Per each dataset, these were built upon the reported mean accuracy across subjects. The three mostly used public datasets (BCI competition IV-2a, BCI competition IV-2b and BCI competition III-4a) were considered there to maximize the number of works to compare. Notably, the first two cases in figure 5 are both related to BCI competition IV-2a with reference to 4-tasks classification and 2-tasks classification, respectively. Interestingly, the cross-validation outperforms hold-out in all cases except BCI competition IV-2b, whereas in that case data were more variegated, e.g. EEG signals were acquired either with and without feedback.

4.4. Promising trends

In comparing the best processing approaches for MI, the datasets exploited at least in 10% of the included studies were considered, namely the BCI competition IV-2a, BCI competition IV-2b, BCI competition III-4a, BCI competition III-3a, and BCI competition IV-I datasets. Note that the BCI competition IV-2a and BCI competition III-3a datasets include up to four classes. Only two studies were excluded in this step because they did not use any of the mentioned datasets, but only the PhysioNet EEG dataset [114, 115]. Per each dataset, a box-plot was built by relying on the respectively reported accuracies (mean across subjects). All possible pairs of classes were considered for the binary classification case (figure 6), which is distinguished from the multi-class case considering all classes (figure 7). Proposals above the 75th percentile (upper quartile) were better investigated as they are associated with the highest classification accuracies (section 4.2). Outliers in the upper part of the box-plots deserved specific attention because they are associated with a performance significantly higher than the other observations. In the current case, one of such outliers indeed appeared. Figures 8 and 9 resume the publication years, processing strategies, and assessment methods related to the binary and multi-class cases, respectively. For binary classification, highest accuracies are concentrated in the last years, with more than half best performances achieved in 2021. Non-BI approaches are dominant, but it should be taken into account that, although they are increasing, there were fewer works exploiting BI approaches in these 5 years. Lastly, the majority of works used cross-validation to assess the performance. Even in the multi-class case, the

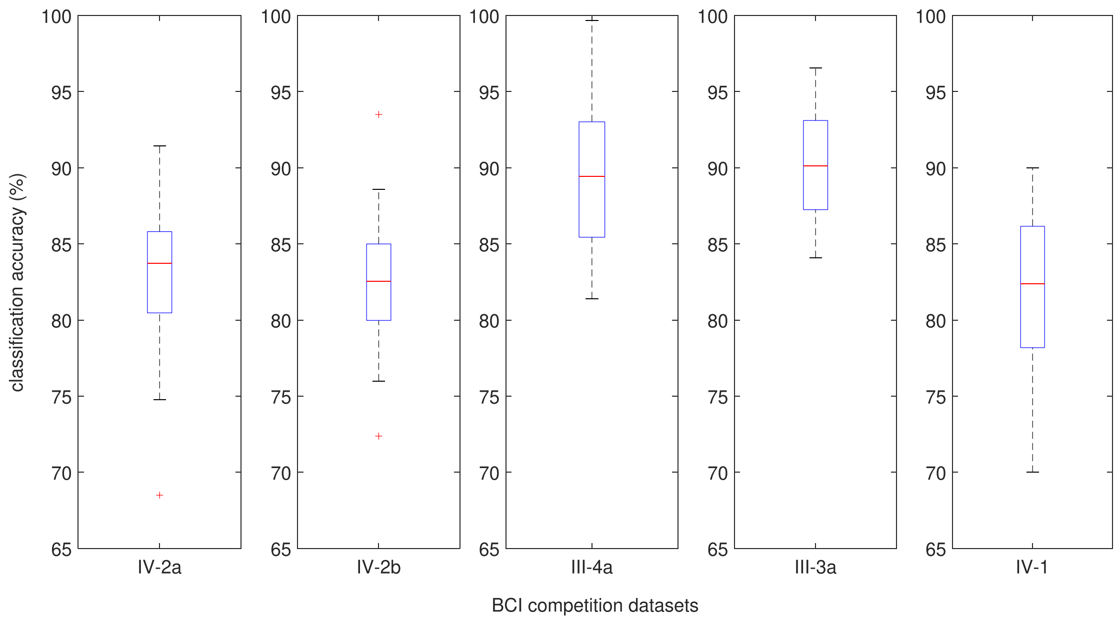

Figure 6. Box-plots of accuracies achieved in binary classification cases using the mostly exploited public datasets.

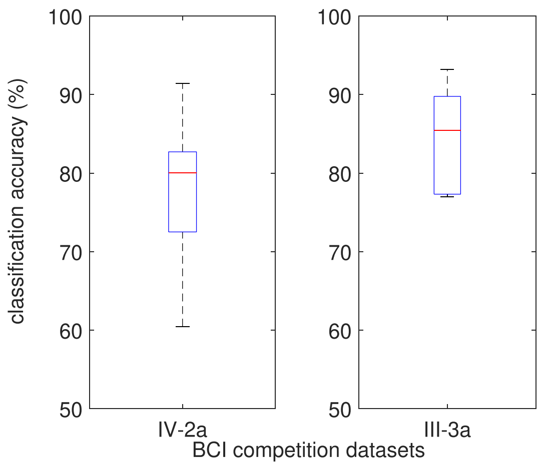

Figure 7. Box-plot of accuracies achieved in the multi-class cases using the mostly exploited datasets.

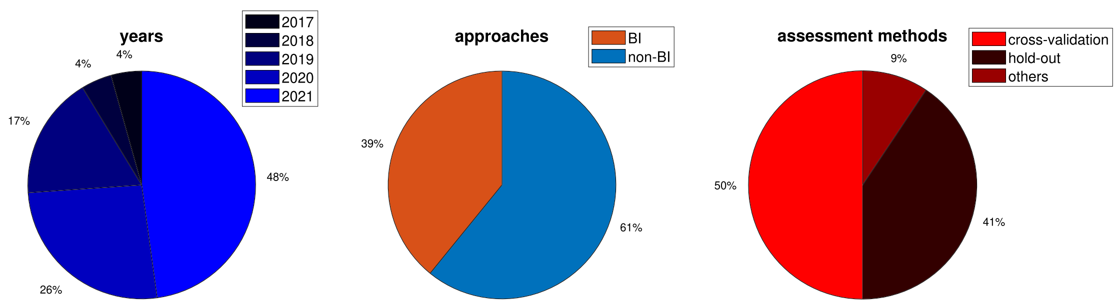

Figure 8. Analysis of most performing approaches in binary classification.

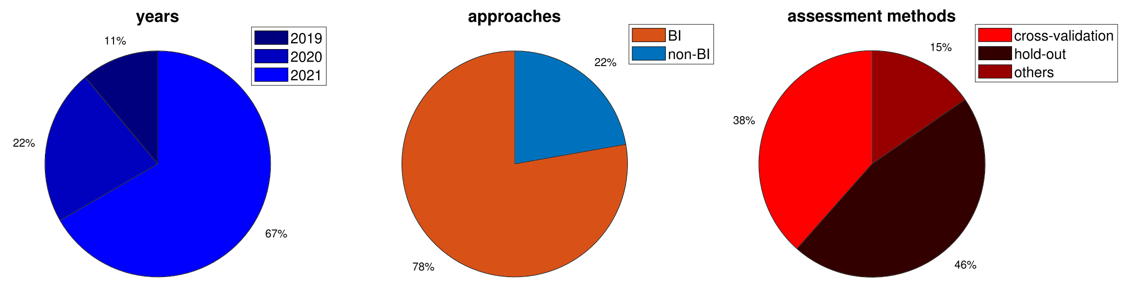

Figure 9. Analysis of most performing approaches in the multi-class case.

best results were obtained in recent years, particularly in 2021. This is also motivated by a greater number of works investigating multi-class problems in recent years. Instead, unlike the binary case, BI approaches stand out, while hold-out was preferred for performance assessment. These two facts are correlated because cross-validation is typically needed in BI approaches to tune the hyperparameters of the network (e.g. define the network structure), while hold-out is then used to test the trained network.

4.5. Most performing approaches

The abovementioned best proposals were analyzed in detail to highlight the best strategies for a performant processing. The reported classification accuracies above the upper quartile are in the 85%–100% range for the binary case and in the 83%–93% range for the multi-class case. With the type-A statistical assessment method, it was calculated that the uncertainty associated with these performance values is up to 6% in binary classification, and up to 5% in the multi-class case. It is worth remarking that type-A uncertainty assessment was here carried out by dividing the standard deviation of the reported results by the square root of the number of available subjects. Therefore, the lower the uncertainty, the more homogeneous the performance is across subjects. The results suggest that there is no substantial difference in terms of uncertainty among best approaches. Moreover, their overall classification performances are compatible since the intervals defined by the respective mean accuracy and associated uncertainty overlap. In terms of repeatability, the classification accuracies vary up to 7% in the binary case and 8% in the multi-class case. These values were retrieved as the difference between maximum and minimum classification accuracies associated with most performing approaches per each dataset. Hence, they correspond to the upper tails of box-plots shown in figures 6 and 7 for the binary and multi-class case, respectively. Interestingly, best repeatability was achieved on the dataset BCI competition III-3a and it resulted equal to 3% for both the binary and multi-class cases. However, performance is assessed in this case on a limited number of subjects, which may not be a statistically significant sample. Finally, if looking at the overall variation of classification accuracies across datasets, reproducibility equals 15% and 10% for the binary and multi-class case, respectively. However, this assessment merely refers to the mostly exploited datasets, i.e. datasets from BCI competitions. Further considerations are prevented by the limited number of studies exploiting different datasets, and, concerning non-public datasets, some information on the experimental setups is also missing, as already shown in table 2.

After these considerations, the analysis was carried on by distinguishing the two main approaches. Overall, the most performing non-BI and BI approaches are resumed in tables 3 and 4, respectively. In non-BI approaches, the achieved accuracies span from 85% to 100% for the binary case and from 86% to 93% for the multi-class case. About eight studies out of ten relied on features extraction strategies based on CSP [116], sometimes in combination with further extraction techniques. About half of the studies used the CSP in combination with classification based on a SVM [117]. This fact suggests that (i) the classical combination of CSP and SVM is effective and (ii) the non-BI studies reaching the highest accuracy are mostly distinguished in terms of pre-processing and feature selection techniques. In the pre-processing step, standing out strategies consist of time segmentation of trials, *a-priori* selection of channels over the sensorimotor area, and band-pass filtering often done over multiple bands [102]. The wavelet transform is frequently exploited too [118, 119]. Moreover, among studies using CSP and SVM, least absolute shrinkage and selection operator (LASSO) [120] was also used to select the most discriminating features. Notably, only two among the mostly performing non-BI proposals were applied to the multi-class case.

In BI approaches, the achieved accuracies span from 82% to 99% for the binary case and from 82% to 95% for the multi-class case. Only a single included work [108] proposed spiking neural networks and the best reported accuracy was 91% in a multi-class case, with 4% uncertainty. All the other works adopted artificial neural networks. Among them, almost seven out of ten used deep neural networks, where

Table 3. Most performing processing strategies using non-BI approach. **B** and **M** indicate the dataset for which the proposal has exceeded the 75th percentile of accuracies distribution for the binary and multiclass cases, respectively (see figures 6 and 7). AR: auto-regressive coefficient, BP: band-pass, CDSF: class discrepancy guided sub-band filter, CSP: common spatial pattern, FB: filter bank, k-NN: k-nearest neighbors, LASSO: least absolute shrinkage and selection operator, LDA: linear discriminant analysis, MAV: mean absolute value, PSR: power spectral ratio, RMS: root mean square, STDF: spatio-temporal discrepancy feature, SVM: support vector machine, BCIc.: BCI competition, GS: Giga Science.

Pre-processing	Feature extraction	Feature selection	Classification	Dataset	
Normalization	Principal component analysis and cross-covariance	Best-first algorithm and correlation-based variable selection	Multi-layer perceptron	BCIc.III-4a (**B**)	[125]
A-priori time windows selection, BP and FB in [8, 30] Hz	CSP	LASSO	SVM	BCIc.IV-1, BCIc.III-3a (**B**), BCIc.III-4a	[126]
Overlapped time windows, dual tree complex wavelet transform, in [4, 8], [8, 16] and [16, 32] Hz	CSP	Regularized neighborhood component analysis	SVM	BCIc.IV-2a (**B**), BCIc.IV-2b, BCIc.III-3a (**B**)	[118]
BP in [8, 30] Hz	CSP	Random Forest	Ensemble of random subspace k-NN	BCIc.IV-2a (**M**), BCIc.III-3a (**M**), BCIc.III-4a (**B**)	[127]
Channel and time window selection, FB in [8, 26] Hz and BP in [8, 24] Hz	Geodesic filtering CSP	Mutual information and Pearson correlation coefficient	SVM	BCIc.III-4a (**B**)	[128]
C3 and C4 *a-priori* selection, 1 s overlapping windows, BP in [8, 14], [14, 27] and [27, 45] Hz	Temporal absolute variations and segment-based bispectrum sums		SVM	BCIc.IV-2b, BCIc.III-4a (**B**), Non-public	[92]
Time domain staging and discrete Fourier transform-based decomposition.	Complex CSP	Energy and power-based selection, CSP sorting, Gaussian weighting, within-class scatter matrix, eigenvalues regularization.	Mahalanobis distance	BCIc.IV-2a (**B**), BCIc.IV-2b (**B**), BCIc.III-3a (**B**)	[129]
Time windows selection and BP in [8, 30] Hz	CSP	LASSO	SVM	BCIc.IV-1 (**B**), BCIc.III-3a, BCIc.III-4a	[130]
Down-sampling, channel selection, BP in [8, 30] Hz, multivariate wavelet transform, arrangement in vertical and horizontal sub-bands	CSP		LDA	BCIc.IV-1 (**B**), GS	[106]

(Continued.)

Table 3. (Continued.)

Pre-processing	Feature extraction	Feature selection	Classification	Dataset	
Overlapping time windows selection, FB in [4, 40] Hz	CSP and multiple-instance logistic regression	LASSO	SVM	BCIc.IV-2a (**B**)	[131]
BP in [4, 40] Hz	Local activities estimation and regularized CSP	Power spectral density in [8, 30] Hz with 4 Hz resolution	SVM	BCIc.IV-1 (**B**), BCIc.IV-2a, BCIc.III-4a, GS	[132]
15 constant-Q filters	CSP plus MAV, variance, and RMS in time domain, plus AR and PSR	ReliefF, Minimum redundancy maximum relevance, Fisher's method	SVM	BCIc.IV-2b (**B**)	[133]
C3, Cz and C4 *a-priori* selection, 2.0 s time windows with 1.9 s overlap, energy spectrum by wavelet packet decomposition and CDSF, FB in [4, 40] Hz	CSP, STDF		Ensemble SVM	BCIc.IV-2a, BCIc.IV-2b (**B**)	[119]
C3 and C4 *a-priori* selection, trial segmentation	STDF, wavelet packet decomposition		SVM	BCIc.IV-2a, BCIc.IV-2b (**B**)	[105]
Overlapping time windows selection, BP in [8, 40] Hz	CSP and AR	Kullback-Leibler divergence or Mutual information	NBPW	BCIc.IV-2a (**M**), BCIc.IV-2b	[134]

Table 4. Most performing processing strategies using BI approach. **B** and **M** indicate the dataset for which the proposal has exceeded the 75th percentile of accuracies distribution for the binary and multi-class cases, respectively (see figures 6 and 7). * outlier in terms of classification performance. BP: band-pass, CNN: convolutional neural network, CSP: common spatial pattern, FB: filter bank, ICA: independent component analysis, LDA: linear discriminant analysis, LGAN: long short-term memory generative adversarial network, NN: neural network. BCIc.: BCI competition, HG: high gamma, ULM: upper limb movements.

Pre-processing	Feature extraction	Feature selection	Classification	Dataset	
2D continuous wavelet transform	CNN pre-trained by AlexNet on ImageNet dataset			BCIc.III-4a (**B**)	[107]
Multi-scale principal component analysis	Successive decomposition index		Feed-forward NN	BCIc.III-4a (**B**)	[135]
BP in [8, 30] Hz	CSP	Particle swarm optimization and extreme learning machine		BCIc.IV-2a (**B**), BCIc.III-3a (**B**)	[136]
Channel selection, FB in [8, 32] Hz and in [6, 30] Hz	CSP and LDA	Sequential backward floating selection	Radial basis function NN	BCIc.III-4a (**B**), Non-public	[94]
Time window selection, BP in [8, 30] Hz	Artifact rejected binary CSP	Self-regulated supervised Gaussian fuzzy adaptive system Art		BCIc.IV-2a (**B**)	[137]
Data augmentation	EEG-inception network (CNN based)			BCIc.IV-2b (**B**), BCIc.IV-2a (**M**)	[32]
Data augmentation with a circular translation strategy	Feature separation network based on adversarial learning			BCIc.IV-2b (**B**), BCIc.IV-2a (**M**)	[31]
FastICA, BP in [4, 50] Hz, data augmentation with LGAN	Wavelet and fast Fourier Transform	Multi-output CNN and attention network		BCIc.IV-2b* (**B**), BCIc.IV-2a (**M**)	[33]
Time window selection, BP in [0, 40] Hz, signal normalization, Morlet wavelet transform	Multiscale space-time-frequency feature-guided multitask learning CNN			BCIc.IV-2b (**B**), BCIc.IV-2a (**M**), HG	[122]
Data augmentation with a circular translation strategy	CNN framework based on the discriminative feature learning strategy			BCIc.IV-2b (**B**), BCIc.IV-2a	[124]
A-priori channel selection, common average referencing, subject-specific time window selection, FB in [7, 30] Hz	CSP	F-score	Spiking NN	BCIc.IV-2a, BCIc.III-3a (**M**)	[108]
FB in [4, 40] Hz	Dipole source estimation, time of interest selection, coordinate transformation, construction of 4D matrix		Three-dimensional CNN	BCIc.IV-2a (**M**), ULM	[138]
FB in 43 bands	CSP	CNN		BCIc.IV-2a (**M**), BCIc.III-3a	[123]
μ and β bands	Multiple frequency CNN				
Overlapping time window selection, FB in [4, 38] Hz	CSP	CNN plus long short-term memory		BCIc.IV-2a (**M**)	[121]

convolutional layers were mostly exploited. Recurrent neural networks like the long short-term memory [121] and adversarial-based methods [31] were used as well. In the multi-class cases, the mean classification accuracy of most performing deep neural networks resulted equal to 85% with 4% uncertainty, while non-deep artificial neural networks did not result among the most performing ones. Instead, non-deep artificial neural networks are associated with a mean classification accuracy equal to 93% (and mean uncertainty 3%) in the binary cases. This should be compared to the 88 $\pm$ 4% associated with deep neural networks.

Overall, if taking into account the uncertainties, classification accuracy is relatively compatible for the different BI approaches. However, there is a difference in terms of mean accuracy. Finally, particularized considerations about repeatability and reproducibility are prevented due to the lack of information for each sub-field of BI approaches. This is also better commented later on. In general, it should be noted that BI approaches are much more effective in multi-class classification than non-BI approaches. Typically, for deep approaches the raw EEG signal [31, 32] or 2D or 3D arrays obtained by Fourier or wavelet transforms [33, 107, 122] were given as input. With a few exceptions [121, 123], features were extracted by hand mainly for the other methods. As for the non-BI approaches, CSP or FBCSP were the most exploited feature extraction algorithms.

In general, channel selection, time segmentation and band-pass filtering were largely used as pre-processing strategies for BI approaches. In addition, unlike non-BI approaches, the need for a larger amount of training data emerged. Half of the best proposals using convolutional networks exploited data augmentation strategies. All of these were deep approaches. The used strategies were: (i) circular translation strategy [31, 124], where samples are circularly shifted by a fixed time step, (ii) the extraction of noise from a trial and its application to another trial [32], and (iii) long short-term memory generative adversarial networks (LGANs), which use random noise and label as input to generated realistic MI data [33]. Transfer learning was adopted as well in [107]. This approach consisted of a CNN architecture (AlexNet) pre-trained on the ImageNet dataset. The idea was to transfer previous knowledge on image classification to feature extraction and classification of EEG. Indeed, the final CNN is fine-tuned on an EEG dataset, but with a limited number of training samples.

In concluding, it should be noted that the study [33] represents a positive outlier among those tested on the BCI Competition IV-2b dataset. In there, the authors proposed the usage of an ICA, a band-pass filter in the 4–50 Hz range, features extraction with wavelet and fast Fourier transforms, a multi-output convolutional neural-network, and an attention network for classifying the MI tasks. The proposal was implemented either with and without the LGAN data augmentation strategy. The highest performance was reached when data augmentation was applied (93%), while the accuracy dropped to 88% without data augmentation, which is still over the 75th percentile. The proposal of [33] thus appears compatible with other performant BI approaches, but the data augmentation would have been a key factor for a superior performance.

4.6. Online experiments

As already mentioned in section 4.1, few studies tested their proposal also on self-collected data. Only in [89] and [70], online experiments were also performed in order to test their algorithm in a more realistic condition. Both synchronous and asynchronous modalities were proposed in [89] to drive a robot. In synchronous control, the robot performed the movement only if the classification result matched the required task. On average, subjects took 41 s to complete 17 instructions. In asynchronous control, instead, the robot performed the movement only after two consecutive equal classification results. On average, subjects reached the predefined goal in 37 s, while no accuracy result was reported. In [70] three subjects exploited the MI to turn a wheelchair reaching a mean accuracy of 84%. The approaches proposed by these two studies were not included in the most performing approaches, even when tested on public datasets. However, all proposals should be tested with online experiments to assess whether they are suitable for real-time applications.

5. Discussion

In analyzing the included studies, state-of-the-art approaches and constitutive elements of a successful processing strategy were investigated. The overall aim is to consolidate knowledge and achievements in MI processing, and thus guide the reader in developing MI-BCIs. However, some limitations arise due to heterogeneity in terms of experimental setups and protocols, e.g. different tasks, variegated number and position of EEG channels, and diversified number of participants. For those reasons, the repeatability and reproducibility of the results are often precluded. On these premises, the current section discusses the achievements of the analysis, but it also addresses the need of further work facing open issues and challenges.

5.1. Achievements

The investigation of successful processing strategies was carried out in two steps. First, approaches and assessment methods were investigated by taking into account the included studies as a whole. Next, the most performing approaches were derived by considering widely used datasets. The reason for such a choice was to have enough evidence for a statistically-based comparison, namely a proper number of proposals to compare. Then, since most studies tested their proposals on multiple datasets, 87 studies out of 89 were considered in this second step because they exploited at least one of the mentioned datasets. Instead, the remaining two studies could not be taken into consideration because they only tested their proposal on the PhysioNet EEG dataset [114, 115]. Given the small number of proposals for that dataset, it was hence not possible to make a statistical analysis. Indeed, for the classification of right-hand versus left-hand imagery, these reported a performance compatible with the median performance achieved on dataset BCI competition IV-2a. However, a dataset-specific analysis should be carried out also for this dataset because accuracy is strongly dependent on the available subjects.

It is worth noting that the included studies for the binary cases are homogeneously distributed over the five years, while the majority of included studies for the multi-class case were published in the last two years. In both cases, the analysis pointed out that most performing approaches are concentrated in the last years. In details, it could be noted that five studies of 2021 [31–33, 122, 127] exceeded the 75th percentile both in the binary and in the multi-class case. Among them, four studies exploited a BI approach. Meanwhile, two other recent studies [129, 136] reached high performance in binary case on all the exploited datasets. Moreover, the approaches proposed in [136, 137] are among the best performing ones for classifying different binary combinations of MI tasks, but the performance reported for the multi-class case resulted below the respective median performance. Unfortunately, few proposals were tested on multiple public datasets and only two proposals were also tested online. Thus, further considerations are precluded by limited evidence.

The reported accuracy results demonstrated that compatible classification performance can be achieved with either BI or non-BI approaches, though nowadays there is an increasing interest in BI ones. Concerning non-BI approaches, literature already reports that the CSP is one of the most exploited feature extraction techniques indeed [37]. In addition to that, the current analysis demonstrated that it is also associated with successful classification. Moreover, the CSP is often used in combination with LASSO and SVM for features selection and classification, respectively. Notably, such approaches attempt to exploit available knowledge about the underlying neurophysiological phenomena. This limits the success of non-BI approaches to binary classification. Then, as an additional drawback, meticulous feature extraction and selection seem required. Meanwhile, the main advantages of these approaches are (i) ease of implementation, (ii) reduced training time, (iii) less parameters, and (iv) reliability even on small datasets.

On the contrary, BI approaches represent a new trend that does not always require *a-priori* knowledge in signal processing. Within BI approaches, two main areas can be distinguished: spiking neural networks and artificial neural networks. Spiking neural networks seem appealing today because they attempt a close reproduction of the human brain operation. Their main advantages are continuous real-time computation capability and the ability to capture multiple dimensions of information (e.g. time, space, frequency, phase) into a single model, as well as to handle large data volumes. Interestingly, they are suitable for hardware implementation and online operation [139]. It is worth noting that the authors only exploited the spiking neural network as a mere classifier but they also recognize its potential as a unique pipeline. Despite that, a single proposal among the best ones [108] was applying this strategy, while others BI proposals used artificial neural networks. Like non-BI approaches, non-deep artificial networks appear limited to binary classification and they rely on more elaborate feature extraction. However, they are easy to implement since they have a limited layers of neurons and consequently few parameters to set. Another trend is the usage of deep learning strategies, namely neural networks with many hidden layers. Especially in full-deep strategies, features are no longer designed, but the network for extraction and classification is built from data [140].

This analysis demonstrated that deep BI approaches outperform the other approaches in multi-class MI classification. These also achieve results compatible with the other approaches in binary classification. Among the benefits of deep strategies, they can accomplish feature extraction, selection, and classification as a unique pipeline. Moreover, raw data can be directly fed into them, even with little or no pre-processing. However, as a drawback, a huge number of parameters must be trained, and this implies a training time increase if compared with other approaches. For this reason, they also require appropriate hardware and they may be unsuitable for online operation. Finally, such an approach requires a large amount of data, which is not easily available in the BCI field. As shown in table 4, the two main strategies adopted to mitigate this issue were data augmentation techniques and transfer learning. Data augmentation has proven to be a powerful tool that increases network robustness and addresses the problem of overfitting when using small data sets. Instead, transfer learning harnesses previously learned knowledge and is extremely useful for reducing network training time and computational cost. Moreover, it is a promising tool in facing one of the main drawbacks of MI-based BCIs, namely the reduction of system calibration for a new subject. Overall, BI strategies appear promising in classifying more than two tasks, which is a significant challenge in the MI-BCI field [11]. Indeed, figure 9 reports that most of the best proposals for the multi-class case fall into the BI category. In particular, spiking neural networks and deep strategies are the most recommended in order to increase the number of commands of a MI-BCI. However, spiking neural networks seems to be most suitable for online experiments as they are faster and more dynamic. Thus, non-BI approaches are well established in the literature and they can be exploited for binary classification. Instead, BI approaches are really promising and allow to answer the main challenges of a MI-BCI such as the need to correctly classify multiple commands in real time. The reader is addressed to explore the pipelines reported in tables 3 and 4 for a successful classification of EEG in MI BCI.

5.2. Open issues

MI classification depends on the available EEG signals indeed. In wanting to replicate the results of a certain strategy, acquisition settings should be under control. Therefore, as a basis for the analysis, public datasets were mainly exploited. They usually report many details on the EEG data acquisition, and it was surely possible to rigorously quantify classification performance in terms of accuracy, uncertainty, and repeatability of the results. However, as already mentioned, current studies often lack reproducibility because experimental conditions are not always completely reported. Moreover, by referring to table 1, only BCI competition datasets were mostly used. Instead, further datasets should be exploited as well, since MI-related EEG data is strongly affected by subjects' training, and validation on a wider sample of subjects would be needed for proposed processing strategies. Ideally, researchers should test the proposed methods either on benchmark datasets like BCI competition ones, other public data, and their own experiments especially online. This would allow comparison with already existing state-of-the-art techniques, but limitations would also be better investigated by analyzing how different performance is associated with different conditions. In these regards, table 2 resumed the main factors for MI-related EEG acquisition, and it could be used as a guide for future experiments if willing to take control over experimental settings. In addition to data-related issues, the need of standardization arises for performance assessment methods too. The literature reports a variegated number of assessment methods, e.g. different cross-validation strategies or different data splits in hold-out methods. This heterogeneity is also limiting results comparison. Interestingly, many recent studies are starting to consider both cross-validation for the definition of the algorithm model (e.g. hyperparameters optimization) and its validation, but also testing on independent data with a hold-out strategy. Surely, challenges are associated with that, e.g. because of EEG non-stationarity, but this also enables a better understanding of MI processing. Concerning standardization, the BCI community has been discussing for years the need to share common paradigms to boost technological advances. Among several initiatives, the Mother of all BCI Benchmarks project is noteworthy because it proposes an open-source software suite as a common framework with benchmark data and state-of-the art processing algorithms [141]. These kind of initiatives would ease throughout comparisons and development of new algorithms, though such frameworks are still poorly exploited. As a final consideration, it is worth emphasizing that classification accuracy is considered almost exclusively when reporting the performance of an MI-BCI. Associated standard deviation and/or uncertainty of the results are often not explicitly reported. Moreover, other metrics are used for BCI systems for assessing speed and/or latency performance, but these are rarely exploited in MI-BCIs. This is justified by the fact that MI detection still requires relatively long time windows for acquisition and processing if compared to other BCI paradigms [38]. However, metrics such as the information transfer rate (ITR) could be interesting to investigate in a MI-BCI as already done in further BCI systems [11]. Reducing the target detection time is one of the objectives of BCI systems to enhance the ITR [2], and it would interesting to highlight control application that could exploit the MI-related phenomena.

6. Conclusions

In this review, an analysis of the state of the art was carried out to consolidate achievements in MI-BCI, with particular regards to the development of a successful classification approach for electroencephalographic signals. The included studies span from 2017 to 2021 and they were selected according to the PRISMA standard. The reported results were compared by means of the rigorous metrological framework defined by the VIM, with a particular focus on uncertainty, repeatability, and reproducibility of the results. The statistically-based analysis highlights that most performing approaches achieved at least 85% and 83% accuracy in classifying EEG signals associated with MI, respectively in the binary and multi-class cases. Essentially, uncertainty calculation demonstrates that most of the best performing approaches lead to compatible results since their associated uncertainty is up to 6%. Then, repeatability of the results on a predetermined dataset is up to 8%, while reproducibility is up to 15%, though its assessment was limited by the datasets effectively used in the included studies.

Increasing attention has been given recently to multi-class discrimination. Promising trends can be mostly distinguished in terms of non-BI and BI approaches. In the former case, a combination of

CSP (or a variant of it), LASSO, and SVM is commonly used and successful in terms of performance, though the specific pre-processing strategies can significantly impact the result. In the latter case, full deep approaches are more and more exploited, while other artificial neural networks or spiking neural networks have been successfully employed too. Among deep approaches, convolutional neural networks are largely exploited indeed. Notably, data augmentation strategies are used to fulfil their need for large amounts of data.

Lastly, this review highlights some issues and challenges that could be addressed in the next feature. Overall, limitations emerge due to heterogeneity in terms of different motor tasks, variegated number and position of EEG channels, and the subjects sample. In particular, assessing the reproducibility of results should be improved by standardizing experimental procedures and performance assessment methods. In conclusion, some recommendation for future developments would be to (i) test the proposals on multiple public datasets other than own data or BCI competitions ones, (ii) take main experimental settings under control (e.g, using the table 2 as a guide), (iii) test the proposals online, and (iv) define standard assessment methods.

Data availability statement

No new data were created or analysed in this study.

Funding

This work was carried out as part of the 'ICT for Health' project, which was financially supported by the Italian Ministry of Education, University and Research (MIUR), under the initiative 'Departments of Excellence' (Italian Budget Law No. 232/2016), through an excellence grant awarded to the Department of Information Technology and Electrical Engineering of the University of Naples Federico II, Naples, Italy.

Conflict of interest

The authors declare no conflict of interests concerning the publication of this manuscript.

ORCID iDs

Pasquale Arpaia https://orcid.org/0000-0002-5192-5922
Antonio Esposito https://orcid.org/0000-0002-3761-2401
Angela Natalizio https://orcid.org/0000-0001-6261-7741
Marco Parvis https://orcid.org/0000-0002-9723-6188

References

[1] Rak R J, Kołodziej M and Majkowski A 2012 Brain-computer interface as measurement and control system the review paper *Metrol. Meas. Syst.* **XIX** 427–44
[2] Ramadan R A and Vasilakos A V 2017 Brain computer interface: control signals review *Neurocomputing* **223** 26–44
[3] Shih J J, Krusienski D J and Wolpaw J R 2012 Brain-computer interfaces in medicine *Mayo Clinic Proc.* vol 87 (Amsterdam: Elsevier) pp 268–79
[4] Yadav D, Yadav S and Veer K 2020 A comprehensive assessment of brain computer interfaces: recent trends and challenges *J. Neurosci. Methods* **346** 108918
[5] Zander T O, Kothe C, Welke S and Rötting M 2009 Utilizing secondary input from passive brain-computer interfaces for enhancing human-machine interaction *Int. Conf. on Foundations of Augmented Cognition* (Springer) pp 759–71
[6] Picton T W, John M S, Dimitrijevic A and Purcell D 2003 Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos *Int. J. Audiol.* **42** 177–219
[7] Vialatte F B, Maurice M, Dauwels J and Cichocki A 2010 Steady-state visually evoked potentials: focus on essential paradigms and future perspectives *Prog. Neurobiol.* **90** 418–38
[8] Zander T O 2008 Enhancing human-machine systems with secondary input from passive brain-computer interfaces *Proc. 4th Int. BCI Workshop, 2008* pp 44–49
[9] Krol L R, Andreessen L M and Zander T O 2018 Passive brain–computer interfaces: a perspective on increased interactivity *Brain–Computer Interfaces Handbook* (Boca Raton, FL: CRC Press) pp 69–86
[10] Thurlings M E, van Erp J B, Brouwer A M and Werkhoven P J 2010 EEG-based navigation from a human factors perspective *Brain-Computer Interfaces* (Berlin: Springer) pp 71–86
[11] Singh A, Hussain A A, Lal S and Guesgen H W 2021 A comprehensive review on critical issues and possible solutions of motor imagery based electroencephalography brain-computer interface *Sensors* **21** 2173
[12] Lotze M and Halsband U 2006 Motor imagery *J. Physiol.-Paris* **99** 386–95
[13] Pfurtscheller G and Da Silva F L 1999 Event-related EEG/MEG synchronization and desynchronization: basic principles *Clin. Neurophysiol.* **110** 1842–57
[14] McFarland D J, Miner L A, Vaughan T M and Wolpaw J R 2000 Mu and beta rhythm topographies during motor imagery and actual movements *Brain Topography* **12** 177–86
[15] Chaudhary U, Birbaumer N and Ramos-Murguialday A 2016 Brain–computer interfaces for communication and rehabilitation *Nat. Rev. Neurol.* **12** 513
[16] Ang K K and Guan C 2016 EEG-based strategies to detect motor imagery for control and rehabilitation *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 392–401
[17] Frolov A A, Mokienko O, Lyukmanov R, Biryukova E, Kotov S, Turbina L, Nadareyshvily G and Bushkova Y 2017 Post-stroke rehabilitation training with a motor-imagery-based brain-computer interface (BCI)-controlled hand exoskeleton: a randomized controlled multicenter trial *Front. Neurosci.* **11** 400
[18] Baniqued P D E, Stanyer E C, Awais M, Alazmani A, Jackson A E, Mon-Williams M A, Mushtaq F and Holt R J 2021 Brain–computer interface robotics for hand rehabilitation after stroke: a systematic review *J. NeuroEng. Rehabil.* **18** 1–25
[19] Li J, Liang J, Zhao Q, Li J, Hong K and Zhang L 2013 Design of assistive wheelchair system directly steered by human thoughts *Int. J. Neural Syst.* **23** 1350013
[20] Bascil M S, Tesneli A Y and Temurtas F 2016 Spectral feature extraction of EEG signals and pattern recognition during mental tasks of 2-D cursor movements for BCI

using SVM and ANN *Australas. Phys. Eng. Sci. Med.* **39** 665–76
[21] Chakladar D D and Chakraborty S 2018 Multi-target way of cursor movement in brain computer interface using unsupervised learning *Biol. Inspired Cogn. Archit.* **25** 88–100
[22] Cao L, Xia B, Maysam O, Li J, Xie H and Birbaumer N 2017 A synchronous motor imagery based neural physiological paradigm for brain computer interface speller *Front. Hum. Neurosci.* **11** 274
[23] Paszkiel S 2016 Control based on brain-computer interface technology for video-gaming with virtual reality techniques *J. Autom. Mob. Robot. Intell. Syst.* **10** 3–7
[24] Bonnet L, Lotte F and Lécuyer A 2013 Two brains, one game: design and evaluation of a multiuser BCI video game based on motor imagery *IEEE Trans. Comput. Intell. AI Games* **5** 185–98
[25] Hortal E, Planelles D, Costa A, Iánez E, Úbeda A, Azorín J M and Fernández E 2015 SVM-based brain–machine interface for controlling a robot arm through four mental tasks *Neurocomputing* **151** 116–21
[26] Leeb R, Lee F, Keinrath C, Scherer R, Bischof H and Pfurtscheller G 2007 Brain–computer communication: motivation, aim and impact of exploring a virtual apartment *IEEE Trans. Neural Syst. Rehabil. Eng.* **15** 473–82
[27] Donati A R *et al* 2016 Long-term training with a brain-machine interface-based gait protocol induces partial neurological recovery in paraplegic patients *Sci. Rep.* **6** 30383
[28] Roc A, Pillette L, Mladenovic J, Benaroch C, N'Kaoua B, Jeunet C and Lotte F 2020 A review of user training methods in brain computer interfaces based on mental tasks *J. Neural Eng.* **18** 1
[29] Saha S and Baumert M 2020 Intra-and inter-subject variability in EEG-based sensorimotor brain computer interface: a review *Front. Computat. Neurosci.* **13** 87
[30] Padfield N, Zabalza J, Zhao H, Masero V and Ren J 2019 EEG-based brain-computer interfaces using motor-imagery: techniques and challenges *Sensors* **19** 1423
[31] Yang L, Song Y, Ma K, Su E and Xie L 2021 A novel motor imagery EEG decoding method based on feature separation *J. Neural Eng.* **18** 036022
[32] Zhang C, Kim Y K and Eskandarian A 2021 EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification *J. Neural Eng.* **18** 046014
[33] Xie J, Chen S, Zhang Y, Gao D and Liu T 2021 Combining generative adversarial networks and multi-output CNN for motor imagery classification *J. Neural Eng.* **18** 046026
[34] Kawala-Sterniuk A, Browarska N, Al-Bakri A, Pelc M, Zygarlicki J, Sidikova M, Martinek R and Gorzelanczyk E J 2021 Summary of over fifty years with brain-computer interfaces—a review *Brain Sci.* **11** 43
[35] Mudgal S K, Sharma S K, Chaturvedi J and Sharma A 2020 Brain computer interface advancement in neurosciences: applications and issues *Interdiscip. Neurosurg.* **20** 100694
[36] Minguillon J, Lopez-Gordo M A and Pelayo F 2017 Trends in EEG-BCI for daily-life: requirements for artifact removal *Biomed. Signal Process. Control* **31** 407–18
[37] Rashid M *et al* 2020 Current status, challenges and possible solutions of EEG-based brain-computer interface: a comprehensive review *Front. Neurorobot.* **14** 25
[38] Saha S, Mamun K A, Ahmed K I U, Mostafa R, Naik G R, Darvishi S, Khandoker A H and Baumert M 2021 Progress in brain computer interface: challenges and potentials *Front. Syst. Neurosci.* **15** 4
[39] Tariq M, Trivailo P M and Simic M 2018 EEG-based BCI control schemes for lower-limb assistive-robots *Front. Hum. Neurosci.* **12** 312
[40] Khan M A, Das R, Iversen H K and Puthusserypady S 2020 Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: from designing to application *Comput. Biol. Med.* **123** 103843
[41] Wierzgała P, Zapała D, Wojcik G M and Masiak J 2018 Most popular signal processing methods in motor-imagery BCI: a review and meta-analysis *Front. Neuroinform.* **12** 78
[42] Craik A, He Y and Contreras-Vidal J L 2019 Deep learning for electroencephalogram (EEG) classification tasks: a review *J. Neural Eng.* **16** 031001
[43] Al-Saegh A, Dawwd S A and Abdul-Jabbar J M 2021 Deep learning for motor imagery EEG-based classification: a review *Biomed. Signal Process. Control* **63** 102172
[44] Brunner P, Bianchi L, Guger C, Cincotti F and Schalk G 2011 Current trends in hardware and software for brain–computer interfaces (BCIs) *J. Neural Eng.* **8** 025001
[45] Jasper H H 1958 The ten-twenty electrode system of the international federation *Electroencephalogr. Clin. Neurophysiol.* **10** 370–5
[46] Jurcak V, Tsuzuki D and Dan I 2007 10/20, 10/10 and 10/5 systems revisited: their validity as relative head-surface-based positioning systems *Neuroimage* **34** 1600–11
[47] Teplan M *et al* 2002 Fundamentals of EEG measurement *Meas. Sci. Rev.* **2** 1–11
[48] Ienca M, Haselager P and Emanuel E J 2018 Brain leaks and consumer neurotechnology *Nat. Biotechnol.* **36** 805–10
[49] Li G, Wu J, Xia Y, He Q and Jin H 2020 Review of semi-dry electrodes for EEG recording *J. Neural Eng.* **17** 5
[50] Liao L D, Wang I J, Chen S F, Chang J Y and Lin C T 2011 Design, fabrication and experimental validation of a novel dry-contact sensor for measuring electroencephalography signals without skin preparation *Sensors* **11** 5819–34
[51] Fiedler P, Pedrosa P, Griebel S, Fonseca C, Vaz F, Supriyanto E, Zanow F and Haueisen J 2015 Novel multipin electrode cap system for dry electroencephalography *Brain Topography* **28** 647–56
[52] Lin C T, Liao L D, Liu Y H, Wang I J, Lin B S and Chang J Y 2010 Novel dry polymer foam electrodes for long-term EEG measurement *IEEE Trans. Biomed. Eng.* **58** 1200–7
[53] Arpaia P, Callegaro L, Cultrera A, Esposito A and Ortolano M 2021 Metrological characterization of a low-cost electroencephalograph for wearable neural interfaces in industry 4.0 applications *IEEE Int. Workshop on Metrology for Industry 4,0 & IoT, 7-9 June, Virtual Conf.* (IEEE) p 000
[54] Kam J W, Griffin S, Shen A, Patel S, Hinrichs H, Heinze H J, Deouell L Y and Knight R T 2019 Systematic comparison between a wireless EEG system with dry electrodes and a wired EEG system with wet electrodes *NeuroImage* **184** 119–29
[55] Hinrichs H, Scholz M, Baum A K, Kam J W, Knight R T and Heinze H J 2020 Comparison between a wireless dry electrode EEG system with a conventional wired wet electrode EEG system for clinical applications *Sci. Rep.* **10** 5218
[56] Wang F, Li G, Chen J, Duan Y and Zhang D 2016 Novel semi-dry electrodes for brain–computer interface applications *J. Neural Eng.* **13** 046021
[57] Halford J J, Schalkoff R J, Satterfield K E, Martz G U, Kutluay E, Waters C G and Dean B C 2016 Comparison of a novel dry electrode headset to standard routine EEG in veterans *J. Clin. Neurophysiol.* **33** 530–7
[58] Sajda P, Gerson A, Muller K R, Blankertz B and Parra L 2003 A data analysis competition to evaluate machine learning algorithms for use in brain-computer interfaces *IEEE Trans. Neural Syst. Rehabil. Eng.* **11** 184–5
[59] Blankertz B *et al* 2004 The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials *IEEE Trans. Biomed. Eng.* **51** 1044–51
[60] Blankertz B, Muller K R, Krusienski D J, Schalk G, Wolpaw J R, Schlogl A, Pfurtscheller G, Millan J R, Schroder M and Birbaumer N 2006 The BCI competition III: validating alternative approaches to actual BCI problems *IEEE Trans. Neural Syst. Rehabil. Eng.* **14** 153–9
[61] Tangermann M *et al* 2012 Review of the BCI competition IV *Front. Neurosci.* **6** 55

[62] Jia T, Liu K, Qian C, Li C and Ji L 2020 Denoising algorithm for event-related desynchronization-based motor intention recognition in robot-assisted stroke rehabilitation training with brain-machine interaction *J. Neurosci. Methods* **346** 108909

[63] Duan L, Li J, Ji H, Pang Z, Zheng X, Lu R, Li M and Zhuang J 2020 Zero-shot learning for EEG classification in motor imagery-based BCI system *IEEE Trans. Neural Syst. Rehabil. Eng.* **28** 2411–9

[64] Casso M I, Jeunet C and Roy R N 2021 Heading for motor imagery brain-computer interfaces (MI-BCIs) usable out-of-the-lab: impact of dry electrode setup on classification accuracy *2021 10th Int. IEEE/EMBS Conf. on Neural Engineering (NER)* (IEEE) pp 690–3

[65] Olivas-Padilla B E and Chacon-Murguia M I 2019 Classification of multiple motor imagery using deep convolutional neural networks and spatial filters *Appl. Soft Comput.* **75** 461–72

[66] Nicolas-Alonso L F and Gomez-Gil J 2012 Brain computer interfaces, a review *Sensors* **12** 1211–79

[67] Peterson V, Wyser D, Lambercy O, Spies R and Gassert R 2019 A penalized time-frequency band feature selection and classification procedure for improved motor intention decoding in multichannel EEG *J. Neural Eng.* **16** 016019

[68] Chen C *et al* 2020 Neural activities classification of left and right finger gestures during motor execution and motor imagery *Brain-Comput. Interfaces* **8** 1–11

[69] Mondini V, Mangia A L and Cappello A 2016 EEG-based BCI system using adaptive features extraction and classification procedures *Computat. Intell. Neurosci.* **2016** 1–14

[70] Tang X, Li W, Li X, Ma W and Dang X 2020 Motor imagery EEG recognition based on conditional optimization empirical mode decomposition and multi-scale convolutional neural network *Expert Syst. Appl.* **149** 113285

[71] Brunner C, Leeb R, Müller-Putz G, Schlögl A and Pfurtscheller G 2008 BCI Competition 2008–Graz data set A *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology* **16** 1–6

[72] Ahn M and Jun S C 2015 Performance variation in motor imagery brain–computer interface: a brief review *J. Neurosci. Methods* **243** 103–10

[73] Jeunet C, Glize B, McGonigal A, Batail J M and Micoulaud-Franchi J A 2019 Using EEG-based brain computer interface and neurofeedback targeting sensorimotor rhythms to improve motor skills: theoretical background, applications and prospects *Neurophysiol. Clin.* **49** 125–36

[74] Leeb R *et al* 2013 Freeing the visual channel by exploiting vibrotactile bci feedback *2013 35th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) pp 3093–6

[75] Jeunet C, Vi C, Spelmezan D, N'Kaoua B, Lotte F and Subramanian S 2015 Continuous tactile feedback for motor-imagery based brain-computer interaction in a multitasking context *IFIP Conf. on Human-Computer Interaction* (Springer) pp 488–505

[76] Wang Z, Zhou Y, Chen L, Gu B, Liu S, Xu M, Qi H, He F and Ming D 2019 A BCI based visual-haptic neurofeedback training improves cortical activations and classification performance during motor imagery *J. Neural Eng.* **16** 066012

[77] Fleury M, Lioi G, Barillot C and Lécuyer A 2020 A survey on the use of haptic feedback for brain-computer interfaces and neurofeedback *Front. Neurosci.* **14** 528

[78] Sollfrank T, Ramsay A, Perdikis S, Williamson J, Murray-Smith R, Leeb R, Millán J and Kübler A 2016 The effect of multimodal and enriched feedback on SMR-BCI performance *Clin. Neurophysiol.* **127** 490–8

[79] Page M J, McKenzie J E, Bossuyt P M, Boutron I, Hoffmann T C, Mulrow C D, Shamseer L, Tetzlaff J M and Moher D 2021 Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement *J. Clin. Epidemiol.* **134** 103–12

[80] Vim: Int. vocabulary of metrology (available at: www.bipm.org/en/committees/jc/jcgm/publications)

[81] Evaluation of measurement data—guide to the expression of uncertainty in measurement (available at: www.bipm.org/en/committees/jc/jcgm/publications)

[82] Cho H, Ahn M, Ahn S, Kwon M and Jun S C 2017 EEG datasets for motor imagery brain–computer interface *GigaScience* **6** gix034

[83] Schirrmeister R T, Springenberg J T, Fiederer L D J, Glasstetter M, Eggensperger K, Tangermann M, Hutter F, Burgard W and Ball T 2017 Deep learning with convolutional neural networks for EEG decoding and visualization *Human Brain Mapp.* **38** 5391–420

[84] Schalk G, McFarland D J, Hinterberger T, Birbaumer N and Wolpaw J R 2004 Bci2000: a general-purpose brain-computer interface (BCI) system *IEEE Trans. Biomed. Eng.* **51** 1034–43

[85] Nikolopoulos S *et al* 2017 A multimodal dataset for authoring and editing multimedia content: the MAMEM project *Data Brief* **15** 1048–56

[86] Ofner P, Schwarz A, Pereira J and Müller-Putz G R 2017 Upper limb movements can be decoded from the time-domain of low-frequency EEG *PLoS One* **12** e0182578

[87] Steyrl D, Scherer R, Förstner O and Müller-Putz G R 2014 Motor imagery brain-computer interfaces: random forests vs regularized lda-non-linear beats linear *Proc. 6th Int. Brain-Computer Interface Conf.* pp 241–4

[88] Tang R, Li Z and Xie X 2021 Motor imagery EEG signal classification using upper triangle filter bank auto-encode method *Biomed. Signal Process. Control* **68** 102608

[89] Ai Q, Chen A, Chen K, Liu Q, Zhou T, Xin S and Ji Z 2019 Feature extraction of four-class motor imagery EEG signals based on functional brain network *J. Neural Eng.* **16** 026032

[90] Xie X, Yu Z L, Lu H, Gu Z and Li Y 2016 Motor imagery classification based on bilinear sub-manifold learning of symmetric positive-definite matrices *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 504–16

[91] Sun B, Zhao X, Zhang H, Bai R and Li T 2020 EEG motor imagery classification with sparse spectrotemporal decomposition and deep learning *IEEE Trans. Autom. Sci. Eng.* **18** 541–51

[92] Sun L, Feng Z, Lu N, Wang B and Zhang W 2019 An advanced bispectrum features for EEG-based motor imagery classification *Expert Syst. Appl.* **131** 9–19

[93] Miao M, Zhang W, Hu W and Wang R 2020 An adaptive multi-domain feature joint optimization framework based on composite kernels and ant colony optimization for motor imagery EEG classification *Biomed. Signal Process. Control* **61** 101994

[94] Bhatti M H, Khan J, Khan M U G, Iqbal R, Aloqaily M, Jararweh Y and Gupta B 2019 Soft computing-based EEG classification by optimal feature selection and neural networks *IEEE Trans. Ind. Inform.* **15** 5747–54

[95] Srinivasan R, Nunez P L, Tucker D M, Silberstein R B and Cadusch P J 1996 Spatial sampling and filtering of EEG with spline laplacians to estimate cortical potentials *Brain Topography* **8** 355–66

[96] Samuel O W, Geng Y, Li X and Li G 2017 Towards efficient decoding of multiple classes of motor imagery limb movements based on EEG spectral and time domain descriptors *J. Med. Syst.* **41** 1–13

[97] Zhang Y, Ji X and Zhang Y 2015 Classification of EEG signals based on ar model and approximate entropy *2015 Int. Joint Conf. on Neural Networks (IJCNN)* (IEEE) pp 1–6

[98] Xu L, Zhang H, Hui M, Long Z, Jin Z, Liu Y and Yao L 2014 Motor execution and motor imagery: a comparison of

functional connectivity patterns based on graph theory *Neuroscience* **261** 184–94
[99] Serdar Bascil M, Tesneli A Y and Temurtas F 2015 Multi-channel EEG signal feature extraction and pattern recognition on horizontal mental imagination task of 1-D cursor movement for brain computer interface *Australas. Phys. Eng. Sci. Med.* **38** 229–39
[100] Tabar Y R and Halici U 2016 A novel deep learning approach for classification of EEG motor imagery signals *J. Neural Eng.* **14** 016003
[101] Lee H K and Choi Y S 2019 Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface *Entropy* **21** 1199
[102] Ang K K, Chin Z Y, Wang C, Guan C and Zhang H 2012 Filter bank common spatial pattern algorithm on BCI competition iv datasets 2a and 2b *Front. Neurosci.* **6** 39
[103] Breiman L, Friedman J H, Olshen R A and Stone C J 2017 *Classification and Regression Trees* (Oxfordshire: Routledge)
[104] Hearst M A, Dumais S T, Osuna E, Platt J and Scholkopf B 1998 Support vector machines *IEEE Intell. Syst. Appl.* **13** 18–28
[105] Luo J, Feng Z and Lu N 2019 Spatio-temporal discrepancy feature for classification of motor imageries *Biomed. Signal Process. Control* **47** 137–44
[106] Molla M K I, Saha S K, Yasmin S, Islam M R and Shin J 2021 Trial regeneration with subband signals for motor imagery classification in BCI paradigm *IEEE Access* **9** 7632–42
[107] Chaudhary S, Taran S, Bajaj V and Sengur A 2019 Convolutional neural network based approach towards motor imagery tasks EEG signals classification *IEEE Sens. J.* **19** 4494–500
[108] Wang H, Tang C, Xu T, Li T, Xu L, Yue H, Chen P, Li J and Bezerianos A 2020 An approach of one-vs-rest filter bank common spatial pattern and spiking neural networks for multiple motor imagery decoding *IEEE Access* **8** 86850–61
[109] Roy Y, Banville H, Albuquerque I, Gramfort A, Falk T H and Faubert J 2019 Deep learning-based electroencephalography analysis: a systematic review *J. Neural Eng.* **16** 051001
[110] Alom M Z, Taha T M, Yakopcic C, Westberg S, Sidike P, Nasrin M S, Hasan M, Van Essen B C, Awwal A A and Asari V K 2019 A state-of-the-art survey on deep learning theory and architectures *Electronics* **8** 292
[111] Sze V, Chen Y H, Yang T J and Emer J S 2017 Efficient processing of deep neural networks: a tutorial and survey *Proc. IEEE* **105** 2295–329
[112] Alom M Z, Taha T M, Yakopcic C, Westberg S, Sidike P, Nasrin M S, Van Esesn B C, Awwal A A S and Asari V K 2018 The history began from alexnet: a comprehensive survey on deep learning approaches (arXiv:1803.01164)
[113] Schlogl A, Kronegg J, Huggins J and Mason S 2007 Evaluation criteria for BCI research *Toward Brain-Computer Interfacing* ch 19, p 327
[114] Fan C C, Yang H, Hou Z G, Ni Z L, Chen S and Fang Z 2021 Bilinear neural network with 3-D attention for brain decoding of motor imagery movements from the human EEG *Cogn. Neurodyn.* **15** 181–9
[115] Kim Y, You J, Lee H, Lee S M and Park C 2018 Correlation assisted strong uncorrelating transform complex common spatial patterns for spatially distant channel data *Computat. Intell. Neurosci.* **2018** 1–9
[116] Blankertz B, Tomioka R, Lemm S, Kawanabe M and Muller K R 2007 Optimizing spatial filters for robust EEG single-trial analysis *IEEE Signal Process. Mag.* **25** 41–56
[117] Shawe-Taylor J and Sun S 2011 A review of optimization methodologies in support vector machines *Neurocomputing* **74** 3609–18
[118] Malan N S and Sharma S 2021 Time window and frequency band optimization using regularized neighbourhood component analysis for Multi-View Motor Imagery EEG classification *Biomed. Signal Process. Control* **67** 102550
[119] Luo J, Gao X, Zhu X, Wang B, Lu N and Wang J 2020 Motor imagery EEG classification based on ensemble support vector learning *Comput. Methods Programs Biomed.* **193** 105464
[120] Tibshirani R 1996 Regression shrinkage and selection via the lasso *J. R. Stat. Soc.* B **58** 267–88
[121] Zhang R, Zong Q, Dou L and Zhao X 2019 A novel hybrid deep learning scheme for four-class motor imagery classification *J. Neural Eng.* **16** 066004
[122] Liu X, Lv L, Shen Y, Xiong P, Yang J and Liu J 2021 Multiscale space-time-frequency feature-guided multitask learning CNN for motor imagery EEG classification *J. Neural Eng.* **18** 026003
[123] Xue J, Ren F, Sun X, Yin M, Wu J, Ma C and Gao Z 2020 A multifrequency brain network-based deep learning framework for motor imagery decoding *Neural Plast.* **2020** 1–11
[124] Yang L, Song Y, Ma K and Xie L 2021 Motor imagery EEG decoding method based on a discriminative feature learning strategy *IEEE Trans. Neural Syst. Rehabil. Eng.* **29** 368–79
[125] Zarei R, He J, Siuly S and Zhang Y 2017 A PCA aided cross-covariance scheme for discriminative feature extraction from EEG signals *Comput. Methods Programs Biomed.* **146** 47–57
[126] Miao Y, Jin J, Daly I, Zuo C, Wang X, Cichocki A and Jung T P 2021 Learning common time-frequency-spatial patterns for motor imagery classification *IEEE Trans. Neural Syst. Rehabil. Eng.* **29** 699–707
[127] Rashid M, Bari B S, Hasan M J, Razman M A M, Musa R M, Ab Nasir A F and Majeed A P A 2021 The classification of motor imagery response: an accuracy enhancement through the ensemble of random subspace k-nn *PeerJ Comput. Sci.* **7** e374
[128] Wang F, Xu Z, Zhang W, Wu S, Zhang Y, Ping J and Wu C 2020 Motor imagery classification using geodesic filtering common spatial pattern and filter-bank feature weighted support vector machine *Rev. Sci. Instrum.* **91** 034106
[129] Mishuhina V and Jiang X 2021 Complex common spatial patterns on time-frequency decomposed EEG for brain-computer interface *Pattern Recognit.* **115** 107918
[130] Jiang J, Wang C, Wu J, Qin W, Xu M and Yin E 2020 Temporal combination pattern optimization based on feature selection method for motor imagery BCIS *Front. Hum. Neurosci.* **14** 231
[131] Collazos-Huertas D, Caicedo-Acosta J, Castaño-Duque G A and Acosta-Medina C D 2020 Enhanced multiple instance representation using time-frequency atoms in motor imagery classification *Front. Neurosci.* **14** 155
[132] Togha M M, Salehi M R and Abiri E 2021 An improved version of local activities estimation to enhance motor imagery classification *Biomed. Signal Process. Control* **66** 102485
[133] Radman M, Chaibakhsh A, Nariman-zadeh N and He H 2021 Feature fusion for improving performance of motor imagery brain-computer interface system *Biomed. Signal Process. Control* **68** 102763
[134] Wang J, Feng Z, Ren X, Lu N, Luo J and Sun L 2020 Feature subset and time segment selection for the classification of EEG data based motor imagery *Biomed. Signal Process. Control* **61** 102026
[135] Sadiq M T, Yu X, Yuan Z and Aziz M Z 2020 Identification of motor and mental imagery EEG in two and multiclass subject-dependent tasks using successive decomposition index *Sensors* **20** 5283
[136] Zhang L, Wen D, Li C and Zhu R 2020 Ensemble classifier based on optimized extreme learning machine for motor imagery classification *J. Neural Eng.* **17** 026004
[137] Jafarifarmand A, Badamchizadeh M A, Khanmohammadi S, Nazari M A and Tazehkand B M 2017 A new self-regulated neuro-fuzzy framework for classification of EEG signals in motor imagery BCI *IEEE Trans. Fuzzy Syst.* **26** 1485–97

[138] Li M A and Ruan Z W 2021 A novel decoding method for motor imagery tasks with 4D data representation and 3D convolutional neural networks *J. Neural Eng.* **18** 046029

[139] Lobo J L, Del Ser J, Bifet A and Kasabov N 2020 Spiking neural networks and online learning: an overview and perspectives *Neural Netw.* **121** 88–100

[140] Dargan S, Kumar M, Ayyagari M R and Kumar G 2020 A survey of deep learning and its applications: a new paradigm to machine learning *Arch. Computat. Methods Eng.* **27** 1071–92

[141] Jayaram V and Barachant A 2018 MOABB: trustworthy algorithm benchmarking for BCIs *J. Neural Eng.* **15** 066011