Neurocomputing 610 (2024) 128577

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Survey paper

Deep learning in motor imagery EEG signal decoding: A Systematic Review

Aurora Saibene [a,b,*], Hafez Ghaemi [c,d], Eda Dagdevir [e]

[a] Department of Informatics, Systems and Communication, University of Milano-Bicocca, Viale Sarca 336, Milano, 20126, Italy
[b] NeuroMI, Milan Center for Neuroscience, Piazza dell'Ateneo Nuovo 1, Milano, 20126, Italy
[c] Department of Computer Science and Operations Research, Université de Montréal, Montreal, QC, Canada
[d] Mila, Quebec Artificial Intelligence Institute, Montreal, QC, Canada
[e] Department of Electronics and Automation, Vocational School of Technical Sciences, Kayseri University, Kayseri, Turkey

ARTICLE INFO

Communicated by R. Yang

Keywords:
Brain–computer interface (BCI)
Motor imagery (MI)
Electroencephalography (EEG)
Deep learning (DL)

ABSTRACT

Thanks to the fast evolution of electroencephalography (EEG)-based brain–computer interfaces (BCIs) and computing technologies, as well as the availability of large EEG datasets, decoding motor imagery (MI) EEG signals is rapidly shifting from traditional machine learning (ML) to deep learning (DL) approaches. Furthermore, real-world MI-EEG BCI applications are progressively requiring higher generalization capabilities, which can be achieved by leveraging publicly available MI-EEG datasets and high-performance decoding models. Within this context, this paper provides a systematic review of DL approaches for MI-EEG decoding, focusing on studies that work on publicly available EEG-MI datasets. This review paper firstly provides a clear overview of these datasets that can be used for DL model training and testing. Afterwards, considering each dataset, related DL studies are discussed with respect to the four decoding paradigms identified in the literature, i.e., subject-dependent, subject-independent, transfer learning, and global decoding paradigms. Having analyzed the reviewed studies, the current trends and strategies, popular architectures, baseline models that are used for comprehensive analysis, and techniques to ensure reproducibility of the results in DL-based MI-EEG decoding are also identified and discussed. The selection and screening of the studies included in this review follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, leading to a comprehensive analysis of 394 papers published between January 1, 2017, and January 23, 2023.

1. Introduction

Electroencephalography (EEG)-based Brain–Computer Interfaces (BCIs) provide a real-time bidirectional communication between their user and an external application, directly exploiting neural signals acquired in a non-invasive manner from the user's scalp [1]. In fact, EEG signals are recorded through sensors, called electrodes, placed on the head surface. These data provide information on brain activity and functions in the time, frequency and spatial domains, while using technologies of low risk, ease of application, and affordability, thus justifying their diffused use for BCI systems [2].

A standard EEG-based BCI system usually consists of three main modules, i.e., signal acquisition, signal processing, and applications that can each contain different stages [3,4]. In particular, the signal processing module usually comprises data pre-processing for noise removal, feature engineering (especially for traditional decoding models), and classification, which can be performed with traditional machine learning (ML) methodologies or with novel deep learning (DL) approaches [4].

Fig. 1 shows a standard BCI system with a detail on a very simple configuration of a DL methodology encased in the signal processing module, starting from a simple input formulation, i.e., raw EEG signals, moving to a schematic representation of a DL architecture, and providing an outcome example.

The EEG signals are recorded with surface electrodes and processed in order to decode the neural information and provide commands for the desired application, and finally, feedback is provided to the user.

BCI applications are very diverse and can be devised for rehabilitation, orthopedic practice, wheelchair control, cursor control, writing systems, gaming, and neuromarketing purposes [5,6], benefiting not only healthcare-related research, but also fields spanning from education to entertainment, and from civil to industrial applications [7].

Therefore, depending on the intended use of a BCI, different experimental paradigms are exploited and in particular three are the ones mainly used [8]: (i) the P300 event-related potential, (ii) the steady-state visual evoked potential, and (iii) Motor Imagery (MI).

* Corresponding author at: Department of Informatics, Systems and Communication, University of Milano-Bicocca, Viale Sarca 336, Milano, 20126, Italy.
E-mail addresses: aurora.saibene@unimib.it (A. Saibene), hafez.ghaemi@umontreal.ca (H. Ghaemi), edadagdevir@kayseri.edu.tr (E. Dagdevir).

https://doi.org/10.1016/j.neucom.2024.128577
Received 14 September 2023; Received in revised form 6 July 2024; Accepted 11 September 2024
Available online 14 September 2024
0925-2312/

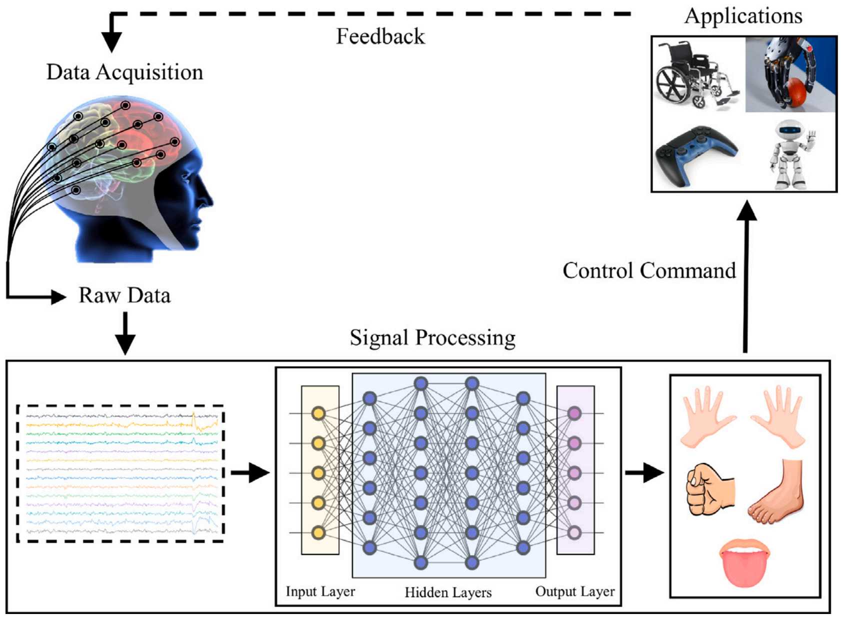


Fig. 1. The main components of a standard BCI system, with a detailed focus on the deep learning part of the signal processing module.

The latter paradigm has been attracting the attention of the BCI community thanks to the neural characteristics it exploits and its wide area of applicability [9]. MI-EEG signals are generated in response to the imagination of motor movements in the sensory motor cortex area of the brain [10]. MI tasks are associated with oscillations in the μ (∼8−13 Hz) and β (∼13−30 Hz) EEG frequency bands [11]. In particular, during MI, a decrease in the μ amplitude (known as event-related desynchronization) and an increase in the β amplitude (known as event-related synchronization) [10] is observed.

Due to its unique capability to decode motor movement imagination in the brain, MI EEG-based BCIs can be exploited in rehabilitation systems to improve disabled people's quality of life [12–15]. Moreover, new MI applications are developed for non-biomedical purposes such as gaming, industry, transport, and art [16].

Creating efficient MI EEG-based BCI systems has a number of challenges which are caused by the low signal-to-noise ratio (SNR) of EEG signals (due to biological or non-physiological noise), the non-stationary nature of MI-EEG signals and their subjectivity, and the differences due to changes in environmental and experimental conditions [17–19].

Therefore, fast and high-performance algorithms are needed to extract meaningful information from these complex signals and decode them. Traditionally, ML algorithms have been employed for MI-EEG decoding [20–23]. Conventional ML methods are usually able to achieve a good performance when the number of data samples is small, or when high generalization abilities are not required from the decoding model. However, in the recent decade, MI-EEG datasets with a large number of recordings have been introduced [24–29], making the decoding task challenging for conventional ML techniques. Furthermore, real-world BCI applications require generalizable decoding models adaptable to new subjects, a feature that traditional ML methods usually do not provide. These issues, alongside the exponential growth of the field of DL [30], the increasing ease of implementation for DL methods, and the availability of graphical processing units (GPUs) for their training have motivated the BCI research community to increase its efforts towards the use of DL-based methodologies to design BCI systems, including MI EEG-based BCIs for decoding motor intentions [9,16,31].

Some review papers have provided overviews on DL-based methodologies for EEG signal decoding [18,31], trying to understand which architectures are the most frequently used in DL-based decoding and in what experimental contexts, alongside the effects that the input type and preprocessing has on the model performance, and how reproducibilty is guaranteed in such applications. In the specific field of MI EEG-based BCI systems, answers to these questions are continuously sought. *Al-Saegh et al.* [9] focus their attention on studies using deep neural networks (DNNs). The authors examine the different datasets used in DL-based MI BCI, input types, preprocessing techniques, and considered frequency intervals related to MI-EEG data to understand which DL architectures work best for MI decoding, how the input data and type impacts the model performance, and which frequency ranges to consider to achieve better performances. This paper reviews 40 articles published between 2015 and 2020, and finds that convolutional neural networks (CNNs) and hybrid CNNs are the most dominant architectures, that usually receive in input raw EEG signals and consider frequencies in the ranges of 6–8 Hz and 38–40 Hz.

Arpaia et al. [32] provide a metrological analysis of the literature in terms of performances achieved in MI EEG-based BCI decoding, the processing trends and their similarities, and the challenges related to uncertainty, repeatability, and reproducibility of the results. The authors conduct their search by following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement [33] and analyze 89 papers published between 2017 and 2021. The outcome of their review can be summarized in three main statements: screened works have compatible results for binary and multi-class cases, trends can be identified in terms of non-brain-inspired and brain-inspired approaches, and reproducibility can be improved by standardizing experimental procedures and performance presentation.

Altaheri et al. [5] propose a similar set of research questions to the previous reviews and thus asks whether prepossessing techniques need to be applied before using DL methods, what input formulations should be used depending on the DL techniques, and what main trends exist in DL-based BCI. The PRISMA guidelines are followed and 89 papers in the last 10 years are selected for the review. The authors find that CNN architectures are widely diffused and that raw EEG signals are usually the input to such models.

A systematic review of MI EEG-based BCIs using only wearable devices is proposed by *Saibene et al.* [4]. Following the PRISMA guidelines, 84 articles published between 2012 and 2022 have been examined with the main aim of understanding if wearable technologies are sufficiently advanced to provide MI EEG-based BCI applications outside of controlled environments. The authors identify the importance of denoising strategies when dealing with wireless and wearable technologies, and highlight the strong presence of strategies that exploit handcrafted features and traditional ML approaches, reaching the conclusion that the use of DL is at an early stage in the wearable domain. A focus on MI experimental paradigms as well as publicly available datasets is also provided, noticing the wide diffusion of diverse MI conditions and the use of literature repositories for comparison with proprietary datasets.

Khademi et al. [34] are mainly interested in the challenges arising when creating MI EEG-based BCIs. The authors compare traditional ML techniques with novel DL approaches, which seem to be apt at dealing with complex and non-stationary signals. However, they highlight that models are usually trained in a subject-dependent manner due to the EEG signal subject-specific nature and the lack of large datasets which leads to the over-fitting phenomenon in DL models.

Wang et al. [35] systematically review 67 papers on DL-based MI-EEG classification methods and compare the performance of each of the reviewed models by running the model source codes provided by the authors of each study. By performing an ablation study on the network architecture, an evaluation is made with 13 representative design models for the most common network architectures. The study concludes that effective feature fusion is crucial for developing accurate multi-stream CNN architectures, LSTM alone is insufficient for classifying MI-EEG signals, dropout is not crucial to improve the model performance, and fully connected layers (besides the output layer) should not be used in classification models. Notice that the focus of the review is on the impact of DL design parameters on model performance and thus no insights across a wide range of datasets are provided.

Finally, *Lionakis et al.* [36] present trends, challenges, and future research directions of various DL techniques for MI-EEG based BCI systems. The authors favored DL approaches due to their ability in handling MI-EEG data spatial and temporal information. While the study provides an overview of DL and hybrid deep learning methods used in MI-BCI systems, the review offers limited considerations on the reproducibility of the reviewed techniques.

Summarizing, besides considering both the currently available reviews and the number of papers screened in our study, the use of DL methodologies is becoming increasingly popular thanks to technological advancements (e.g., better GPUs), the availability of DL libraries that can be easily incorporated in other source codes, the development of novel optimization techniques that allow a reduction of the time required for computation, and the possibility of letting the algorithm learn novel patterns for neural data interpretation. Moreover, DL approaches are benefiting from the ever increasing presence of datasets with a high number of EEG data, on which they usually provide better decoding performance compared to the traditional ML techniques.

Starting from these premises regarding EEG signals, BCIs, DL, and the MI experimental paradigm, and also considering the reported state-of-the-art survey papers, this work systematically reviews the DL approaches used for MI-EEG decoding, strictly following the PRISMA guidelines. Additionally, this review considers the research papers from the early 2017 until the early 2023 (period during which this review project started). In fact, 2017 is a landmark year for the EEG community working on DL approaches, having that it corresponds to the year of publication of important benchmarks in DL-based EEG decoding such as Shallow and Shallow/Deep ConvNet [37] as well as EEGNet [38].

Compared to other review papers, this study focuses on reproducibility by considering only works employing publicly available datasets and providing well documented DL approaches. Moreover, great attention is given to the type of decoding paradigms used, i.e., subject-dependent (SD), subject-independent (SI), global and transfer learning (TL) approaches. This will help early-career researchers by guiding them through different topics related to the proper development of these methodologies and provide references to possible models of interest. Due to all these contributions, the present work is a quite comprehensive and inclusive roadmap to MI-EEG decoding through DL approaches.

There are four main research questions driving this review paper, considering both the challenges identified by previous survey studies (e.g., [32,34]), and the ones identified by the authors of this review.

The first research question is formulated as follows,

RQ1: *What are the most frequently used publicly available datasets for MI EEG decoding, specifically DL-based decoding?*

To answer this question, all publicly available MI datasets that are used in the screened studies are identified and described, wanting to provide references to possible test beds for novel DL models and to allow comparisons with different proposals using the same dataset.

The second research question is formulated as

RQ2: *What are the current trends, strategies, and architectures used in DL-based MI-EEG decoding?*

This question is not only concerned with the frequently used DL approaches and architectures, but also aims to provide detailed comments on procedures used in training and evaluation of decoding models. All reviewed studies are categorized and detailed in terms of decoding paradigms, i.e., the previously cited subject-dependent (SD), subject-independent (SI), transfer learning (TL), and global evaluation, and MI datasets (the ones resulting from RQ1).

The main reason guiding this question stems from the necessity of providing an updated overview of these topics in respect to previous survey studies. The identification of these decoding strategies may be particularly useful for young researchers approaching this field for the first time.

The third research question is formulated as

RQ3: *What are the DL architectures most frequently used as baseline models or as inspirations for new decoding methodologies?*

The main aim of this question is to identify widely used and reliable DL models that are utilized for baseline performance comparison, or as a base model that can be modified to create new decoding models. In fact, the use of baseline models is valued as particularly important for a matter of reproducibility and reliability of novel proposals.

Finally, the fourth research question that will be addressed is given as follows.

RQ4: *How should methodologies and results be reported in a DL-MI decoding study to ensure the reproducibility of results?*

One of the main contributions of this review is identifying and describing studies that make the codes corresponding to their work publicly available, and thus that have addressed the problem of reproducibility to a large degree. Specifically, 27 papers are identified among the reviewed studies as the ones that have made their codes publicly available (open-source). These studies are listed in Tables 4 and 5. In particular, some guidelines will be provided in Section 5 after having analyzed the reported studies.

To answer these research questions, this review is organized as follows. Section 2 explains the procedure by which the 394 studies included in this review have been selected for this review, following the guidelines provided by the PRISMA statement. Section 3 introduces all the publicly available MI datasets that have been employed by the reviewed studies, and provides detailed descriptions of each dataset. This section can be consulted by researchers who want to find a suitable dataset to evaluate their MI decoding models. Section 4 represents the core of this review, providing a detailed overview of the reviewed papers in terms of datasets and decoding paradigms. In fact, this section is divided into subsections each corresponding to one of the previously reported datasets to examine the current DL trends in MI-EEG decoding for every decoding paradigm. Section 5 discusses the answers to the aforementioned research questions by analyzing the materials presented in the previous sections. Finally, conclusions are given in Section 6.

2. Systematic review methodology

This systematic review follows the guidelines provided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement [33]. The following sections provide the necessary information related to the identification, selection, and interpretation of the studies included in this review.

2.1. Eligibility criteria

Studies presenting the use of EEG signals acquired during the MI experimental paradigm and DL decoding approaches have been included in this systematic review according to the following criteria.

1. Non-English papers were excluded.
2. Only papers published in the last five years (from January 1, 2017 to January 23, 2023) are considered. Older papers were excluded.
3. Only studies published as journal articles and conference proceedings are considered (studies published as abstracts, book chapters, posters, reviews, Master's theses, and Ph.D. dissertations were excluded).
4. Papers that did not use at least one publicly available dataset to evaluate their decoding methods were excluded.
5. Papers that did not report an MI decoding performance measure, e.g., classification accuracy or Cohen's kappa (κ) value on at least one of the MI datasets considered were excluded.
6. Papers that did not use a DL architecture (any type of neural network with more than one hidden layer) in at least one of the stages of MI decoding, including preprocessing, feature extraction, and final decoding, were excluded.
7. Papers that did not present sufficient information to reproduce their results, e.g., a model architecture or a valid model evaluation strategy were excluded.
8. Papers that did not report their decoding performance in a way that can be meaningfully compared with other studies were excluded. This exclusion criterion encompasses papers that remove a number of subjects from the considered MI dataset without justification, or the ones that report average performance on subjects from multiple datasets without reporting detailed subject-by-subject performance.
9. Papers that reported a lower performance than the chance level were excluded.

These criteria aim to provide the review with studies that have reproducible results on publicly available datasets, and also have decoding pipelines that are described with a sufficiently high level of detail to enable their re-implementation even when the code is not publicly available.

Studies were initially collected by exploiting scientific search engines and applying the keyword search filters mentioned in Section 2.2. The first three criteria were applied automatically using search engine or spreadsheet filters, while the rest required a detailed examination of each paper as detailed in Section 2.3.

2.2. Databases and search strategy

The documents to be screened were collected from two scientific search engines, i.e., PubMed, and Scopus. The last consultation date of these search engines was January 23, 2023. These studies were extracted by querying

- Scopus as follows: *TITLE-ABS-KEY (((brain–computer AND interface) OR bci OR (brain–machine AND interface) OR bmi) AND ((motor AND imagery) OR mi) AND ((neural AND network) OR (deep AND learning)) AND (electroencephalography OR eeg)) AND (LIMIT-TO (LANGUAGE, "English"))*
- PubMed as follows: *(((Brain–Computer Interface[Title/Abstract]) OR BCI[Title/Abstract] OR Brain–Machine Interface[Title/Abstract]) OR BMI[Title/Abstract]) AND ((Motor Imagery[Title/Abstract]) OR MI[Title/Abstract]) AND ((Neural Network[Title/Abstract]) OR (Deep Learning[Title/Abstract])) AND (Electroencephalography [Title/Abstract] OR EEG[Title/Abstract])*

The search provided 1047 records, of which 998 entries were collected from Scopus, and 192 were collected from PubMed. Duplicate papers were removed automatically by merging the two extracted databases from PubMed and Scopus based on the PubMed ID field in each database and 143 duplicate entries were removed. A final manual screening to check for any remaining duplicate publication was subsequently performed and a number of 897 studies remained for further screening.

2.3. Screening process

In this section, the screening process to apply the eligibility criteria listed in Section 2.1 is explained. After duplicate removal as reported in the previous section, in order to apply criteria No. 4 to 9, we required a careful manual examination of each paper. Therefore, the remaining studies were equally divided among the authors for further screening. Details on the studies were reported by each author and checked by at least one of the other authors to ensure that the provided descriptions were not too biased towards the domain knowledge of the author in charge of analyzing a specific paper. In case this was true, the author providing the inspection read through the paper to find other information and/or confirm the reported descriptions.

Firstly, to apply criteria No. 4 to 6, abstract, keywords, and data-related sections were read to check if the employed datasets were publicly available and if the study was using DL-based approaches. Afterwards, the provided performance measures were screened. If no performance measure was present, i.e., accuracy or κ value, or if the performance was not obtained on a publicly available MI dataset the paper was immediately discarded.

After having performed these steps, the papers were thoroughly read to apply the eligibility criteria No. 7 to 9, and to have a better understanding of the proposed strategies as well as of how the EEG data were used in MI decoding. Specifically, the description of model architectures were analyzed and papers were only retained if the information about the models were sufficient to allow, at least, a partial reproduction of the study. Furthermore, if the model evaluation was not described clearly, a part of the MI dataset was removed without a clear justification, or the performance was reported in a way not comparable to other studies, the paper was discarded. Finally, the papers reporting a lower performance than the chance level were removed from further consideration.

During the screening process, all authors carefully examined their allocated papers to apply the eligibility criteria and extract the information required for reviewing the studies that remained after the screening process, making their decisions based on the overall evaluation of each paper. In cases where a paper was considered borderline or it was unclear if it met the eligibility criteria according to one of the authors, the other authors were notified to examine the considered study and reach a decision collectively.

2.4. Screening outcome

In this section, the final results of the screening outcome is reported. After applying the eligibility criteria, 394 papers remained to be included in the review. Fig. 2 summarizes the complete screening process based on the PRISMA guidelines. Fig. 3 shows the number of remaining document for each year from 2017 to 2023 (the datum related to year 2023 corresponds to the number of papers up to January 23). The increasing trend of DL-based MI-EEG decoding is easily identifiable by observing Fig. 3. More statistics regarding the final outcome of screening, specifically, considering the MI-EEG datasets and MI decoding paradigms are given in Sections 4 and 5.

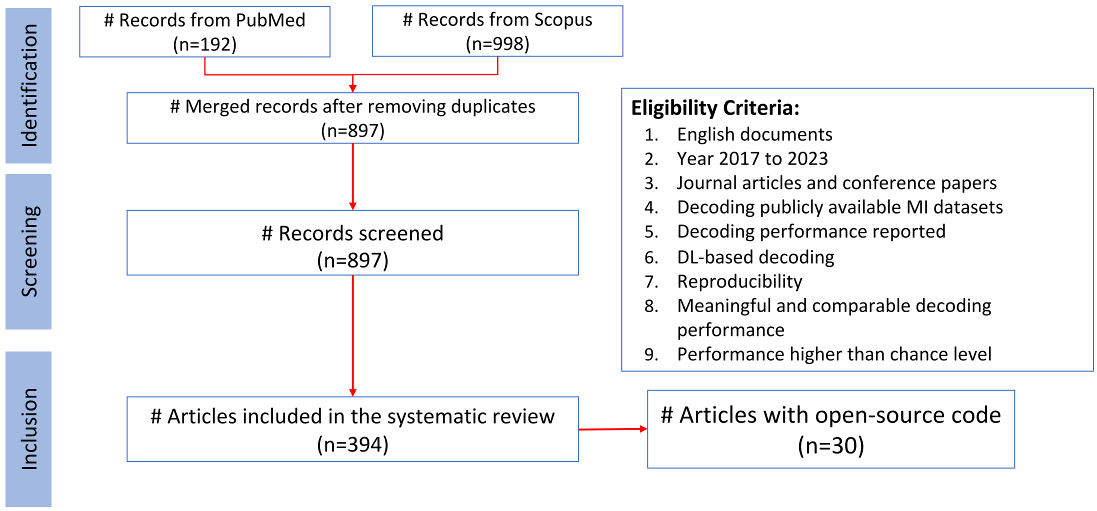


Fig. 2. The PRISMA flow diagram for the systematic review.

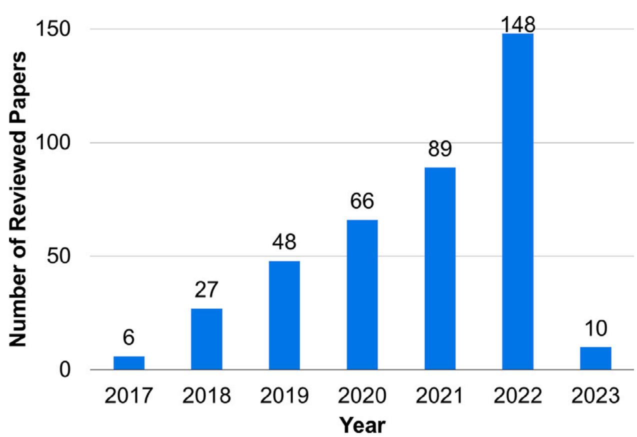


Fig. 3. The total number of remaining studies (394) after screening per year (year 2023 refers only to the period between January 1, 2017 and January 23, 2023).

3. Datasets

In this section, all of the publicly available datasets employed by the 394 reviewed papers are presented and described. The 22 resulting datasets are listed in Table 1, alongside the acronyms that will be used in the rest of this review paper, the related references, and the links to their repositories (last accessed August 21, 2023). Table 2 complements the previous table by providing a brief description of each dataset in terms of the number of electrodes used for EEG recording, the experimental paradigms, and subject-related details, i.e., the number of participants and their age, gender, health status, and their dominant hand. Blank spaces represent missing information.

Considering the experimental paradigm column in Table 2, left/right hand imagination appears to be a very popular paradigm for EEG-based MI datasets. In fact, based on our analysis of the reviewed datasets, about 40% of the reviewed papers use datasets with experimental paradigms specialized in cue-based left/right hand imagination [24,26,27,29,40–42,48], while about 90% consider datasets that involve motor tasks related to tongue [41,42], foot [25,41,42,49], fist [25] and random letter generation [41] in addition to left/right hand imagination. A particular case is represented by *MED-62*, which presents an online BCI system utilizing left, right, and up-down imagination for cursor control [28].

The *SameLimb* dataset [46] moves away from traditional MI-related datasets and proposes a multi-channel EEG recording during motor imagery of different joints (hand or elbow) from the same limb, utilizing kinesthetic motor imagery. Notice that, besides *SameLimb*, some of the reported datasets present paradigms that are very diverse compared to the more diffused ones concerning left/right hand and tongue MI. *Nikki2021* [27] proposes cue-based cylindrical, spherical, and lumbrical MI grasps, while *Kaya2018* [44] introduces five different BCI interaction paradigms including both motor execution and imagination: (i) CLA, consisting of closing and opening fists MI or showing a circle for passive response; (ii) HaLT, including left/right hand, left/right leg, and tongue imagery (brief movements of the leg or foot for leg imagery, imagining a distinct letter or sound for tongue imagery, and a circle for passive state); (iii) 5F, considering finger flexion according to the number shown on the fingers of a virtual hand depicted on a screen; (iv) FreeForm, involving the pressing of the *d* and *l* keys on the keyboard using the right or left hand in the desired order; (v) NoMT, focusing on consistency check and baseline determination with passive participants. This particular dataset is selected by six of the reviewed papers [50–55].

Considering other datasets, that differ from the classical left/right hand MI paradigm, *Jeong2020* [43] designs 11 different upper extremity movement tasks, including arm-reaching along six directions, hand-grasping of three objects, and wrist-twisting with two different motions. *MIDistraction* [45] proposes the left/right hand motor imagery data collection procedure with six variations: without distractions, eyes closed, listening to new sequences, searching for letter-number combinations, watching flickering videos, vibrotactile stimulation on both forearms with carrier frequencies of 50 and 100 Hz, modulated at 9, 10 and 11 Hz. Instead, in the kinesthetic-based dataset *UpperLimb* [47], motor execution and MI experiments are performed, considering right upper limb, elbow flexion/extension, forearm supination/pronation, and hand open/close paradigms.

Thanks to the availability of diverse experimental paradigms the variety of MI studies in the literature is increasing. Also, as can be observed from Table 2, more recent datasets have a larger number of trials which makes them more suitable for DL-based decoding. Another observation that can be made by analyzing Table 2 is related to the number of subjects involved in the experiments for each of the reported datasets. Newer datasets usually consider a greater number of subjects, as introduced in Section 1. This trend can be noticed by observing Fig. 4 that shows the number of subjects enrolled for the datasets used in the reviewed papers, providing a brief reference to the year of dataset publication. Besides *eegmmidb* published in 2009 and providing the dataset with the higher number of subjects (109), *Jeong2020* and

Table 1
References and links to the publicly available datasets used by the reviewed papers.

Acronym	Full name	Link
2020-International-BCI	2020 International BCI Competition Track #4: Upper-limb movements decoding in a single-arm [39]	https://osf.io/pq7vb/?view_only=08e7108d89fd42bab2adbd6b98fb683d
BCI-II-III	BCI Competition II dataset III [40]	https://www.bbci.de/competition/ii/
BCI-III-IIIa	BCI Competition III dataset IIIa [41]	https://www.bbci.de/competition/iii/
BCI-III-IVa	BCI Competition III dataset IVa [41]	https://www.bbci.de/competition/iii/
BCI-III-IVb	BCI Competition III dataset IVb [41]	https://www.bbci.de/competition/iii/
BCI-III-V	BCI Competition III dataset V [41]	https://www.bbci.de/competition/iii/
BCI-IV-1	BCI Competition IV dataset 1 [42]	https://www.bbci.de/competition/iv/
BCI-IV-2a	BCI Competition IV dataset 2a [42]	https://www.bbci.de/competition/iv/
BCI-IV-2b	BCI Competition IV dataset 2b [42]	https://www.bbci.de/competition/iv/
Cho2017	EEG datasets for motor imagery brain–computer interface [26]	http://gigadb.org/dataset/100295
eegmmidb	EEG Motor Movement/Imagery Dataset [25]	https://physionet.org/content/eegmmidb/1.0.0/
Jeong2020	Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions [43]	https://academic.oup.com/gigascience/article/9/10/giaa098/5918864
Kaya2018	A large electroencephalographic motor imagery dataset for electroencephalographic brain–computer interfaces [44]	https://www.nature.com/articles/sdata2018211
MBT-42	Shared data for exploring training effect in 42 human subjects using a noninvasive sensorimotor rhythm-based online BCI [29]	https://figshare.com/articles/online_resource/Shared_data_for_exploring_training_effect_in_42_human_subjects_using_a_noninvasive_sensorimotor_rhythm-based_online_BCI/7959572
MED-62	Human EEG Dataset for Brain-Computer Interface and Meditation [28]	https://figshare.com/articles/dataset/Human_EEG_Dataset_for_Brain-Computer_Interface_and_Meditation/13123148
MIDistraction	BCI under distraction: Motor imagery in a pseudo realistic environment [45]	https://depositonce.tu-berlin.de/items/0f01eb46-4e6e-427a-9a68-b264a839615f
Nikki2021	Psychological and Cognitive Factors in Motor Imagery Brain Computer Interfaces [27]	https://dataverse.nl/dataset.xhtml?persistentId=doi:10.34894/Z7ZVOD
OpenBMI	EEG Dataset and OpenBMI Toolbox for Three BCI Paradigms [24]	http://gigadb.org/dataset/100542
SameLimb	Multi-channel EEG recording during motor imagery of different joints from the same limb [46]	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/RBN3XG
UpperLimb	Upper limb movements can be decoded from the time-domain of low-frequency EEG [47]	https://zenodo.org/record/834976
WCCI2020-Glasgow	Clinical Brain Computer Interfaces Challenge WCCI 2020 Glasgow [48]	https://github.com/5anirban9/Clinical-Brain-Computer-Interfaces-Challenge-WCCI-2020-Glasgow
Weibo2014	Evaluation of EEG oscillatory patterns and cognitive process during simple and compound limb motor imagery [49]	https://paperswithcode.com/dataset/motor-imagery-dataset-from-weibo-et-al-2014

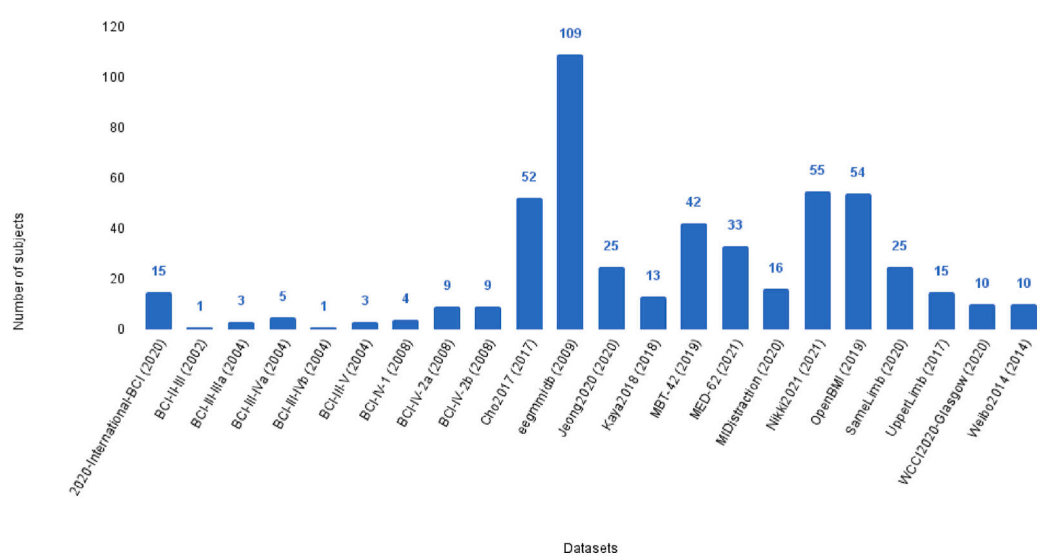


Fig. 4. Number of subjects involved per dataset. The dataset year of publication is reported in parenthesis.

SameLimb (2020) present 25 subjects, the 2019 datasets *MBT-42* and *OpenBMI* as well as *Nikki2021* and *Cho2017* provide data collected from more than 40 subjects.

To conclude the observations in Table 2, 13/22 and 12/22 datasets report the subjects' age and gender, respectively. All the datasets provide data collected from healthy subjects, besides *WCCI2020-Glasgow* which consider hemiparetic stroke patients. Notice that *MBT-62* considers both a control group and subjects trained in mindfulness-based stress reduction (MBSR). Also, the information on the dominant hand is reported in 9/22 dataset descriptions.

An interesting datum present in Table 1 is related to the number of employed electrodes for EEG signal recording. It can be observed that this number varies greatly from a minimum of 3 to a maximum of 128 sensors. This information is extremely correlated with the hardware details reported in Table 3, that presents also the used sampling frequency for EEG signal acquisition, and if a filtering technique was

Table 2
Summary of the characteristics of the datasets used by the reviewed papers.

Dataset	Number of electrodes	Experimental paradigm	Details on subjects				
			Number of subjects	Ages	Gender	Healthy or not	Hand
2020-International-BCI	60	Cylindrical, spherical, lumbrical grasp	15	20–34		Healthy	Right
BCI-II-III	3	Left/right hand	1	25	F	Healthy	
BCI-III-IIIa	64	Left/right hand, foot, tongue	3			Healthy	
BCI-III-IVa	118	Left/right hand, right foot.	5			Healthy	
BCI-III-IVb	128	Left hand, right foot, tongue	1			Healthy	
BCI-III-V	32	Left/right hand MI + words	3			Healthy	
BCI-IV-1	59	Left/right hand, foot	4			Healthy	
BCI-IV-2a	22	Left/right hand, both feet and tongue	9			Healthy	
BCI-IV-2b	3	Left/right hand	9			Healthy	Right
Cho2017	64	Left/right hand	52	24.8 (3.86)	19 F 33M	Healthy	Right except for 2 both handed
eegmmidb	64	Left/right hand, feet/fists	109				
Jeong2020	60	Arm-reaching, hand-grasping, wrist-twisting	25	24–32	10F 15M	Healthy	Right
Kaya2018	19	5 different BCI interaction paradigms[a]	13	20–35	5F 8M	Healthy	
MBT-42	62–64	Left and right hand	42	18–50	18F 24 M	Healthy	37 right 5 left handed
MED-62	62	Up, down,left,right cursor control	62			33 MBSR, 29 control	
MIDistraction	63	Left or right hand with various distractions	16	22–30	6F 10M	Healthy	
Nikki2021	16	Left, right hand	55	20.71 (3.52)	36F 19 M	Healthy	
OpenBMI	62	Grasping with left/right hand	54	24–35	25F 29 M	Healthy	
SameLimb	64	Hand or Elbow from the same limb	25	19–27	6F 19M	Healthy	Right
UpperLimb	61	Right upper limb, elbow flexion/extension, forearm supination/pronation, hand open/close	15	22–40	9F 6M	Healthy	Right except for 1 left handed
WCCI2020-Glasgow	12	Left/right hand	10	47.5(15.31)	4F 6M	Hemiparesis	3 right 7 left
Weibo2014	21	Left hand, right hand, feet	10	23–25	7F 3M	Healthy	Right

[a] Please, see details on this experimental paradigm in the text.

applied by the dataset authors. The first field (*Dataset*) allows the cross-referencing with the other tables by reporting the acronym of each dataset. The *Sampling rate* field presents the sampling frequency (Hz) and down sampling frequency (Hz) between parenthesis. Notice that most of the datasets consider a 250 Hz sampling frequency, but that the majority of the authors downsample the data to 100 Hz or 200 Hz. Besides five datasets not mentioning it, the *Filtering* field reports a vast use of notch filters and bandpass filters. Usually the lower bound of the filter is either 0.1 or 0.5 Hz, while the upper bound is either 100 or 200 Hz. The last column (*Used device and/or system*) of Table 3 presents the systems used for data collection. EEG caps are included, if mentioned in the dataset description file. Most of the presented datasets use Brain Products GmbH devices, while there is an increase use of g.tec GmbH technologies especially in more recent data collections.

4. Deep learning approaches

This section and the following (Section 5) represent the core of this review, providing an overview and a discussion on the studies remaining after the screening process described in Section 2. As previously discussed (Section 1), the main aim of this review is not to present a treatise on DL architectures but to provide pointers to the EEG-based MI datasets commonly used to train and test decoding models, to show the current trends in the DL-MI field, to discuss the models used for benchmarking purposes, and to better understand how the results are presented to the research community. Fig. 5 gives an overview of the distribution of the MI datasets that are used for DL-based MI-EEG decoding,[1] showing a diffused use of the *BCI-IV-2a* and the *BCI-IV-2b* datasets.

[1] As some papers have considered multiple studies to evaluate their models, the sum of the numbers reported on this plot exceeds the number of papers reviewed (which is 394).

Therefore, this section is divided in subsections corresponding to the datasets presented in Section 3. These subsections can be the source of studies with which researchers can compare newly devised MI-based applications. Also, the experimental paradigm used to gather each dataset could be gauged against different methods of MI decoding. We perform a more fine-grained categorization for each dataset and divide each sub-section with respect to the MI decoding paradigms. Note that one can perform even more fine-grained categorization based on the different stages in a DL-based MI decoding pipeline, such as pre-processing and feature engineering or post-processing. However, as such categorization would be rather subjective, the details of the DL pipelines are given individually for each study when they are presented by the authors of the study.

The categorization based on decoding paradigms is intended to better understand the applicability of the DL approaches. Each paradigm characterizes the specifics of either personalized applications, e.g., subject-tuned BCIs or more general and scalable models that can be used on different subjects and datasets. To clarify this point, descriptions for these decoding paradigms are given below.

The main decoding paradigms in MI are the *subject-dependent* (SD) and *subject-independent* (SI) ones. The other two paradigms present in the reviewed studies are *transfer learning* (TL) and *global*.

- **Subject-dependent**: This paradigm is the traditional and most frequently used one in MI EEG decoding. For each subject, a different decoding model is built for training and evaluation using only the data of the individual subject under analysis. As the model is specifically fit to a single subject's data, it may lack generalization and have a low performance when tested on another subject's data.
- **Subject-independent**: This paradigm usually considers a decoding approach based on leave-one-subject-out cross-validation (LOSO-CV), i.e., for each subject a model is trained on the data

Table 3
Additional information on the presented datasets, including sampling frequency, filtering techniques, and employed hardware.

Dataset	Sampling rate (downsampling)	Filtering	Used device and/or system
2020-International-BCI	250 Hz	Notch (60 Hz)	BrainAmp, Brain Products GmbH
BCI-II-III	128 Hz	Bandpass (0.5–30 Hz)	EEG system by g.tec medical engineering GmbH
BCI-III-IIIa	250 Hz	Bandpass (1–50 Hz) and notch (50 Hz)	Neuroscan
BCI-III-IVa	1000 Hz (100 Hz)	Bandpass (0.05–200 Hz)	BrainAmp, Brain Products GmbH
BCI-III-IVb	1000 Hz (100 Hz)	Bandpass (0.05–200 Hz)	BrainAmp, Brain Products GmbH
BCI-III-V	512 Hz	Not mentioned	Biosemi cap system
BCI-IV-1	1000 Hz (100 Hz)	Bandpass (0.05–200 Hz)	BrainAmp MR plus and Easycap, Brain Products GmbH
BCI-IV-2a	250 Hz	Bandpass (0.5–100 Hz) and notch (50 Hz)	BrainAmp MR plus and Easycap, Brain Products GmbH
BCI-IV-2b	250 Hz	Bandpass (0.5–100 Hz) and notch (50 Hz)	BrainAmp MR plus and Easycap, Brain Products GmbH
Cho2017	512 Hz	Not mentioned	Biosemi ActiveTwo system and BCI2000 system 3.0.2
eegmmidb	160 Hz	Not mentioned	BCI2000 system
Jeong2020	2500 Hz	Notch (60 Hz)	BrainAmp and ActiCap EEG, Brain Products GmbH
Kaya2018	1000 Hz (if files present the string *HFREQ.mat* 100 Hz, otherwise 200 Hz)	Notch (50 Hz) and if sampling rate is 100 Hz, bandpass 0.53–100 Hz, otherwise 0.53–70 Hz	Neurofax 1200 EEG
MBT-42	Experiment 1, 2: 1000 Hz (100 Hz); Experiment 3: 1024 Hz (128 Hz)	Experiment 1, 2: bandpass (0.5–200 Hz) and notch (60 Hz); Experiment 3: bandpass (0.16–100 Hz) and notch (60 Hz)	Experiment 1, 2: Neuroscan; Experiment 3: Biosemi ActiveTwo system
MED-62	1000 Hz (250 Hz)	Bandpass (0.1–200 Hz) and notch (60 Hz)	BCI2000 system, SynAmps RT amplifiers and Neuroscan software
MIDistraction	1000 Hz (100 Hz)	Not mentioned	BrainAmp and Easycap Brain Products GmbH
Nikki2021	250 Hz	Bandpass (0.5–30 Hz) and notch (48–52 Hz)	g.Nautilus, g.tec medical engineering GmbH
OpenBMI	1000 Hz	Not mentioned	BrainAmp, Brain Products GmbH
SameLimb	1000 Hz (200 Hz)	Highpass (0.1 Hz) and lowpass (40 Hz)	Neuroscan SynAmps2
UpperLimb	512 Hz	Bandpass (0.01–2000 Hz) and notch (50 Hz)	EEG System by g.tec medical engineering GmbH and 5DT Data Glove for arm movements
WCCI2020-Glasgow	512 Hz	Bandpass (0.1–100 Hz) and notch (50 Hz)	g.USBamp, g.tec medical engineering GmbH
Weibo2014	1000 Hz (200 Hz)	Bandpass (0.5–100 Hz) and notch (50 Hz)	Neuroscan SynAmps2

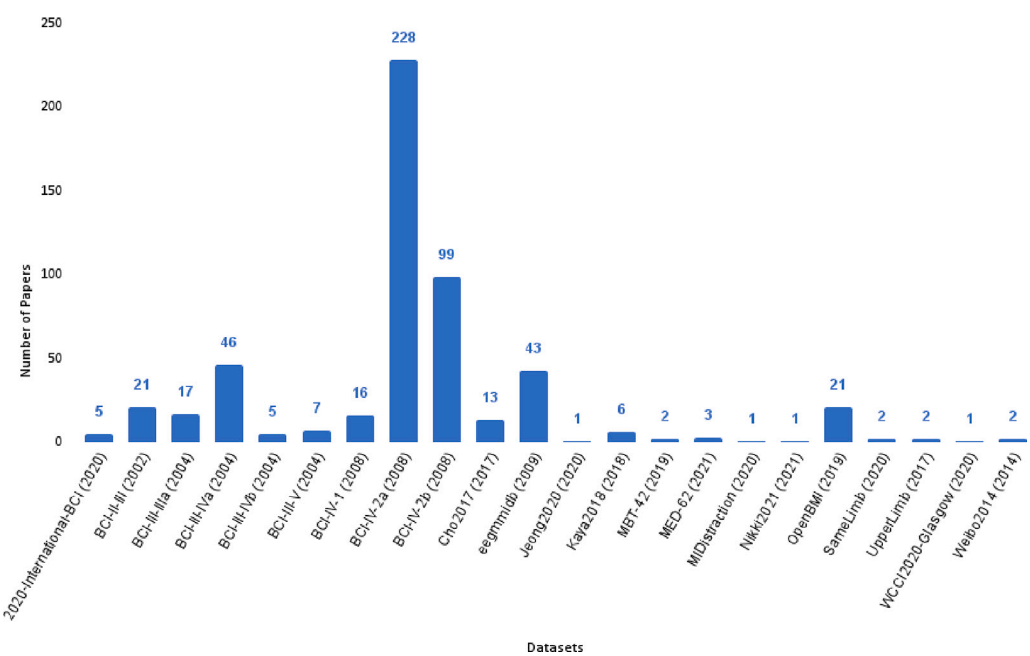


Fig. 5. Number of reviewed studies considering different datasets for DL-based MI-EEG decoding.

related to all other subjects, and the final model is evaluated on the subject considered, without any fine-tuning on this subject's data. This process is repeated for all subjects, and the mean decoding measure is reported as the final SI performance. Therefore, as the test subject's data is never seen during the training process, the SI decoding performance is a good representation of the generalization capabilities of the decoding model.

- **Transfer learning**: TL and domain adaptation are popular techniques used in machine learning and deep learning that exploit an already trained model on a given task to boost performance on related tasks [56,57]. In DL MI decoding, TL is usually applied by combining the SD and SI paradigms. In particular, the decoding model is firstly trained on a subset of subjects. Afterwards, it is fine-tuned on a subset of the test subject's data before

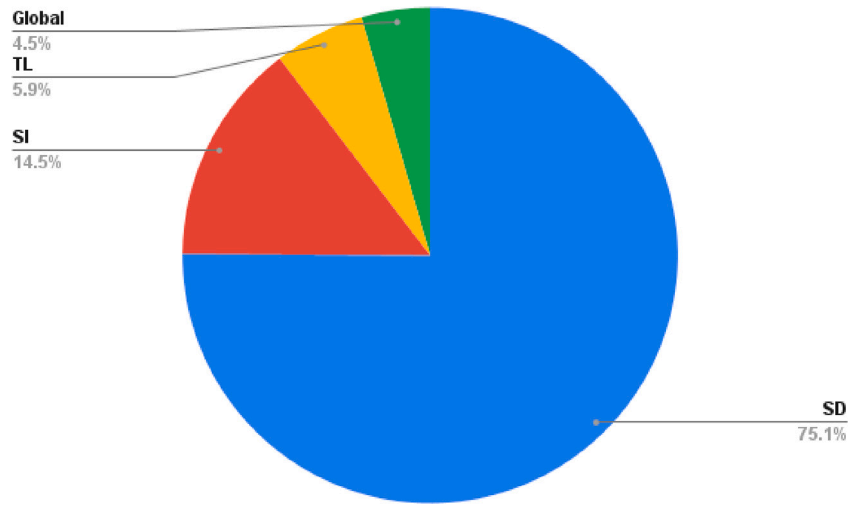


Fig. 6. Distribution of different MI decoding paradigms.

being evaluated on another unseen subset of this subject's data. Therefore, this approach can be useful when a small number of recordings is present for the test subject and the goal is to increase the model decoding performance. This TL approach is known as subject-transfer learning. Another TL approach in MI decoding is session-transfer learning, in which the model is first trained on the data related to one MI recording session of a given subject, and subsequently fine-tuned and tested on another session of the same subject.
- **Global**: In this paradigm, the evaluation process does not take into account the subject index corresponding to a recording. Instead, the whole EEG dataset consisting of all subjects' data is shuffled and a monolithic dataset is formed. Afterwards, a hold-out or CV strategy is used to evaluate the model. This paradigm can also provide information on a model generalization capabilities as it is not calibrated on a specific subject's data.

Fig. 6 gives an overview of the distribution of each decoding paradigm in DL-based MI decoding, while Fig. 7 shows the paradigms in relation to the datasets reported in Section 3. Notice that the majority of the reviewed papers add an SD decoding paradigm to their analyses. Moreover, the use of the SD approach as the only paradigm employed is evident for the *BCI-II-III*, *BCI-III-V*, *WCCI-2020Glasgow*, *Jeong2020*, *MIDistraction*, *SameLimb*, *UpperLimb*, and *Weibo2014* datasets. Instead, a greater variety of decoding paradigms is adopted for other datasets, e.g., *BCI-IV-2a* and *BCI-IV-2b*, and *eegmmidb*.

Considering the relation between decoding paradigms and the evaluation strategies, note that the SD paradigm can use either holdout (considering the train, test, and – sometimes – the validation split) or K-fold CV for evaluation as it deals with a single subject's data for both training and testing. Instead, the SI paradigm is most suited for 1-fold CV which is usually called LOSO-CV, where one subject is left out for testing, and the remaining subjects are used for training. As previously discussed, in the global paradigm, all the subjects' data are shuffled randomly as a monolithic dataset, and either k-fold CV or hold-out can be used for evaluation. Finally, the use of transfer learning usually considers a hold-out evaluation strategy, as the model is first trained on the data of a subset of subjects, fine-tuned on a portion of the test subject's data, and finally tested on the rest of the test subject's data.

Finally, particular attention has been given to the studies that reported available online resources or the ones that provided a clear description to reproduce their proposed approach. Out of 394 works remaining after the screening process, 30 provide references to the codes used for their studies. One paper reported Python codes in the appendix. However, the codes are unavailable in the peer-reviewed publication and there are some inconsistencies between the available paper versions. Therefore, the paper is discarded from further analyses. Other works excluded from the review provided links that land to unavailable repositories. The remaining 27 linked resources are reported in Tables 4 and 5. Notice that the links have been accessed on July 10, 2023, and that the tables are organized to contain the code reference papers, the employed datasets, the topic related to the presented approaches, and the complete link for each resource. Table 4 lists the works published from 2018 to 2021, while Table 5 the ones in 2022 and 2023 for better readability.

In what follows, studies that report decoding performances on each dataset are discussed and further categorized based on the decoding paradigm (i.e., SD, SI, TL, and global). The most common performance measure reported by the reviewed papers is accuracy. Sometimes standard deviation over multiple runs or subjects is provided and thus denoted in this review paper as average accuracy $\pm$ standard deviation % (some studies also report other performance measures that are common in ML model evaluation such as F1-score, recall, and precision, yet as most MI datasets considered have balanced class distributions, the main comparative measure in this review is the mean accuracy). Some studies have also reported Cohen's kappa (κ) values alongside or instead of accuracy.[2]

4.1. 2020-International-BCI

Being very recent, there are a few studies remaining after the screening process (only five) related to the *2020-International-BCI* dataset [39] with respect to other benchmarking databases present in the literature (e.g., BCI Competitions datasets described in Section 3). The related papers have been published in 2021 or 2022. The proposed approaches mainly consider the use of CNNs [82,85–87] to learn deep features and address the multi-class problem of discriminating between different types of MI grasps: cylindrical, spherical, or lumbrical.

Two out of five papers consider an SI approach [82,85,88]. In particular, *Partovi et al.* [85] propose the use of a simple CNN, obtaining the best average accuracy among the three SI works, but with a large standard deviation ($61.55 \pm 16.23\%$). Instead, *Han and Jeong* [88] employ transfer learning using a backbone residual network (ResNet) obtaining an accuracy of $51.06 \pm 12.29\%$.

[2] Notice that the reported performance values have been denoted with the same decimal number precision (i.e., two decimal numbers for the accuracy values and four for the κ ones) for consistency. However, the reviewed studies may provide their results with different precision.

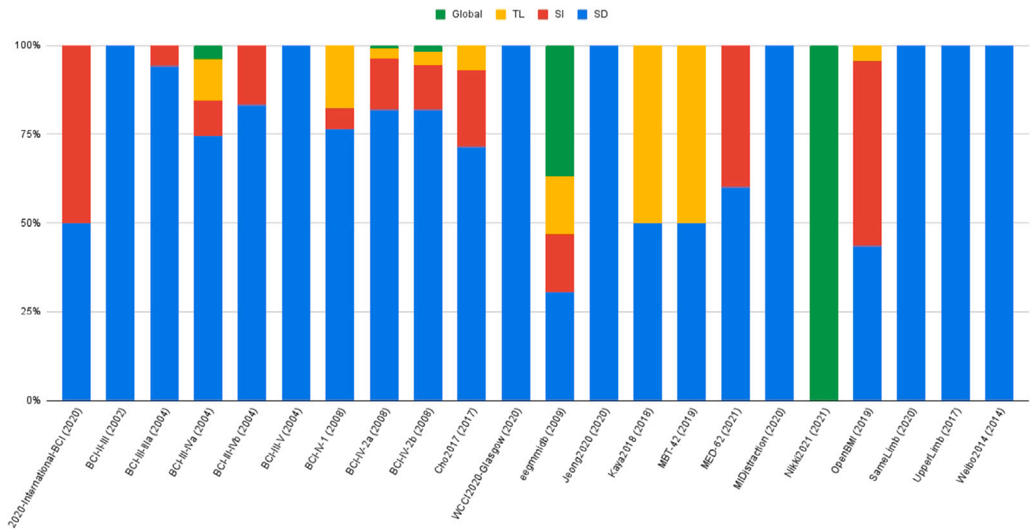


Fig. 7. Distribution of different MI decoding paradigms with respect to the identified publicly available MI-EEG datasets.

Table 4
References to available online codes per publication (2017–2021).

Reference	Datasets	Topic	Link
Dose et al. [58]	eegmmidb	Simple CNN model for classification	https://github.com/hauke-d/cnn-eeg
Kumar, Sharma, and Tsunoda [59]	BCI-IV-1 and Cho2017	OPTICAL subject-dependent predictor based on CSP and LSTM	https://github.com/ShiuKumar/OPTICAL
Mousavi and de Sa [60]	BCI-IV-2a	Temporally adaptive common spatial pattern NN	https://github.com/mahtamsv/TA-CSPNN
Tayeb et al. [61]	BCI-IV-2b	Validation of DL models, i.e., LSTM, spectrogram-based CNN, and recurrent CNN	https://github.com/gumpy-bci/gumpy-deeplearning
Roots, Muhammad, and Muhammad [62]	eegmmidb	EEGNet fusion for cross-subject classification	https://github.com/rootskar/EEGMotorImagery
Zhang et al. [63]	BCI-IV-2a and eegmmidb	Graph based convolutional recurrent attention model	https://github.com/dalinzhang/GCRAM
Autthasan et al. [64]	BCI-IV-2a, MED-62, OpenBMI	MIN2Net as end-to-end multi-task learning model, cosidering SI approaches	https://github.com/IoBT-VISTEC/MIN2Net
Collazos-Huertas [65]	Cho2017	Deep and wide transfer learning	https://github.com/dfcollazosh/DWCNN_TL
Jia et al. [66]	BCI-IV-2a and BCI-IV-2b	Multi-branch multi-scale CNN considering inter-subject and time variability	https://github.com/jingwang2020/ECML-PKDD_MMCNN
Pals et al. [67]	BCI-IV-2b	Mapping DL-based decoders to spiking networks	https://github.com/Matthijspals/neuromorphic-EEG
Pei et al. [68]	BCI-III-IVa	Data augmentation with brain-area-recombination and use of EEGNet	https://github.com/vlawhern/arl-eegmodels
Rammy et al. [69]	BCI-IV-2a	LSTM-based model to consider temporal information	https://github.com/waseemabbaas/Motor-Imagery-Classification.git
Rasheed and Mumtaz [70]	WCCI2020-Glasgow	Classification based on EEGNet	https://github.com/SulemanRasheed/EEG-HandGrasp-Classification
Xu et al. [71]	BCI-IV-2a, Cho2017, eegmmidb, and Weibo2014	Transfer learning enhancement, exploiting pre-trained networks	https://github.com/Mrswolf/alignment-methods-and-adabn
Zancanaro et al. [72]	BCI-IV-2a, BCI-IV-2b, and eegmmidb	DynamicNet, tool for the development of DL models based on CNNs	https://github.com/jesus-333/Dynamic-PyTorch-Net

Yang et al. [82] considers both SD and SI paradigms, and provide their codes to the research community (Table 4). The authors propose an end-to-end deep CNN combining EEGNet [38] and a temporal convolution network, achieving an average accuracy of 56.73% for the SD approach and 55.24% for the SI one.

Considering the SD approaches, *Kwon et al.* [87] provide results considering intra- or inter-session approaches. The authors use the BBCI toolbox and openBMI software to pre-process the signals, and apply data augmentation through the sliding window methodology. The achieved accuracy values are $69.68 \pm 10.10\%$ for intra-session and $52.76 \pm 12.23\%$ for the inter-session approach. Han et al. [86], obtain an SD accuracy of the 62.58% using a ResNet with a one-session-hold-out.

4.2. *BCI competition II dataset III*

For the left/right hand MI *BCI-II-III* [40], only studies using an SD paradigm have been considered, having that it has been collected on a single subject. The screening resulted in 21 papers and besides some variations of standard NNs [89–94] or CNNs [95–101] using either handcrafted or deep features extracted from raw signals or their time–frequency images, works dealing with *BCI-II-III* and using less common approaches are reported in finer details.

A data augmentation technique prior to classification is proposed by *Zhang et al.* [102]. The authors create artificial EEG frames by swapping the intrinsic mode functions obtained by the application of empirical

Table 5
References to available online codes per publication (2022–2023).

Reference	Datasets	Topic	Link
Altaheri, Muhammad, and Alsulaiman [73]	BCI-IV-2a	Physics-informed attention temporal convolutional network	Complete url[a]
Chen, Yu, and Yang [74]	BCI-IV-2a	Meta transfer learning with symmetric positive definite CNN	Complete url[b]
Dolzhikova et al. [75]	OpenBMI	Multi-subject ensemble CNN	https://github.com/irinadolzhikova/MS-En-CNN
Faria et al. [76]	BCI-IV-2a and BCI-IV-2b	Analysis of different data augmentation methods plus application of EEGNet	https://github.com/gbrlfaria/bci-data-augmentation
Huang et al. [77]	BCI-IV-2a	CNN-based classification with fuzzy fusion	https://github.com/Fuminides/Fancy_aggregations
Ko et al. [78]	BCI-III-IIIa, BCI-III-IVa, and BCI-IV-2a	Semi-supervised generative and discriminative adversarial learning	http://deepbci.korea.ac.kr/opensource/opensw/
Mattioli, Porcaro, and Baldassarre [79]	eegmmidb	1D CNN and transfer learning	https://github.com/Kubasinska/MI-EEG-1D-CNN
Salami, Andreu-Perez, and Gillmeister [80]	BCI-IV-2a and OpenBMI	Explainable Inception temporal convolutional network	https://github.com/AbbasSalami/EEG-ITNet
Song et al. [81]	BCI-IV-2a and BCI-IV-2b	EEG Conformer as a compact convolutional transformer	https://github.com/eeyhsong/EEG-Conformer
Yang et al. [82]	2020-International-BCI, OpenBMI, and SameLimb	End-to-end deep CNN combining EEGNet and a temporal convolution network	https://github.com/ingod/DDCLS-MI-EEG-BCI-
Alnaanah, Wahdow, and Alrashdan [83]	BCI-IV-2a and eegmmidb	CNN models, particularly multi-band 1D CNN	https://github.com/malnaanah/coleeg
Nouri et al. [84]	OpenBMI	Convolutional CSP network for SI approaches	https://github.com/Singular-Brain/CCSPNet

[a] https://github.com/Altaheri/EEG-ATCNet/tree/79e5c0ce27218f4e43d45a29bb46ea9eea06063a.
[b] https://github.com/sabinechen/SPD-CNN-Using-Meta-Transfer-Learing-EEG-Cross-Subject-learning.

mode decomposition. Moreover, the data are transformed in tensors exploiting the Morlet wavelet transform and the authors propose the use of a CNN and a wavelet neural network (WNN). Using a 5-fold cross-validation (CV), the CNN and WNN models achieve an accuracy of 90.00% and 85.20%, respectively. Notice that 50% of the original data were used for training and 50% for testing. Artificial frames were generated to train the proposed DL model.

Other approaches consider the use of a temporal spatial CNN combined with stacked autoencoders [103], or the use of a deep transfer learning Alexnet-based model, while exploiting the feature extracted through continuous wavelet transform (CWT) [104]. These two approaches obtained an accuracy of 90.60% and 96.43% (κ = 0.9286), respectively.

An ensemble approach using a factorization machine to combine models learning features from multiple domains is proposed by *Wen* [105]. The author concatenates the outputs of a common spatial pattern (CSP) plus linear discriminant analysis (LDA), a WT plus CNN, and a long-short term memory (LSTM) model and inputs this concatenation to the factorization machine. After 5-fold CV, 85.00% average accuracy is achieved. A similar approach is related to a time- and frequency-domain dual-stream CNN [106], which obtains 90.71% accuracy.

A novel approach is proposed by *Malibari et al.* [107], in which model comprises (i) denoising through a multi-scale principal component analysis based algorithm, (ii) CWT based decomposition, (iii) feature extraction through RetinaNet [108], (iv) hyperparameter optimization through an arithmetical optimization algorithm, and (v) the application of an ID3 classifier. The obtained accuracy is of the 86.53%. Another use of DL models is presented by *Salimpour et al.* [109], who propose a CNN for deep two-dimensional time–frequency maps feature extraction and consider a majority voting classification based on traditional ML techniques, e.g., support vector machine (SVM), achieving 94.91% accuracy. Finally, an interesting case is represented by *Xu et al.* [110] who consider a dual alignment-based multi-source domain adaptation framework. The proposed approach has some promising characteristics, like the selection of informative samples and the use of a multi-branch deep network, but it obtains only a 69.45% accuracy.

As a final remark notice that the data division, when not specified, corresponds to the one provided by the dataset authors, i.e., 140 trials for training and 140 for testing.

4.3. BCI competition III dataset IIIa

The *BCI-III-IIIa* [41] collects the data recorded during a cue-based experiment of left/right hand, tongue and foot MI. Only one out of 17 studies consider an SI approach [111], which proposes a novel methodology for feature extraction by transforming the EEG signal into the weight vector of an autoencoder. A one-vs-all strategy was applied and a SVM classifier used to discriminate the MI conditions, achieving 95.33% accuracy.

The remaining works employ an SD approach and consider the use of different DL strategies. Accuracy values of 69.50% (κ = 0.6950) with one-versus-rest validation on the four-class task, 80.68%, 89.45%, 85.30% (κ = 0.8040), and 91.85% using 10-fold CV, are respectively achieved by a functional link NN [112], a Fisher discrimination dictionary extreme-learning machine [113], a multi-frequency brain network where each layer corresponds to a specific frequency band [114], a temporal constrained sparse group Lasso regularization with EEGNet [115], and a spatial-frequency-temporal 3D CNN model [116].

Data augmentation strategies are also proposed to complement the DL models, starting from standard approaches like the sliding window method applied before using one-vs-rest CSP and a CNN (91.90%) [117], or by adding noise to the training data, before using CSP and classifying with a CNN (88.89%, κ = 0.8519) [118]. A more complex strategy is proposed by *Choo and Nam* [119], who use CWT for image feature extraction, augment the data through a deep convolutional generative adversarial network (GAN) and use a CNN to perform the classification. Through 5-fold CV, the authors achieve 83.04% accuracy.

Another approach considers the modification of the images obtained by transforming the EEG signal after having extracted features through quick-response eigenface analysis [120]. A CNN is finally applied and an accuracy of 91.11% is obtained.

Another novel method considers the use of EEG zero-time windowing to track sensorimotor rhythms temporally and extract features from them using CSP, before using a CNN classifier (88.14%) [121]. The training and validation sets are chosen from the two dataset sessions. Instead, *Liu et al.* [122] propose a bispectrum-based hybrid NN and combine a CNN with squeeze-and-excitation modules obtaining an accuracy of 74.44% considering the binary classification of left and right hand MI (10-fold CV). Among the studies presented in

Table 5, the one by *Ko et al.* [78] reports semi-supervised generative and discriminative adversarial learning. The authors developed three different GAN architectures exploiting deep recurrent spatio-temporal NN, Deep and Shallow ConvNet [37]. Moreover, they consider different percentages of data for training. By using the 75% data for training, the authors achieved the $79 \pm 9.00\%$ accuracy using Deep ConvNet (with semi-supervised generative adversarial learning).

Finally, the remaining studies consider the evaluation of spiking neural models [123], the use of an ANN [124], the development of a multi-class support matrix machine based on evolutionary optimization [125], and the use of LDA after the application of a graph CNN [126].

4.4. BCI competition III dataset IVa

The MI conditions (classes) presented by *BCI-III-IVa* [41] are three: left/right hand and foot motor imagery. After the screening procedure, 46 studies remained, out of which 34 use an SD paradigm, three consider an SI approach, two consider both SD and SI, five use TL approaches, one uses the global paradigm, one reports both SD and global performances, and one reports TL and SD performances.

Starting from TL studies, *Fahimi et al.* [127] propose a deep convolutional generative adversarial network (GAN) for training data augmentation and apply a CNN model with a leave-one-out-subject transfer learning strategy. The discriminator receives subject-independent samples and generated subject-dependent samples, alongside the subject's feature vector to predict the labels. The generator gets the subject's feature vector and add random noise. Notice that *BCI-III-IVa* is used to test the model trained on the authors' proprietary dataset, obtaining 71.14% accuracy. Instead, *Chen et al.* [128] start from an EEGNet structure, considering classification and domain adaptation loss. Afterwards, they exploit the adaptation loss that has a class-wise and a time-wise difference component between subjects. The devised model achieves an accuracy of the 82.61%. *Zaremba and Atyabi* [129] consider cross-subject and cross-dataset subject transfer settings, with the aim of (i) highlighting the effectiveness of subject-transfer, and (ii) evaluating the hypothesis concerning the presence of unique MI patterns. The authors apply a CNN and consider three experiments. The first considers a baseline single trial EEG design with 5-fold CV. The second is a cross-subject within-dataset subject transfer experiment. The third is a cross-subject cross-dataset subject transfer experiment. The final accuracy values are as follows: $85 \pm 1.40\%$ for the first scenario, $85 \pm 1.40\%$ for the second, and $85.80 \pm 1.40\%$ for the last scenario. *Liu et al.* [130] propose a TL approach for subject adaptation using a CNN and achieve a performance of $81.71 \pm 9.27\%$. Considering the binary classification of right hand vs. right foot, *Theng and Atyabi* [131] apply session- and subject-transfer, on the data processed with wavelet transform. The authors use a CNN as the predictive model and achieve the best average single-trial accuracy (db10) equal to 70.28% and the best average subject-transfer accuracy (db7) equal to 75.90%.

For the studies considering both SD and SI paradigms, strategies exploiting standard feed-forward NNs (FFNNs), multilayer perceptrons (MLPs), and CNNs have been proposed and high performances are obtained by the authors of these works. Firstly, [94] consider time and frequency features which are extracted using improved empirical Fourier decomposition passed as inputs to a FFNN classifier. After 10-fold CV, the model achieves 99.82% and 82.70% for the SD and SI paradigm, respectively. Finally, *Sadiq et al.* [132] present a simple MLP with input features extracted from empirical wavelet transform and reduced by dimensionality reduction methods. The model obtains 100% and 97.80% for the SD and SI paradigm, respectively.

Considering the global paradigm, *Zhang et al.* [133] use a spatio-temporal CNN architecture and achieve an accuracy of $85.32 \pm 3.00\%$ using 10-fold CV. Moving to the studies using the SI approach, a CNN-based framework achieving an SI 94.66% accuracy is proposed by *Ortiz-Echeverri et al.* [134], while an MLP-based model with entropy and energy features achieves a performance of $81.71 \pm 9.27\%$ with 10-fold CV, maintaining low computational costs [135]. Autoencoders are also exploited to provide features for random forest classifiers, achieving $99.92 \pm 0.08\%$ accuracy [136].

Balim et al. [137] apply a constant-Q time–frequency transform to better detect the energy variations in low frequencies. Afterwards, they concatenate the images obtained on each channel and feed them to a CNN. Again, 10-fold CV is applied and the following results obtained: $76.80 \pm 4.80\%$ and $77.30 \pm 1.60\%$ for the SD and global paradigms, respectively.

Regarding the SD studies, considering their numerosity (34 studies), a more detailed description will be provided only to works using novel DL strategies or reporting higher accuracy values compared to the majority of the studies. For example, a very high accuracy (99.35%, $\kappa = 0.9869$) has been obtained by *Chaudhary et al.* [138] considering the binary classification of right hand vs. right foot MI. The authors provide time–frequency image representations of the EEG signals using CWT and inputs them to a CNN model, considering 80% of the data for training. Another binary classification is performed by *Stephe and Kumar* [139], who propose a deep generative adversarial network and obtain 95.29% accuracy.

The semi-supervised generative and discriminative adversarial learning models proposed by *Ko et al.* [78] and presented in the *BCI-III-IIIa* dedicated Section 4.3 are also used on *BCI-III-IVa* and achieve around $77 \pm 9.00\%$ accuracy considering the application of Shallow ConvNet in the semi-supervised generative adversarial learning scenario. Instead, *Sadiq et al.* [140] propose a pipeline comprising a multiscale principal component analysis for denoising, the 2D modeling of the modes obtained by empirical wavelet transform, the extraction of geometrical features, and a cascade FFNN obtaining 95.30% accuracy. *Dokur and Olmez* [118] augment the dataset with noise, afterwards apply CSP and feed the CSP features to a CNN. The output of the CNN is compared to a predetermined Walsh matrix to predict the class. The final classification accuracy reported is 98.50% ($\kappa = 0.9700$).

Particular proposals are presented by *Jin et al.* [141] and *Lin et al.* [142]. In the first study, the authors create a graph for each EEG signal based on mutual information and the distance between electrodes and between different classes. Afterwards, they use a special CNN that converts graphs into embeddings and performs convolution. The output of the CNN is given to a fully-connected layer for classification. After 10-fold CV, their proposal achieves a $95.94 \pm 4.40\%$ accuracy. The second study adopts a simplified distributed dipoles model. Firstly, the signals are filtered in multiple sub-bands, and are ranked based on average energy. Afterwards, the sLORETA algorithm [143] is applied to assign a dipole to each of the obtained top sub-bands in a simplified brain cortex model. Afterwards, 68 scouts (regions of interest in the brain) are created with a 3D coordinate and time series. For each N top scout and the corresponding coordinates, a 3D CNN is trained. Finally, the resulting features are fused and classified, obtaining an accuracy of $97.98 \pm 0.82\%$ ($\kappa = 0.9596$) after 10-fold CV.

DL models are also used to learn features that are provided as inputs to traditional ML algorithms. In fact, a CNN model is used to extract features before performing the classification through an ensemble SVM-based voting system (96.34% accuracy with 10-fold CV) [144]. Another example is represented by the use of a graph CNN prior to LDA application (89.14% accuracy with 10-fold CV) [126].

For completeness, other approaches using an SD paradigm considered the evaluation of spiking neural models [80], a multi-kernel extreme-learning machine [145], and a Fisher discrimination dictionary extreme-learning machine (ELM) [113]. Other proposals regard models based on CNN architectures using tensor-based feature representation through wavelet transform and tensor discriminant analysis [146], CWT image features after data augmentation through deep convolutional GAN [119], CSP features after EEG zero-time windowing [121], or spatial–spectral–temporal [147], spatial-frequency [148], time–frequency [149–152], or frequency features [153] as inputs. CNNs are also proposed in combination with an autoencoder [154] or considering benchmark models like the Shallow ConvNet [155]. Other deep architectures are a spatial-frequency-temporal 3D CNN [116], a brain-area-recombination EEGNet-based model [68], a bispectrum-based hybrid NN where a CNN is combined with squeeze-and-excitation modules [122], a temporal-rearrange based MI-EEG network called *TRMINet* [156], an AlexNet [157] based CNN [158], a MLP preceded by a GAN data augmentation [159], and a FFNN applied after channel and neuron number selection [160]. Deep NNs presenting LSTM layers and considering time–frequency auto-regressive domain features are also present among the screened studies [161–163].

4.5. BCI competition III dataset IVb

BCI-III-IVb [41] presents recordings collected during the MI of left hand, right foot, and tongue. Only five studies remain after screening and present some basic architectures, where the one by *Sadiq et al.* [132] presented in the previous Section 4.4, is the only work using both SD and SI paradigms, and others work on the SD paradigm. The accuracy values obtained by the MLP model fed with the reduced empirical wavelet transform (EWT) features are equal to 100% and 92.20% for the SD and SI approaches, respectively. Considering the SD paradigm, the same group of authors [160] achieve 99.52% accuracy by applying a FFNN after 10-fold CV.

Other approaches consider the application of an adaptive neuro-fuzzy inference system [89], and the use of an AlexNet-based CNN [158]. A more complex procedure is proposed by *Yu et al.* [94] who apply multi-scale principal component analysis for denoising, and improved empirical Fourier decomposition to obtain different modes, before inputting the data into a FFNN. The 10-fold CV obtains 93.33% accuracy.

4.6. BCI competition III dataset V

The motor imagery experimental paradigm considered by *BCI-III-V* [41] consists of the left and right hand imagination. Of the seven (all SD) related studies, some basic approaches proposed for other datasets [94,123,158,160] are present and achieve $62.90 \pm 2.60\%$ (spiking neural models), 99.33% and 88.08% (FFNNs), and 89.90% (AlexNet-based CNN) accuracy, respectively.

Other methodologies rely on simple architectures based on LSTM given power spectral density and CSP features (68.51% accuracy) [164]. *Tiwari et al.* [165] propose MIDNN, which is a deep neural network (DNN) devised to detect left and right hand MI. The authors divide the signal in different frequency sub-bands by bandpass filtering the data, and compute the power spectral density. The final accuracy value reported is equal to 82.48%. Finally, *Sadiq et al.* [166] propose the exploitation of pre-trained networks and provide the time–frequency representations of the signals through CWT. Different CNN architectures are applied and the authors obtain the best accuracy (97.77%) with ShuffleNet [167].

4.7. BCI competition IV dataset 1

The MI conditions considered by *BCI-IV-1* [42] vary among subjects, having that the subjects had the option to choose two conditions from the left/right hand and foot MI tasks. Among the 16 studies remaining after screening, one considers both SD and SI paradigms, three present a TL approach, and the others work on the SD paradigm.

Yang et al. [168] propose the use of a multi-layer CNN to spatially represent the data. The invariant spatial representations are extracted by considering inter-subject training with a spatial autoencoder. Notice that the signals are divided in segments of 2s with 0.5s overlap. The final reported accuracy are equal to 85.10% and 84.70% for the SD and SI paradigm, respectively.

A few studies [129–131] consider the use of TL strategies (see Section 4.4 for details of these models). *Liu et al.* [130] obtain an accuracy of $79.35 \pm 12.01\%$. Instead, *Zaremba and Atyabi* [129] approach uses cross-subject and -dataset subject transfer and achieves accuracy values of $83.20 \pm 1.50\%$ in the single-trial EEG design, $82.20 \pm 1.90\%$ in the cross-subject within-dataset experiment, and $84.40 \pm 1.80\%$ in the cross-subject cross-dataset scenario. Also, *Theng and Atyabi* [131] achieve final accuracy values of 67.12% on the single-trial experiment, and 73.20% on the subject transfer experiment in the TL paradigm.

Concerning the remaining SD works (12 papers), some studies consider the use of FFNNs, or standard CNNs [169–173].

Kumar et al. [59] make their codes publicly available (see Table 4) and use the *OPTICAL* (Optimized CSP and LSTM based) predictor. This method combines CSP and LSTM to extract features that are reduced via LDA and are given to an SVM for classification. After 10-fold CV, $82.52 \pm 8.17\%$ accuracy ($\kappa = 0.6500$) is obtained. Another proposal [126] regarding the use of DL strategies before classification, applies a graph CNN to provide a channel selection based on brain activity. Afterwards, CSP and LDA are used to extract features and classify the resulting data. The accuracy achieved after 10-fold CV is equal to 79.76%.

Instead, *Wu et al.* [174] propose *TransEEG*, which is a hybrid network exploiting convolutional operations and self-attention mechanisms. Its main aim is to model both local and global dependencies for EEG signal processing by exploiting a CNN encoder and three transformer blocks (with graph embeddings). The final reported accuracy is $77.40 \pm 6.00\%$.

Han et al. [175] use a parallel CNN considering the data collected only on subjects performing left/right hand MI, and obtain a 84.50% accuracy ($\kappa = 0.6900$).

Concerning the remaining SD studies, a bispectrum-based hybrid NN [122] achieves 73.25% accuracy. Finally, *Ou et al.* [156] obtains 73.14% using a temporal-rearrange based MI-EEG network based on the Shallow ConvNet model.

4.8. BCI competition IV dataset 2a

BCI-IV-2a [42] is widely used as a benchmark dataset to test novel processing and classification approaches. Having that the remaining studies after screening are 228 and wanting to avoid an extremely lengthy section, *BCI-IV-2a* entries have been treated as follows:

1. Priority is given to the studies that report publicly available codes (whose references are presented in Tables 4 and 5).
2. The remaining entries are divided per paradigm (i.e., SD, SI, both SD and SI, and other mixed approaches, if any).
3. The works that have novel approaches among the top 20 in terms of performance are presented with a higher level of detail.
4. The remaining papers are clustered by DL approach and further details are provided only for novel proposals.

Notice that the highest reported accuracy referred in the third point above may not be directly compared in all cases, due to the presence of different performance evaluation strategies (e.g., hold-out or k-fold CV). Therefore, we try to specify the validation method to the extent discussed in the original work for these studies.

Considering this analysis procedure, this review section presents 16 studies reporting available codes, 170 using an SD approach, 16 an SI one, 14 employing both paradigms. The remaining papers consider TL, global or a combination of decoding paradigms.

4.8.1. Studies reporting available codes

Observing the code links and references reported in Tables 4 and 5, 16 out of 27 table entries use *BCI-IV-2a* [60,63,64,66,69,71–74,76–78, 80,81,83].

Salami et al. [80] consider an SI paradigm and propose *EEG-ITNet*, an explainable Inception temporal convolutional network that achieves

$76.74 \pm 11.48\%$ accuracy after 10-fold CV. *Ko et al.* [78], described in Section 4.3, reports an accuracy of $66 \pm 14.00\%$ (5-fold CV) using Shallow ConvNet with their semi-supervised generative adversarial learning strategy, in an SD configuration. This paradigm as well as the SI one are applied in another study [73] using a physics-informed attention temporal convolutional network. The final achieved accuracy values are equal to 85.38% and 70.97% for the SD (hold-out with 50% trials for training and 50% for testing) and SI (LOSO validation) paradigm, respectively.

Zhang et al. [63] propose a graph based convolutional recurrent attention model. The EEG spatial information is modeled through a graph, and subsequently a convolutional recurrent attention model is deputed to learn temporal and spatial features. Using an SI approach with a LOSO validation, the authors obtain a best average accuracy of $60.11 \pm 9.96\%$ considering an N-Graph. Instead, *Song et al.* [81] consider a data augmentation step (during training) previous to the application of the proposed EEG Conformer, which is a compact convolutional transformer composed by 1D temporal and spatial convolution layers, a self-attention module, and fully-connected layers. Considering an SD paradigm, the model achieves 78.66% accuracy (in this case, the division intended by the dataset authors is respected, thus the trials of the first session were used for training, while the ones from the second session for testing). Another approach [77] proposes a CNN-based classification with fuzzy fusion and obtains $67.93 \pm 0.13\%$ and $53.28 \pm 0.16\%$ accuracy for the SD and SI paradigm, respectively.

Faria et al. [76] are particularly interested in testing different data augmentation approaches (i.e., sliding window, segmentation and recombination, empirical mode decomposition, noise addition, and amplitude perturbation) and consider within- and cross-session EEGNet classification for a same subject. The authors obtain $70.80 \pm 14.80\%$ and $80.80 \pm 15.30\%$ accuracy, considering the multi-class and binary (left/right hand MI) classification problems. *Autthasan et al.* [64] propose *MIN2Net*, a model that exploits end-to-end multi-task learning, integrating deep metric learning into a multi-task autoencoder. This allows to provide an efficient representation of the EEG signals and classify them at the same time. The model achieves $65.23 \pm 16.14\%$, and $60.03 \pm 9.24\%$ accuracy for the SD and SI approaches, respectively.

Instead of proposing a novel architecture, *Zancanaro et al.* [72] design *DynamicNet*, a tool to develop CNN-based DL models in a quick and flexible manner, and their best average achieved accuracy is of the 70% for the left/right hand MI classification task (SD paradigm). Considering the multi-class problem, $67.88 \pm 1.19\%$, $72.50 \pm 1.80\%$, $59.35 \pm 1.97\%$ accuracy values are obtained by using an SD paradigm, applying an SI strategy considering all the subjects test data for testing, or a LOSO validation. *Jia et al.* [66] focus on a strategy that can avoid signal preprocessing, while considering the inter-subject and time-variability. Therefore, the authors propose a multi-branch multi-scale CNN and verify that considering a greater number of channel usually provides more information to classify the EEG data. Their model achieves $81.40 \pm 1.17\%$ with 5-fold CV in an SD configuration. Instead, accuracy of 59.20% and 69.20% are obtained by a temporally adaptive CSP NN (10-fold CV) [60] and by a multi-band 1D CNN [83], respectively.

A more complex approach is proposed by *Xu et al.* [71]. The source dataset is used to train EEGNet and Shallow ConvNet considering four preprocessing strategies, i.e., channel/trial normalization and Euclidean/Riemannian alignment. Afterwards a transfer performance is applied on the target datasets using the pre-trained networks. Interval co-variate shift is reduced with adaptive batch normalization. Notice that the performance models were evaluated using a 5-fold stratified sampler and considering a random trial division as follows: training (60%), validation (20%), and test (20%) set. Moreover, three experiments were performed: (1) within-subject classification, (2) cross-dataset classification, and (3) baseline comparison. The considered datasets are *BCI-IV-2a*, *Cho2017*, *eegmmidb*, and *Weibo2014*. Concerning *BCI-IV-2a*, the within-subject best accuracy (around 89%) is achieved using EEGNet with Riemannian alignment. Instead, the cross-dataset experiment obtains the best accuracy 78.80% using as source dataset *eegmmidb* and EEGNetv4 with Euclidean alignment and adaptive batch normalization. A high accuracy ($81 \pm 7.90\%$) is also achieved by considering a simple LSTM model taking CSP features as input [69]. Finally, *Chen et al.* [74] study describes a meta transfer learning with a symmetric positive definite CNN architecture and fine-tuning the model for each subject obtains an average accuracy of $47.44 \pm 4.10\%$.

4.8.2. Studies using a subject-dependent paradigm

Given the large number of papers in this paradigm, the papers with the highest reported accuracy (as discussed at the beginning of Section 4.8) are reported. Afterwards, papers with novel and interesting approaches according to the authors' perspective are discussed.

Starting from the top 20 studies, the reported results span from 89.68% to 99.96% accuracy. *Jayashekar and Pandian* [176] use CSP and a CNN to extract features, and afterwards classify the data using a multi-SVM obtaining 89.68% accuracy. *Lee et al.* [177] focus on the data dimensionality problem and propose a data-driven data augmentation based on ensemble empirical mode decomposition. Moreover, the authors exploit filter bank CSP (FBCSP) [22] and CNN for classification. The best average accuracy ($91.24 \pm 5.67\%$ with 5-fold CV) is achieved by augmenting the dataset to five times the original size.

Kim et al. [178] employ a sequential transfer learning approach obtaining $91.34 \pm 5.45\%$ accuracy with 10-fold CV, while [179] propose a novel lightweight feature fusion network based on an attention mechanism and tensor decomposition. The authors augment the data through an enhance-super-resolution GAN and obtain 91.58% accuracy, considering the division of the dataset in the ratio 6:2:2 for the training, validation, and test set. Instead, *Kamhi et al.* [180] focus on finding the best hyperparamter through Bayesian optimization for the CNN average ensemble strategy they devised and obtain 92% accuracy. A similar value (92.40%, with 50% training data) is also achieved applying FBCSP before using an artificial neural network (ANN) for classification [181], or exploiting a depth separable convolution bidirectional convolution LSTM model based on an attention mechanism (92.60% accuracy) [182]. The use of an attention mechanism is also proposed by *Lashgari et al.* [183] using a CNN-based model and data augmentation, which achieves 93.60% accuracy using hold-out evaluation. A simpler classifier is instead considered by *Ferreira et al.* [184], who extract features through wavelet energy spectrum and average power with discrete Fourier transform, and then apply a simple MLP ($94.06 \pm 1.37\%$). Notice that 70% of the data were used for training while the remaining data for model validation.

Another simple approach [185] employs features computed through continuous wavelet transform and CSP as inputs to a CNN, obtaining $94.44 \pm 2.18\%$ accuracy (with 5-fold CV) for the binary classification of left/right hand MI. Instead, *Li and Ruan* [142], described in Section 4.4 and which proposed a simplified distributed dipole model exploiting LORETA, achieving 94.53% accuracy applying a 10-fold CV. A similar result is achieved by applying non-negative matrix factorization and a CNN classifier obtaining 94.58% accuracy with 10-fold CV (and 99.53% accuracy considering left/right hand MI only) [152].

Wankhade and Chorage [186] propose *RideNN* classifier, a NN using the rider optimization algorithm and receiving in input the feature extracted by CSP, tunable Q-wavelet transform, and holo-entropy based wavelet packet decomposition. The final achieved accuracy is equal to 95.32%. *Judith et al.* [187] obtains 96% accuracy, considering 70% data for training, 15% for validation, and 15% for testing, providing a two phase classification starting from the use of an ANN and finishing with the application of an adaptive SVM. Another approach [188] considers the selection of an optimal frequency band to better characterize the EEG data computing the energy in different sub-bands through discrete wavelet transform. Power spectral density is used to extract features, while a visual geometric group network based CNN achieves 96.21% accuracy (augmenting the data through the sliding window methodology). *Sun et al.* [189] propose an end-to-end deep learning framework called EEG channel active inference NN, based on graph CNNs. The authors main aim is to exploit temporal and spatial domain correlations. Notice that the model architecture is composed as follows: temporal feature extraction module, channel active reasoning module, repeated in sequence for other two times, plus a final temporal feature extraction module, followed by a flatten and softmax layer. Moreover, it has been tested only on the left/right hand MI conditions to allow comparison with their proprietary dataset. The proposed framework obtains $96.90 \pm 3.50\%$ accuracy (with 10-fold CV).

Choi et al. [120] create multiple images for each EEG signal and use quick-response eigenface analysis for feature extraction. Data augmentation is applied before using a CNN for classification, obtaining 97.87% accuracy. A data augmentation strategy based on linear interpolation is instead proposed by *Li et al.* [190], who use a plain CNN to achieve $98.23 \pm 1.60\%$ accuracy (10-fold CV). *Balmuri et al.* [191] focus on the use of an enhanced grasshopper optimization algorithm for optimum statistical feature selection and apply an extreme learning machine (ELM) to obtain 99.12% accuracy. A similar value (99.40% with 5-fold CV) results from a cascade network composed of a CNN and a gated recurrent unit (GRU) [192], while $99.96 \pm 0.04\%$ accuracy is achieved by classifying the data through a random forest algorithm, after having used an autoencoder for feature extraction [192].

Besides these (accuracy-based) top 20 studies, some other works present peculiar approaches and/or analyses. Starting from papers reporting Cohen's kappa (κ) as the only performance measure, *Razzak* [125] proposes the use of a multi-class support matrix machine based on evolutionary optimization, achieving a κ value of 0.6560 (5-fold CV), while *Chen et al.* [193] propose a semi-supervised DL approach based on a stacked variational autoencoder, obtaining κ = 0.6300 (70% data for training). Instead, *Liu et al.* [194] consider frequencies in the range 4–38 Hz and propose a spatial–spectral feature learning neural network to improve the FBCSP feature separability ($\kappa = 0.7000$). Another work uses a spatio-temporal CNN with residual connections, achieving $\kappa = 0.4500$ (10-fold CV) [195], while also considering the possibility of analyzing the learned network weights from a neurophysiological point of view, and thus observing MI-related neural patterns.

Another couple of peculiar works are the ones by *Jin et al.* [141] and *Wang et al.* [196]. In the first study, as described in Section 4.4, the authors propose the generation of graphs for each EEG signal and then converts them into embeddings to feed to a CNN. After 10-fold CV they achieve $92.47 \pm 11.50\%$ and 90.33% accuracy considering the binary classification of left/right MI and all the experimental conditions, respectively. Instead, *Wang et al.* [196] perform the binary classification of all the different combinations of the dataset tasks using a variational sample-LSTM. Their proposal obtains an average binary accuracy value of left vs. right hand, left hand vs. foot, left hand vs. tongue, right hand vs. foot, right hand vs. tongue, and foot vs. tongue equal to 92.52%, 94.72%, 94.34%, 94.31%, 95.33%, 96.44%, respectively.

Virgilio et al. [197] employ a spiking neural network and provide two experiments: (i) considering input features with constant values, and (ii) considering input features with temporal information. The classification task consider binary combination of rest, left/right hand, foot, and tongue MI. The average accuracy on all the classification tasks is 81.36% using 75% data for training and the remaining ones for testing. *Sorkhi et al.* [198] use multi-scale FBCSP to extract features and propose a model based on compact CNN plus a LSTM, which hyparameters are tuned by applying Bayesian optimization. The model achieves 89.26% accuracy with 10-fold CV. Instead, *Li et al.* [199] consider the use of fast Fourier and Clough–Tocher interpolation to generate the spatio-frequency images, which are fed to a modified VGGNet obtaining 88.87% accuracy ($\kappa = 0.7800$).

A new approach is provided by *Zhang et al.* [200], who propose *EEG Inception*, i.e., a CNN architecture based on an inception-time network. The model achieves $88.39 \pm 7.06\%$ accuracy considering a data augmentation step, which exploits data over 100 Hz to add noise between trials. Another novel proposal is a multi-level generative deep learning based methodology [201]. A deep belief network model composed by multiple basic restricted Boltzmann machine blocks as well as a stacked sparse autoencoder model consisting of different stack autoencoder blocks are considered. The best average accuracy is achieved with the stacked sparse autoencoder and is equal to 87.99%.

Bang et al. [202] consider first pre-defined and subject-optimized (using mutual information) filters. Then create a 3D normalized sample covariance matrix from the filter outputs and pass them as inputs to a 3D CNN, which reconstructs the feature map that are visualized topographically. The authors use a hold-out validation strategy and obtain $87.15 \pm 7.31\%$ accuracy. An accuracy of 86.80% is achieved by a temporal and channel attention convolutional network in a framework named TCACNet [203]. Notice that 80% data were used for training (of which 20% for validation), and 20% for testing.

Instead, the use of an Euclidean data alignment as a preprocessing method and a multi-scale CNN with a TL approach achieve $86.03 \pm 0.55\%$ accuracy [204].

Xue et al. [114] consider multi-frequency brain network already applied on *BCI-III-IIIa* (Section 4.3) obtains 83.83% accuracy; similarly, *Musallam et al.*'s [205] *TCNet-Fusion* achieves 83.73% accuracy. *TCNet-Fusion* is a fixed hyperparameter-based CNN model that exploits temporal convolutional networks, separable and depth-wise convolution, as well as layer fusion. In *Altuwaijri and Muhammad* [206], the authors use fusion multi-branch EEGNet with convolutional block attention module and knowledge-driven feature component [207] to achieve 83.68% and 88.63% accuracy, respectively.

Yang et al. [208] propose a 3D representation of the data and exploit a two-branch (spatial and temporal feature learning branches) 3D CNN. Also, the authors introduce a 3D data augmentation based on cyclic translation in the time dimension and obtain $83.20 \pm 11.52\%$ accuracy. 83% and 82.87% accuracy are instead achieved by a time-contained spatial filtering and spatial–temporal analysis network [209], and a multi-branch EEGNet with squeeze-and-excitation blocks [210], respectively.

Zhang and Yang [211] introduce an adaptive layer into the fully connected layer of a deep CNN. This new layer objective function considers the minimization of the local maximum mean discrepancy, the prediction error and the distance within each class, while maximizing the distance between classes within each domain. Applying 10-fold CV for ten times, the proposed model achieves a best average accuracy of the $82 \pm 0.13\%$. *Yang et al.* [212] focus on improving the feature vector by proposing a discriminative feature learning strategy. Notice that a CNN framework including a circular translation data augmentation is applied and $81.85 \pm 10.15\%$ accuracy achieved with 10-fold CV. Instead, *Hong et al.* [213] propose a dynamic joint domain adaptation network based on adversarial learning strategy to learn domain-invariant feature representations. The source domain is represented by the session 1 data, while the target domain the session 2 ones. The final obtained average accuracy is equal to 81.52%.

Another use of DL models is proposed by *Chen et al.* [214], who consider a multi-attention CNN for feature extraction and tune it by inter-session discriminator loss to reduce inter-session variability. Notice that the features are given to another layer of the CNN for classification, which achieves 81.48% accuracy ($\kappa = 0.7530$). *Gao et al.* [215] also focus on features and augment them by considering the combination of a GRU and a CNN. The classification is performed by a simple ANN and obtains 80.70% accuracy ($\kappa = 0.7400$) with 10-fold CV.

A densely feature fusion CNN achieves 79.90% accuracy [216], which is also obtained by a multi-scale fusion CNN based on an attention mechanism that also extracts spatio-temporal multi-scale features [217]. Instead, *He et al.* [218] develop *S-CAMLP-Net* to predict next EEG segments using a self-supervised learning approach. This mechanism consists of a 1D CNN encoder, and LSTM and convolutional layers. The weights of the encoder are given to a downstream classification task encoder. Afterwards, a channel attention MLP mixer and an ANN classifier are applied. The authors obtain $79.43 \pm 1.73\%$ accuracy with 5-fold CV. Another interesting approach is proposed by *Huang* [219]. The author focuses on the fact that CNNs are generally unable to handle the low SNR of EEG signals and thus proposes a novel residual shrinkage block to construct a CNN model called *RSBConvNet*. Moreover, this work verifies the denoising efficacy of the model and applies both trial-wise and trial-cropping classification strategies. An accuracy of 79.17% is achieved considering a one-vs-rest classification of the multi-class problem.

The best average results obtained by *Wang and Li* [220] is equal to 77.54% (10-fold CV). The authors propose a channel importance based image method (called *CIBI*) by computing the power in the 8–30 Hz range for each channel with discrete Fourier transform. The power is given in input to a random forest algorithm to find channel contributions considering the μ and β activity. Time–frequency images are then generated and fed to a dual branch fusion CNN. Instead, *Ju and Guan* [221] propose a model called *Tensor-CSPNet*, which is a geometric DL framework trying to model the temporal, spatial and frequency patterns of EEG data. In particular, this framework consists of a tensor stacking stage, the use of CSP, the application of a temporal convolution, and the classification. Considering a 10-fold CV strategy and an hold-out validation, the model obtains $74.92 \pm 14.63\%$ and $72.96 \pm 14.98\%$ accuracy, respectively. *Ai et al.* [222] apply an ensemble learning strategy for the SD approach, and also report a TL subject-transfer accuracy. A serial parallel CNN is used and 74.90% and 56% best accuracy achieved for the SD and TL configurations.

Another interesting approach is proposed by *Liao et al.* [223], who consider three DL models that are adaptations of Shallow ConvNet: (i) multiple-local spatial convolution, (ii) global spatial convolution, and (iii) a parallel architecture combining (i) and (ii). Moreover, the authors consider a topographical representation of the EEG signal and augment the data by using a cropping and voting approach (sliding windows of length 2s, obtaining 74.6% using session 1 as the training data and session 2 as the test ones). *Liu et al.* [224] propose a SincNet-based hybrid neural network to improve the use of EEG information. The EEG signal is segmented and CSP applied. SincNets are used as filter bank bandpass filters, while the sparse representation of the data is obtained by a squeeze-and-excitation block. A CNN is used for deep feature representation and 74.26% accuracy achieved. Instead, *Ling et al.* [225] apply CWT using a Bump wavelet and generate time–frequency images of the various signals for each channel. A VGG-16 network is used and binary classifications deriving from the combination of the different conditions are considered, obtaining an average accuracy of $68.33 \pm 12.75\%$ (70% data used for training).

Zou et al. [226] propose a combination of broad learning and CSP, where *broad learning* is defined by the authors as "an effective and efficient incremental learning algorithm with simple neural network structure". The average accuracy is equal to 66.91%. Instead, *Xu et al.* [227] focus on different feature extraction strategies and provide deep multi-view feature learning through the usage of a deep restricted Boltzmann machine network improved by t-distributed stochastic neighbor embedding. SVM is used to classify the deep-learned features, obtaining 78.51% accuracy (with 10-fold CV).

Other works propose a dual alignment-based multi-source domain adaptation framework [228], a temporal-spectral-based squeeze-and-excitation feature fusion network [229], a temporal constrained sparse group Lasso regularization with EEGNet [115], a multi-scale convolutional transformer [230], and a natural evolution optimized DL method [231]. Other studies provide a combination of a feature-level graph embedding method with EEGNet [232], a temporal spatial convolution neural network fused with mutual information [233], a multi-domain CNN that learns subject-specific and electrode dependent features (time, spatial and phase domain) [234], a mutual graph network [235], and a filter bank Sinc-ShallowNet, while augmenting the data through an empirical mode decomposition based mixed noise adding methodology [236]. Other works propose a CNN combined with bidirectional GRU [237], a channel-wise convolution with channel mixing [238], an auxiliary multi-scale input CNN [239], a CNN-LSTM combined with a dynamic channel selection approach based on Davies–Bouldin index [240], a deep adversarial domain adaptation with few-shot learning model [241], and a DL method based on multi-task learning composed by representation, reconstruction and classification modules [242]. Moreover, some authors use publicly available architectures like InceptionV3 [243], Shallow ConvNet and/or EEGNet [244,245], and consider the combination of CNN and Riemannian geometry [246]. *Amin et al.* [247] consider firstly an attention-based Inception CNN to model spatial information, and then apply a BiLSTM to exploit the temporal ones. The best average accuracy is equal to 82.80% with a hold-out validation strategy.

Other studies consider the use of ANNs [92,112,248–256], different CNN models [96,106,118,121,126,147,151,155,210,250,254,257–306] that could present a TL mechanism [131], LSTM architectures [307–309], and a combination of CNN and LSTM models [310–313] or MLPs [314]. Moreover, some attention mechanisms are sometimes considered when proposing different DL models [247,315–319] as well as self-supervised learning [320], ELMs [113,321], or neuro-fuzzy classifiers [322].

4.8.3. Studies using a subject-independent paradigm

Besides the two works presenting available codes and previously described [63,80], other 16 papers use an SI approach and the reported results span from 76.52% to 99.96% accuracy. *Lee et al.* [323] propose a lightweight EEG-Inception squeeze-and-excitation network constituted by a depth-wise convolution to learn channel-wise features and squeeze-and-excitation blocks to manage such features. In fact, the authors' main aim is to decode channel-wise dependencies while having a small network. They use a 10-fold and a 5-fold CV for the training and evaluation data, respectively, and finally obtain 76.52% accuracy.

Kostas and Rudzicz [324] propose a CNN with semi-isolated temporal filtering and achieve 77.74% accuracy with a LOSO validation. A similar accuracy (77.40%) is obtained by applying a LOSO validation to a semi-supervised framework combining self-supervised contrastive learning and adversarial training [325].

Simpler methods appear in the next positions after this method. *Machida et al.* [326] use a CNN for classification after employing a transposed convolution as a pre-processor to set the window width and number of output features, achieving 77.83% accuracy. An accuracy of 78% is instead obtained by avoiding subject-specific model selection through the application of highway networks [327].

Siamese NNs obtains 79.05% accuracy considering the experimental conditions in pairs, while 80.07% accuracy is obtained using 5-fold CV on the left/right hand conditions performing analyses of continuous asynchronous on-line application of EEGNet [328].

Xie et al. [329] combine a LSTM GAN with a multi-output CNN, obtaining 83.99% accuracy (considering the dataset test sessions for testing their model), while a very similar result (84.00% accuracy) is achieved by a temporal attention convolutional network [330]. Instead, 92% accuracy is obtained through a hybrid neural network consisting of a CNN followed by a LSTM [331]. *Samanta et al.* [136] use an autoencoder for feature extraction and random forest for classification, obtaining $99.96 \pm 0.04\%$ accuracy (60% data for training).

Zaremba and Atyabi's approach [129], described in Section 4.4, and tested on numerous datasets, achieves $73.80 \pm 1.30\%$ accuracy considering a single-trial EEG. Moreover, it obtains $71.80 \pm 2.06\%$ and $67.20 \pm 2.10\%$ accuracy considering cross-subject within-dataset and cross-subject cross-dataset configurations, respectively.

Other studies propose an ensemble of CNNs [332] and achieve 64.34% accuracy with LOSO-CV, or consider a signal alignment before performing CNN regularization [333] obtaining 62.87% accuracy. Other approaches propose a two layers FFNN [334] or a supervised autoencoder [335].

4.8.4. Studies using both a subject-dependent and -independent paradigm

In this case, all the 14 studies remaining after screening will be reported without applying any kind of ranking. Details are not provided for works exploiting standard architectures, i.e., using simple CNN variations [336–339].

Considering the other studies, *Jia et al.* [340] propose a novel metric-based spatial filtering transformer and apply EEG pyramid for data augmentation. In particular, their model achieves 86.11% accuracy for the SD scenario, 61.92% for the SI one. Moreover, the authors train the feature extractor with *BCI-IV-2a* and fine-tune the classifier with *BCI-IV-2b* obtaining 83.38% average accuracy. Instead, *Yacine et al.* [341] develop *ARK-ANN*, a novel adaptive Riemannian kernel ANN, where a MLP classifies the covariance matrices of the EEG signals related to the MI tasks. The model achieves 87.40% and 77.30% for the SD and SI paradigm, respectively. Notice that for the SI case, a LOSO validation is applied. Another approach [342] considers a CNN with an end-to-end serial-parallel structure and afterwards applies a transfer learning strategy, obtaining $72.13 \pm 12.79\%$ for the SD paradigm (10-fold CV) and 56.02% with transfer learning for the SI paradigm.

Another proposal is provided by *Milanes et al.* [343], who develop an ensemble Shallow ConvNet using a Monte Carlo dropout to ameliorate the classification and provide uncertainty estimation. The final achieved accuracy values are 80.46% and 69.76% for the SD and SI paradigm, respectively. *Bria et al.* [344] propose a novel DL model called *Sinc-EEGNet*. They design it with four layers combining FBCSP and EEGNet. The first layer is set to use parameterized sinc functions that implement band pass filters. Notice that for the cross-subject experiment, the data are normalized using Z-score. Moreover, a LOSO validation is applied, considering only the training data of the subjects diverse from the target one for the training set, and the test data of the target subject for the test set. The SD paradigm obtains 70.56% accuracy, while the SI one 58.98%. Another novel proposal develops an attention-based 3D densely connected cross-stage-partial network model based on CNN [345]. The model achieves 84.45% and 64.53% accuracy for the SD and SI (LOSO validation) paradigms. Instead, *Wang et al.* [346] use one dimension-aggregate approximation to provide a suitable representation of the EEG signals to LSTM networks. They also consider the dataset conditions in pair, performing binary classifications: right vs. left hand, left hand vs. feet, left hand vs. tongue, right hand vs. feet, right hand vs. tongue, feet vs. tongue. The final achieved accuracy per classification task and considering the SD paradigm are as follows (5 times 5-fold CV): 69.70%, 69.40%, 69.30%, 69.50%, 69.60%, 77.30%. Similarly, for the SI paradigm the following accuracy values are achieved: 74.80%, 74.20%, 72.50%, 76.30%, 74.30%, 79.60%.

Another study considers the use of a TL-based multi-scale feature fused CNN and obtains $89.20 \pm 0.96\%$ and $91.61 \pm 0.88\%$ for the SD and SI (LOSO validation) paradigm, respectively. Instead, *He et al.* [347] firstly employ a spatio-temporal CNN module to extract features, then apply an ANN and consider a multi-task learning loss by supervised learning as well as an unsupervised metric learning. Their proposal achieves $71.52 \pm 12.61\%$ and $65.08 \pm 6.82\%$ for the SD (10-fold CV) and SI paradigm, respectively. *Liu et al.* [348] extract multiscale temporal and spatial features to better represent the MI neural pattern using an end-to-end 3D CNN. Therefore, the authors propose a mechanism to assign higher weights to channels and time points related to the motor activity. The devised model achieves 93.06% and 63.27% for the SD and SI paradigm, respectively.

4.8.5. Studies reporting TL-based or global performances

Considering other studies using TL or global paradigms, *Zhang et al.* [349] propose a hybrid deep neural network (CNN plus LSTM to decode spatial and temporal features) using transfer learning. The authors fine-tune the parameters for a single subject at a time and obtain an average TL accuracy of 81%. *Zhao et al.* [350] presents a deep representation-based domain adaptation method, which is constituted by modules deputed to feature extraction, classification and domain discrimination, achieving an accuracy of 74.75%. *Wu and Chan*'s [351] *Reptile-EEG* TL-based meta-learning model is proposed. This model achieves an accuracy of $66.03 \pm 12.28\%$ using a LOSO validation strategy and five-shot fine-tuning on the test subject. *Phadikar et al.* [111] (previously mentioned in Section 4.3) presented methodology based on the extraction of features through an autoencoder and a classification with an SVM achieves 97% global hold-out accuracy. *Lian et al.* [352] work in a global paradigm and apply a Shallow ConvNet and a BiLSTM to a merged global dataset of all subjects' session one data that is also augmented by an autoencoder, and report an average classification accuracy of $82.70 \pm 5.57\%$ on the session two (test) data. Few-shot learning with a convolutional attention module and a relation module is instead considered by *An et al.* [353], obtaining an accuracy value of $59.10 \pm 11.10\%$. Notice that the relation module is used to provide the final classification, which is based on relation scores computed between a support set and a query signal. Instead, the twin-cascaded softmax CNN devised by *Luo et al.* [354] achieves 80.03% global accuracy.

Jeong et al. [355] apply a subject-transfer decoding method based on a CNN. The CNN is pre-trained with other subjects' data and fine-tuned on the target subject. They augment the data of the 50% and propose a multi-model CNN concatenating a Shallow ConvNet, a Deep ConvNet and an EEGNet. Using session one for training (fine-tuning) and session two for testing, the authors obtain $86.54 \pm 7.78\%$ accuracy with subject-transfer. A test accuracy of $83.43 \pm 12.06\%$ is achieved by using CSP and a session-to-session domain adversarial NN [356]. The use of a data augmentation step is also presented by *Li et al.* [357], who obtain 86.12% accuracy by considering multiple spatial convolution kernels and sliding windows of 500 samples with step 20 samples on all the data (5000 entries for the test set) in the global paradigm. Finally, *Zhang et al.* [358], obtain 78.38% accuracy by proposing a filter bank Wasserstein adversarial domain adaptation framework. Notice that feature extraction is performed through a filter bank CNN model. Instead, *Chen et al.* [55] propose the use of a CNN and a mixup strategy to generate new samples, achieving 60.69% accuracy with LOSO validation.

4.9. BCI competition IV dataset 2b

The analyses of the papers remaining after the screening process for *BCI-IV-2b* [42] follow the rationale described at the beginning of Section 4.8. The 99 studies have the following characteristics: six works report available codes, 78 studies use the SD paradigm, 10 work on SI, five works use both paradigms, and the remaining studies consider a TL or a global approach.

4.9.1. Studies reporting available codes

As for Section 4.8.1, the studies here reported are listed in Tables 4 and 5. No ranking criteria have been applied, since the number of works providing codes and remaining after the screening process are six.

Starting from the less recent paper, *Tayeb et al.* [61] use Gumpy BCI toolbox[3] for signal preprocessing and develop three DL models, i.e., a (i) LSTM, a (ii) spectrogram-based CNN, and a (iii) recurrent CNN. The best average accuracy among the proposed models is achieved by employing the second one and is equal to $84.24 \pm 14.69\%$ with an SD approach (5-fold CV). Notice that by applying a Deep ConvNet the authors obtain $92.28 \pm 1.69\%$ accuracy. Instead, *Jia et al.*'s [66] multi-branch multi-scale CNN (Section 4.8.1) achieves $84.40 \pm 7.50\%$ accuracy (5 times 5-fold CV).

Pals et al. [67] apply an SI approach and using a spiking NN and a 5-fold CV obtain $75.63 \pm 12.25\%$ accuracy. A similar result (76% accuracy with LOSO validation) is achieved by *Zancanaro et al.* [72] *DynamicNet* and considering an SD paradigm. Notice that this and the

[3] https://github.com/gumpy-bci (last accessed on August 02, 2023).

following studies have been described in the previous Section 4.8.1. *Faria et al.*'s [76] approach, exploiting different data augmentation techniques, obtains $76.40 \pm 14.20\%$ accuracy, while *Song et al.* [81] EEG Conformer achieves 84.63% accuracy.

4.9.2. Studies using a subject-dependent paradigm

The majority of the studies remaining after the screening process consider the use of an SD paradigm, thus the papers with the highest reported accuracy are firstly discussed and other studies with novel and interesting approaches from the authors' perspective are then introduced.

Starting from the top 20 works, the reported results span from 87.33% to 96.82% accuracy. *Shen et al.* [284] multi-scale Siamese CNN is related to the first reported accuracy value, while *Huang et al.* [269] consider the use of an EEGNet-based CNN and supervise the training process by monitoring the model loss. By considering a data division as for the competition (i.e., training and evaluation set), they obtain an accuracy of the $87.52 \pm 8.54\%$.

Lashgari et al. [183] apply a CNN-based neural network with an attention mechanism, achieving 87.83% accuracy, while a time-contained spatial filtering and spatial–temporal analysis network obtains 88% accuracy, considering the first session trials for training and the second session ones for testing [209]. *Zhang et al.* [200] *EEG Inception* model, using data augmentation on the training set, achieves $88.58 \pm 5.50\%$ accuracy, *Dokur and Olmez* [118] proposed data augmentation with noise addition on the training set, CSP application, and CNN use, obtaining 88.60% accuracy ($\kappa = 0.7720$) considering 80% data for training and 20% for testing.

Tian and Liu [359] achieve the same accuracy (89.07% with 10-fold CV), while firstly using the short-time Fourier transform outputs combining time, frequency and location information to extract band-pass and power spectral density features, which are given as inputs to a CNN. Secondly, the authors test two different CNNs: (i) a single-input CNN considering as input the results of a short-time Fourier transform, and (ii) a multiple-input CNN with a 3D input [360]. The use of a combination of frequency, temporal, and spatial features is also considered by *Liang et al.* [126], who apply a CNN and obtain 89.19% accuracy (10-fold CV).

A very different approach is proposed by *Gomes, Rodrigues, and dos Santos* [361], who firstly pre-process the signal by windowing and filtering it. Afterwards, the authors apply two feature extraction methods: (i) using sinogram images and applying VGGNet and LeNet, and (ii) extracting numerical features directly from the signals. The features are then selected using evolutionary search and a random forest classifier is applied for the final classification. Note that SMOTE (synthetic minority oversampling technique) is used for data augmentation and $92.47 \pm 1.19\%$ accuracy achieved ($\kappa = 0.8500$) with 10-fold CV.

Instead, 92.56% and 93.08% accuracy values are obtained by a dual alignment-based multi-source domain adaptation framework [110], and a CNN receiving the outputs of a short-time Fourier transform [98], respectively. *Wankhade and Chorage* [186] *RideNN* classifier, described in Section 4.8.2, achieves 93.30% accuracy, while *Cai et al.* [362] combine a CNN with a gated recurrent unit, obtaining 93.57% accuracy. *Tang et al.* [363] consider seven subjects of *BCI-IV-2b* and propose a semi-supervised KNN-based smooth autoencoder model achieving 94.81% accuracy (with 5-fold CV), while *Malibari et al.* [107] proposal, described in Section 4.2, obtains 96.14% accuracy.

Two other studies achieve 96.50% accuracy with a 10-fold CV. The first one [190] presents a data augmentation step based on linear interpolation before using a plain CNN, while the second one [141] considers graph embeddings and CNNs (a detailed description is provided in Section 4.4). *Wang et al.* [196] obtain 96.77% accuracy, using variational sample-long short term memory to perform multi-band decomposition and spectral discriminative analysis for MI classification. Finally, *Li et al.* [199] use fast Fourier transform and Clough–Tocher interpolation to generate the spatio-frequency images, inputted to a modified VGGNet, and obtain 96.82% accuracy ($\kappa = 0.9400$) with 10-fold CV.

Besides these top 20 studies according to the accuracy ranking, other interesting approaches have been presented for *BCI-IV-2b*. In particular, *Xu et al.* [364] propose a deep transfer CNN framework based on VGG-16. The inputs are images with dimension $224 \times 224 \times 3$, where the third dimension corresponds to the C{3,4,z} channel images obtained through short-time Fourier transform. Notice that the authors exploit the datset in two ways: (i) considering the signals from second 0.5 to 3.5 after the stimulus, and (ii) considering the signals from second 1 to 4 after the stimulus. The best achieved average accuracy is equal to 74.20%. *Bang et al.* [202] proposal, based on filters and 3D CNNs and tested on *BCI-IV-2a* (details in Section 4.8.2), achieves $75.85 \pm 12.80\%$ accuracy, while *Xie et al.*'s approach [365] obtains 78.22%. The authors propose a novel method, called parallel stacking encoded convolutional network, to extract features and classify them. The authors combine a stacked denoise autoencoder with a CNN, and thus consider both an unsupervised and a supervised learning approach.

Another study [366] presents the use of sparse spectro-temporal decomposition to extract time–frequency features, and then the application of a CNN with squeeze-and-excitation blocks, which are used to recalibrate channel-wise feature responses. The final achieved accuracy is equal to $79.30 \pm 1.60\%$. Instead, *Chen, Wang, and Song* [367] bandpass (8–30 Hz) the signals and apply wavelet transform to generate time–frequency images. Notice that the authors use only the signals recorded by the C3 and C4 electrodes. Afterwards, image subtraction of these two channels is performed and the resulting image fed to the model, which is a mix of convolutional blocks and attention module. After 10-fold CV, the proposed strategy obtains $79.60 \pm 1.80\%$ accuracy ($\kappa = 0.5920$). Similarly, CWT is used to produce concatenated time–frequency images of each channel, which are then fed to a CNN model, obtaining the best average accuracy $83.00 \pm 1.60\%$ (10-fold CV considering only the first three sessions) considering both μ and β rhythms [99].

An accuracy of the 80.78% is instead achieved by exploiting a CNN model for deep two-dimensional time–frequency maps feature extraction and different traditional ML models used for classification [109]. Similarly, a multi-attention CNN is used for feature extraction [214] and tuned by an inter-session discriminator loss to reduce inter-session variability. The classification resulted in an accuracy of the 82.54% ($\kappa = 0.6510$). *Hong et al.* [213] dynamic joint domain adaptation network based on adversarial learning (Section 4.8.2) achieves 83% accuracy, while *Han et al.* [175] parallel CNN (Section 4.7) obtains $83.00 \pm 3.40\%$ accuracy ($\kappa = 0.6590$), after applying for five times 10-fold CV.

Huang et al.'s [219] *RSBConvNet*, described for *BCI-IV-2a* (Section 4.8.2), achieves 83.04% accuracy, dividing the dataset in train (80%) and test (20%) sets. Another model described for the same dataset, i.e., a SincNet-based hybrid neural network [224], obtains 83.49% accuracy. *Tao et al.* [368] use short-time Fourier transform to produce time–frequency graphs and give them in input to a deep convolutional GAN. The proposed model achieves 85.70% accuracy considering 70% data for training. Instead, *Yang et al.* [212] discriminative feature learning strategy and a deep adversarial domain adaptation with few-shot learning framework [241] (Section 4.8.2) obtain $85.46 \pm 10.44\%$ and 84.63% accuracy, respectively. To conclude, *Zhao et al.* [350] deep representation-based domain adaptation method achieves an accuracy of 83.98% ($\kappa = 0.6796$).

Finally, the remaining screened studies are quickly reported for the sake of completeness. Different works use CNN-based architectures [100,259,262,265,292,306,315,336,369–380] with transfer learning strategies [381], variational or stacked autoencoders [103, 382,383], generative adversarial networks [384], and pyramid pooling [289]. Moreover, benchmark models are also considered as they are or with some variations, like Shallow ConvNet [97,286]. Other studies propose a frequential deep belief network [385], the use of filter bank CSP as inputs to gate recurrent units or LSTM models [309], a weight sharing CNN-LSTM model [386], a LSTM autoencoder [387], a distribution-based learning network [388], multi-kernel or sparse Bayesian ELMs [145,389], capsule NNs [390–392], a dual-attention-based adversarial network [319], and the temporal-rearrange based MI-EEG network *TRMINet* [156].

4.9.3. Studies using a subject-independent paradigm

The studies remaining after screening and considering a SI paradigm are 10, thus no ranking will be applied. Moreover, the works described in previous sections will not be detailed, but cross-references provided to avoid an excessive paper length. A CNN ensemble [332] (Section 4.8.3) achieves $64.34\pm8.31\%$ with a LOSO validation strategy. Other studies reported in Section 4.8.3 are *Xie et al.* [329] LSTM GAN with a multi-output CNN, achieving 94.31% accuracy; *Li et al.* [357] spatial convolution kernels, obtaining 75.53% accuracy; and *Chen et al.* [55] CNN model, which achieves 76.65% accuracy. *Liu et al.*'s [130] subject adaptation CNN, which is described in Section 4.4, obtains $86.42\pm5.42\%$ accuracy.

The remaining studies present diverse approaches. *Zhang et al.* [393] propose an instance transfer subject-independent framework based on perceptive Hash algorithm to measure the similarity between the spectrogram of the EEG signals of different subjects and combine it with a CNN. This proposal achieves $94.70 \pm 2.60\%$ accuracy with 10-fold CV. Instead, *Roy*'s [394] multi-scale CNN obtains 93.74% accuracy performing data augmentation and considering intrinsic feature integration. Notice that the first two sessions have been used to train the model, the third session for validation and last two sessions for evaluation.

Yang et al. [395] devise a framework capable of capturing spatial and spectral dependencies of the signals and consider a combination of CNN and a recurrent NN LSTM. The data are augmented and $76.44 \pm 6.60\%$ accuracy is achieved, considering first three sessions for training and the remaining two for testing. Finally, Bayesian hyper-parameter optimization is applied to a CNN, obtaining $76.37 \pm 13.91\%$ accuracy [396].

4.9.4. Studies using both a subject-dependent and -independent paradigm

Considering the use of both SD and SI paradigms, *Wu et al.* [339] propose a parallel multiscale filter bank CNN and obtain 84.30% and 84.70% accuracy for the two paradigms, respectively. *Jia et al.* [340] approach, described in Section 4.8.4, achieves 83.38% accuracy considering a cross-task classification, while 88.39% is achieved with a SI configuration. Instead, *Milanés-Hermosilla et al.* [343] ensemble Shallow ConvNet, described as well in Section 4.8.4, provides the following accuracy values: 78.58% for the SD and 78.20% for the SI paradigm.

Zhu et al. [397] employ CSP handcrafted features and propose a separated channel CNN. By applying a LOSO validation strategy, the authors obtain 64% and 83% accuracy for the SD and SI approaches, respectively. Finally, a Shallow ConvNet based CNN [155] considering the last two sessions of the dataset as the training set, achieves an accuracy of 77.50% for the SD paradigm and of 75.33% for the SI one.

4.9.5. Studies reporting TL-based or global performances

Considering other studies using decoding paradigms diverse from the SD and SI ones, the twin-cascaded softmax CNN [354] reported in Section 4.8.5 achieves 88.05% global accuracy. *An et al.* [353] few-shot learning algorithm (Section 4.8.5) obtains instead $74.60\pm10.20\%$ global accuracy with a 9-fold CV.

Finally, *Sun et al.* [398] propose a network constituted by a generator and a CNN. The authors' main aim is to transfer the features of BCI-friendly subjects to BCI-illiterate ones, obtaining $80.70 \pm 6.10\%$ accuracy with 10-fold CV.

4.10. Cho2017

Cho et al. dataset [26] provides data of left/right hand MI collected on 52 subjects, thus with a greater numerosity in respect to the previously reported datasets. The remaining studies using *Cho2017* are 13, of which two consider a SI paradigm, two a transfer learning strategy, and the rest work on the SD paradigm. TL is used by *Collazos-Huertas et al.* [65] (which also appears on Table 4). CWT is applied to generate signal topograms to be used by a CNN for feature extraction. Afterwards, a TL strategy is applied for classification with an MLP. The pre-training is based on kernel matching with subjects' features and with the questionnaires available as part of the dataset. Applying 10-fold CV, and considering 10% of the data for the test set, the authors obtain $79.50\pm10.80\%$ accuracy considering an SD approach without TL, while $82.60 \pm 8.40\%$ applying the TL strategy.

Another TL approach is employed by *Wu and Chan* [351], who develop a model called *Reptile-EEG*, which exploits the meta-learning algorithm Reptile and integrates it to a deep neural network. Notice that a LOSO validation is performed and that the authors declare that the best accuracy is achieved by EEGNet, used as benchmark model. *Reptile-EEG* best accuracy is equal to $70.73 \pm 12.18\%$.

A methodology exploiting a dataset-transfer learning mechanism is also proposed by *Xu et al.* [71], whose method was discussed in Section 4.8.1 and code link listed on Table 4. In the SD paradigm, the best accuracy using EEGNetv4 with Euclidean alignment is around 74%. Instead, in the cross-subject (SI) scenario, the best accuracy was obtained by using *BCI-IV-2a* as the source dataset, Shallow ConvNet with Euclidean alignment, and adaptive batch normalization, and is around 71.60%.

The SI paradigm is present in two studies [399,400]. In the first case [399], the highest average accuracy (62.60%) is obtained by the application of Shallow ConvNet, considering 80% data for training and 20% for testing. Instead, *Jeon et al.* [400] use a deep network to estimate the mutual information between feature representations by decomposing features in class-relevant and -irrelevant ones, and enhancing class discrimination of feature representations. Notice that each subject is considered as one domain, and that two DL methods were used, i.e., Deep ConvNet and EEGNet. The final layer is the global encoder, while the previous one a local encoder. The author exploit a concat-and-convolve architecture and considered a cross-subject learning and zero-training scenario. Notice that the best average accuracy ($76.60 \pm 12.48\%$) in the first scenario is obtained by EEGNet, which achieves the best average accuracy also for the zero-training scenario with $73.73 \pm 13.75\%$ accuracy.

Considering the SD studies, *Kumar et al.* [59] proposal (Section 4.7) achieves $68.19 \pm 9.06\%$ accuracy ($\kappa = 0.3740$). Instead, *Wankhade and Chorage* [186] *RideNN* classifier described in Section 4.8.2, achieves 92.02% accuracy.

Santos, San-Martin, and Fraga [401] consider two feature extraction methods, i.e., CSP and LORETA. The authors verify that they can achieve better results using CSP features and in particular they obtain $91.60 \pm 8.90\%$ (10-fold CV) feeding them to a MLP.

Finally, CNN models are applied by the remaining works, considering channel selection [402] or data augmentation preceding the classification [50], or proposing the use of CWT [166] and CSP features [185, 403].

4.11. EEG motor movement/imagery database

eegmmidb[4] is the third most frequently used dataset for the evaluation of MI BCI decoding models with 43 remaining studies after the

[4] This dataset is sometimes called the "PhysioNet" dataset as it is part of the large PhysioNet (https://physionet.org/) physiological signals database. This dataset consists of 109 subjects out of which four, five, or six subjects have been reported to have data annotation errors, and therefore have been excluded by some of the studies during MI-EEG decoding. Note that we only considered studies that use the motor imagery portion of this large dataset.

screening process. Among the 43 studies, there are eight works that have made their code publicly available. Categorizing the studies based on the decoding paradigms, 11 studies work on the SD paradigm, eight studies on the SI paradigm. The remaining works report global or TL-based performance measures. Similarly to previous sections, only cross references are provided for studies and model architectures already discussed to avoid repetitions.

4.11.1. Studies reporting available codes

As for Sections 4.8.1 and 4.9.1, the studies with an available code are referenced to Tables 4 and 5. In what follows, eight studies from these lists and using *eegmmidb* are described in chronological order.

Dose et al. [58] utilize temporal and spatial 1D-CNN layers for feature extraction followed by pooling and fully connected layers for classification. They report decoding accuracy values for both the global and the TL paradigm (with fine-tuning on the test subject) considering the classification of two, three, and four classes of *eegmmidb* and using 5-fold CV. They report a mean global accuracy of 59.71% and a mean TL accuracy of 68.93% for the classification of four MI classes. *Roots et al.* [62] propose a multi-branch CNN with EEGNet style branches, termed EEGNet Fusion. The model achieves an SD mean accuracy of 83.80% on 103 subjects considering only two classes of left and right hand with hold-out train/validation/test split. *Zhang et al.* [63], using the model architectures described in Section 4.8.1, report the best SI accuracy of $74.41 \pm 4.19\%$ using 95 subjects for training and 10 subjects for testing. Using a 10-layer 1D CNN architecture and SMOTE for data augmentation, *Mattioli et al.* [79] split the EEG channels into six regions of interests (ROIs) and report global accuracy values for each ROI. The authors claim that one of these ROIs achieves a performance of 99.38% for the global paradigm.

Considering the global paradigm, *Yue et al.* [404] map filtered EEG signals (μ and β bands are preserved using a wavelet transform) to images by a trivariate Clough–Tocher scheme [405] that also takes into account the geometry of electrodes on the head. The authors propose Denoised-ConvNet consisting of color space and spatial transformations, and CNN with fully-connected layers, for classification. The reported global accuracy using a one-vs-rest classification scheme for the five-class (MI plus rest state) *eegmmidb* task is 99.06%. *Zancanaro et al.* [72] using their proposed EEGNet-based tool DynamicNet (see Section 4.8.1) achieve a performance of 83% on the SI binary classification task of right and left hand movements with LOSO-CV. Instead, *Xu et al.* [71] achieve an SD performance of 73.10% with EEGNetv4, Euclidean alignment and adaptive batch normalization, using the same approach discussed in Section 4.10. Furthermore, they obtain a cross-dataset TL accuracy of 67.10% with *BCI-IV-2a* as the source dataset, EEGNet with Riemannian alignment and adaptive batch normalization as the model architecture. *Alnaanah et al.* [83] (see Section 4.8.1) report an SD accuracy of 58% and $\kappa = 0.4750$. Finally, *Fadel et al.* [406] apply a Clough–Tocher interpolation algorithm to derive 2D images from EEG signals, and then use a CNN followed by an LSTM for classification. They report a mean SI accuracy of 70.64% in a LOSO-CV scenario with another subject's data as the validation set on 103 subjects.

4.11.2. Studies using a subject-dependent paradigm

The papers providing an SD performance are reported in descending order with respect to the performance value, giving priority to the papers that consider the three and four-class tasks instead of a binary (left/right hand MI) classification task. Remind that due to different performance evaluation strategies, e.g., hold-out or cross-validation, these performance measures may not be directly compared in all cases.

There are five studies with performance evaluation on the classification task with four classes of the dataset, and one on a classification task with three classes of the dataset. *Sorkhi et al.* [407] extract features through multi-scale FBCSP and pass them through a CNN after applying a Hilbert transform. Their hyperparameter selection is done through Bayesian optimization which helps them achieve a high SD performance (98.02%) using 10-fold CV. *Li and Ruan* [142] (see Section 4.4 for model details) achieve an SD accuracy of 95.09% on 109 subjects with 10-fold CV. *Lomelin-Ibarra, Gutierrez-Rodriguez, and Cantoral-Ceballos* [408] take advantage of a backbone ResNet-18 [409] followed by a three-layer CNN to classify spectrogram images. They report as their best performance an accuracy of 93.32% using 80/10/10 train/validation/test split on 105 subjects. *Wang et al.* [410] propose a residual learning attention CNN which consists of CNNs with residual connections followed by attention modules and max pooling layers. They also use a backbone CNN with residual connections for feature extraction from only nine channels relevant to MI. Excluding four subjects with erroneous annotations, they report SD accuracy values on the binary (left/right hand), three-class (left/right hand, and rest), and four-class (left/right hand, rest, and both feet) tasks on 105 subjects with 84.69%, 84.81%, 75.85% reported accuracy values, respectively.

Alwasiti, Yusoff, and Raza [411] propose a deep metric learning framework with a DenseNet-121 [412] as its body. Stockwall time–frequency transform, which is a generalization of short-time Fourier transform is used to create images from EEG signals. For deep metric learning, the anchor output is compared to positive and negative instance images and the loss is calculated to make the output embedding similar to the positive instance and different from the negative one. The authors report an SD accuracy of $64.70 \pm 1.22\%$ with a 80/20 hold-out evaluation method. Finally, considering a three-class task (left/right hand and rest), *Alwasiti and Yusoff* [413] proposed mix-up data augmentation with ResNet-18 [409] and DenseNet-101 [414] backbones to classify EEG spectrograms. They obtain the best performance ($93 \pm 1.00\%$) using the DenseNet-101 backbone and hold-out SD evaluation.

Concerning studies that work on the left/right hand classification task of the *eegmmidb* dataset in the SD paradigm, *Nekrasova, Kanarskii, and Sudareva* [415] use a simple MLP fed by CSP features and report a 4-fold CV 99% accuracy on 103 subjects. *Awais et al.* [416] use 91 out of 109 subjects, declaring the presence of too contaminated recordings. The authors extract channel connectivity features using partial directed coherence and directed transfer functions. The classification is performed using a probabilistic NN. The authors report a high SD accuracy of 98.65% with 10-fold CV.

Jin et al. [141] (see Section 4.4 for details) achieve an accuracy of 80.20% using 6-fold CV. *Kang et al.* [417] apply non-linear analyses such as sample entropy, permutation entropy, and recurrent quantification analyses to extract subject-specific features from EEG signals. Afterwards, they compare the performance of FBCSP with three DL models including EEGNet, Shallow and Deep ConvNet. The best performance is achieved by Shallow ConvNet ($63.50 \pm 1.16\%$) in the binary SD task using 5-fold CV on 104 subjects. Finally, an interesting approach used in the SD and TL paradigms is the one by *Ju et al.* [418], who consider a federated learning framework consisting of three non-federated stages for feature extraction including a manifold reduction layer, a common embedded space, and a tangent projection layer, followed by a federated layer (i.e., a NN whose weights are updated by federated aggregation). Considering the binary classification task, the authors report a performance of 60% in the SD paradigm.

4.11.3. Studies using a subject-independent paradigm

As the number of studies using the SI evaluation paradigm is only 11, all of them are discussed in this subsection. *Kostas and Rudzicz* [324] consider a CNN with semi-isolated temporal filtering and a novel learning technique called multi-domain learning with an SI performance of 67.60% (LOSO-CV with one subject's data used as the validation set), and a TL performance of 78.73% on 105 subjects. *Wang et al.* [419] develop a simplified EEGNet architecture and deploy it on low-power microcontroller units that are suitable for edge-computing. The authors report an accuracy of 65.07% considering 5-fold CV with each fold being a subset of 21 subjects among 105 subjects of the dataset. The authors also provide statistics regarding computation power and time as the study is intended to measure the capabilities of edge-computing in BCI. Instead, *Vivek et al.* [420] propose a graph neural network to extract spatio-temporal features. Afterwards, the features are transformed into temporal embeddings and their graph representations are given in input to a graph NN, from which node embeddings are produced and given to a final classifier. The authors report an accuracy of 82.92% using 5-fold CV on the binary classification task of right and left hand MI.

Xie et al. [421] propose a combined temporal CNN-transformer architecture and measure the effect of different positional embeddings on SI performance with 3 or 6s EEG signal segments. The best performance (68.54%) is obtained using 6s segments and relative positional embeddings for the four-class task (left/right hand, eyes open, and both feet). *Strahnen and Kessler* [328], using the same methodology explained in Section 4.8.3, report an accuracy of 77.40%. *Özdenizci et al.* [422] design an adversarial variational auto-encoder and train it on 90 subjects' data. The trained variational auto-encoder is used in both SI and TL paradigms for the other 13 subjects. Their mean SI accuracy on the test subjects is 56.90%. Finally, *Sun et al.* [189] (see Section 4.8.2 for architecture details) consider a SI paradigm with 10 out of 105 subjects as the test set and report an accuracy of $82 \pm 3.60\%$ over ten independent runs and on the binary classification of left/right hand MI.

4.11.4. Studies reporting TL-based or global performances

For the final section on the *eegmmidb* dataset, the studies considering TL and global paradigms are discussed.

Considering some of the studies reported for the previous Sections 4.11.2 and 4.11.3, *Ju et al.* [418] consider a TL approach and achieve an accuracy of 54.90%. *Wang et al.* [419] report a transfer learning accuracy by using a model trained on other subjects and fine-tuning it on a specific subject. Using 4-fold CV they report a TL accuracy of 70.83%. Similarly, *Özdenizci et al.* [422] provide a mean TL accuracy of 63.80%.

Considering the global paradigm, *Chu and Zhang* [423] use a combination of BiLSTM, an attention module, and a CNN layer for decoding, and report an accuracy of 88.30% on the binary classification task of left/right hand MI using 70/30 hold-out train/test split. *Shah, Albishri, and Lee* [424] used a slightly modified EEGNet architecture and proposed an Internet of things framework. Excluding six subjects with erroneous annotations, they report a global accuracy of 72.82% on the binary classification task of both hands versus feet using a 80/20 train/test split for model evaluation. *Sun, Xie, and Zhou* [425] consider transformer-based models and report global accuracy values using 21 recordings from each subject and 5-fold CV. The best global accuracy on the four-class categorization (68.54%) task is achieved using a combination of CNN and temporal transformers. *Wu and Chan* [351] Reptile-EEG meta learning strategy (see Section 4.10) with a five-shot fine-tuning strategy and using 5-fold CV on 102 subjects (six subjects excluded having erroneous annotations), achieves $68.49 \pm 13.91\%$ and $68.60 \pm 14.36\%$ on the binary classification tasks of left/right hand and fists/both feet MI, respectively. *Zhang et al.* [133] proposal (see Section 4.4) achieves an accuracy of $93.36 \pm 1.68\%$ using 10-fold CV on the global decoding paradigm. *Dose et al.* [426] utilize a shallow spatio-temporal CNN similar to the one used by *Schirrmeister et al.* [37] and report a global accuracy of 65.73% on the four-class task using 5-fold CV and 6s segments of the EEG signals. A global accuracy of $58.59 \pm 14.67\%$ is obtained when considering the first 3s of the signals. The authors adopt a TL approach to fine-tune the global model on each subject before testing for five epochs. The TL accuracy for the four-class task (left/right hands, both fists, and both feet) is reported as $68.51 \pm 12.56\%$ for 3s segment signals. With a similar evaluation strategy, *Hernandez-Ruiz et al.* [427] using an EEGNet-style architecture implemented on low-energy FPGA hardware achieve a global test accuracy of 83.15%, 75.74%, and 65.75%, on four, three and two-class tasks, and 93.10%, 93.21%, and 89.23% in the TL paradigm. Considering the binary task of classifying right and left hand MI, *Wang and Li* [428] propose a parallel deep CNN whose input is spatio-temporal representation of the EEG signals. Each branch of the parallel CNN corresponds to one of the extracted sub-band spatio-temporal features (based on Clough–Tocher interpolation) and achieve a global accuracy of 90.52% using hold-out validation. *Du, Liu, and Tian* [429] apply data augmentation to generate more training data by superimposing (summing) and normalizing EEG signals, and afterwards use simple spatio-temporal CNN for classification. The data augmentation approach proves effective, achieving a global accuracy of 66.36% on 104 subjects via 5-fold CV. A mean TL accuracy of 74.60% is obtained using 4-fold CV and the fine-tuning the global model for different subjects.

Considering a 3D CNN, *O'Neill et al.* [430] achieve a global test accuracy of 92.23% on the five-class classification task (left/right hands, both feet, fists, and rest), while *Li et al.* [370] (see Section 4.8.2) obtain a global accuracy of 88.62% and a $\kappa = 0.7700$ on 109 subjects using 10-fold CV. *Li, Li, Yang, and Du*, and *Li, Shi, and Li* [431,432] use a spatio-temporal CNN, once followed by a BiLSTM and once by GRU units for classification achieving a mean global performance of 98.09% and 97.76% on 108 subjects, respectively using 75/25 hold-out evaluation. *Khetrapal and Kadambari* [433] apply three different feature extraction methods, i.e., power of the signal, fast Fourier transform, and power spectral density. These features are fed to a multi-branch CNN. The authors report a global accuracy of 98.39% using 70/30 train/test split.

Li et al. [434] use gradient-class activation mapping [435] for channel EEG selection and perform classification using a CNN followed by GRU units. The authors report a global performance of 97.36% on the five-class task (left/right hands, both feet, fists, and eyes closed) on 108 subjects with 75/25 train/test split. Finally, *Wang and Li* [220] (see Section 4.8.2) report a global accuracy of 96.91% on the binary classification task using 10-fold CV.

4.12. Kaya2018

Kaya2018 [44] considers different MI conditions, comprising hands, feet, and tongue imagined movements. Of the six studies remaining after the screening procedure, two consider a TL strategy and one a SI paradigm. For this last case, *Pérez-Velasco et al.* [54] propose a DL architecture, called *EEGSym*, with the main aim of overcoming the inter-subject variability problem. The final achieved accuracy is of the $85.10 \pm 9.50\%$.

Considering the TL strategies, *George et al.* [52] apply TL with three different networks, i.e., a multi-branch CNN with final concatenation and convolutional layers, a Deep ConvNet, and a Bi-GRU. Transfer learning is applied considering two scenarios. The first is within-subject TL, where cross-session transfer is applied and then a fine-tuning is performed on another session. The second is cross-subject TL in which the model is trained on all of the subjects except the target one, and then is fine-tuned on the target subject. The within-subject cross-session TL achieved an accuracy of $76.17 \pm 14.62\%$ using Deep ConvNet, and the cross-subject TL obtained an accuracy of $80.01 \pm 10.00\%$ using the Bi-GRU model.

Chen et al. [55] cross-subject TL strategy achieves 68.01% accuracy with LOSO validation. In this work, a CNN is shared between the source and the target domain, and a novel mix-up strategy is also developed to provide new artificial samples by mixing two random samples in the frequency domain. Concerning the SD paradigms, two works from similar groups of authors [50,51] consider different data augmentation approaches followed by the application of one ore more CNN models. Another peculiar data augmentation strategy preceding CNN classification is proposed by *Zahra et al.* [53], who consider a data augmentation of the training set through sliding windows, noise addition, or their combination. Notice that their best average accuracy is achieved with the combined augmentation and is equal to $57.50 \pm 7.90\%$ (10-fold CV).

4.13. OpenBMI

Another binary MI experimental paradigm is proposed for *OpenBMI*, which contains data of 54 subjects rehearsing the grasping of left and right hand. The screening on *OpenBMI* resulted in 21 studies, of which eight with an SD paradigm, 10 with a SI paradigm, and three using both SD and SI approaches.

Considering the SD only works, the best result in terms of accuracy (93.81%) is achieved by an AlexNet-based CNN [158] after applying a 10-fold CV. The authors employ a multi-scale principal component analysis, a novel empirical Fourier decomposition signal resolution method with Hilbert transform, and use four pre-trained CNNs for automatic feature extraction and selection.

Besides the use of CNN architectures with spectrally localized time-domain representation of the signals [436], filter-bank or end-to-end filter-bank multi-scale CNN models [287,437], other SD approaches rely on TL strategies. For example, *Zheng and Yang* [211] idea of introducing an adaptive layer into the fully connected layer of a deep CNN, described for *BCI-IV-2a* (Section 4.8.2), achieves a best average accuracy of the $76 \pm 0.13\%$ by applying 10-fold CV for 10 times.

Instead, *Ju and Guan* [221] geometric DL framework *Tensor-CSPNet* described in Section 4.8.2, considering a cross-validation strategy on the first and second sessions of *OpenBMI* and a hold-out validation using the second session as the test set, obtains $74.95 \pm 15.27\%$, $75.92 \pm 14.63\%$, and $69.65 \pm 14.97\%$ accuracy, respectively.

In *Bang et al.* [202] work, cited in Section 4.8.2, $70.37 \pm 17.09\%$ accuracy is obtained by using the proposed 3D CNN and applying a hold-out validation strategy. Finally, *Yang et al.* [82] work already described in Table 5 and in the *2020-International-BCI* dedicated Section 4.1, achieves 75.40% accuracy with the proposed end-to-end CNN (5-fold CV).

For what concerns the SI-based studies, two of them present available codes [75,80], as reported in Table 5. The first work [80] proposes a novel model called explainable Inception temporal convolutional network and achieves 76.19% accuracy by applying a 10-fold CV. The same validation strategy is proposed by the second study [75], which considers a multi-subject ensemble CNN with majority voting and obtains $85.56 \pm 11.63\%$ accuracy.

Other studies use different CNN models like *EEGsym* [54], ResNet [88], considering relevance-based channel selection [438], or a hemisphere discrepancy network [439].

Additional works analyze the efficacy of network pruning. *Zhang and Li* [440] propose a DL model composed by a convolution (time), deep-wise (space), average pooling, flatten, and fully connected layer, considering exponential linear unit as the activation function. The authors obtain 62.70% accuracy after using a fast recursive algorithm to prune redundant parameters. Instead, *Vishnupriya et al.* [441] use Deep ConvNet and apply a magnitude-based weight pruning on pre-trained models. The evaluation is conducted through the LOSO methodology, and the data from 53 subjects are divided in the training (85%) and validation (15%) sets. The remaining subject's data are used as the test set. The model with the highest validation accuracy is saved and pruned for different sparsity levels (10%, 20% to 90%). The pruned models are evaluated on the last 100 trials of the target subject's data. The average accuracy obtained on the baseline model without pruning is equal to $84.46 \pm 11.39\%$, while the best one with pruning (sparsity level 70%) is equal to $85.53 \pm 11.33\%$.

Another peculiar approach proposed by *Xu, Yao, and Ni* [442] focuses on an attention mechanism. The authors custom-define a form of sequence inputs with spatial and temporal dimensions. This form is adopted for dual headed attention via a deep convolution network, which simultaneously learns temporal and spatial features. Afterwards, the features of spatial attention on each input head are divided in two parts for spatial attention learning. The best average accuracy achieved by the proposed model is equal to $75.52 \pm 11.72\%$ with 10-fold CV.

The final work using only an SI paradigm is the one from *Jeon et al.* [400], presented in the *Cho2017* dedicated Section 4.10. The model achieves $76.67 \pm 13.01\%$ accuracy considering a cross-subject learning scenario and using Deep ConvNet, while it obtains $73.32 \pm 13.55\%$ accuracy using the same DL model in the zero-training scenario.

The last part of this section is dedicated to the remaining three studies using both SD and SI paradigms. *Ko, Jeon, and Suk* [443] propose a reinforcement learning assisted DL framework. Its pipeline follows three main steps: (i) estimation and selection of reliable signals, (ii) use of an actor critic model, (iii) application of a reinforcement learning based feature selection. For the DL embedding networks the authors exploit Shallow/Deep ConvNet, EEGNet, and multi-scale NN. Using the multi-scale NN with the proposed strategy they achieve $77.26 \pm 13.92\%$ and $75.24 \pm 17.40\%$ accuracy for the SD training/test and SI (LOSO-CV) approaches, respectively.

Autthasan et al. [64] provide their codes to the research community to reproduce the results achieved by their novel DL model, called *MIN2Net*. The filtered EEG data are passed through a CNN autoencoder. Afterwards, the latent vector of the encoder is used to train a deep metric learning network to keep vectors corresponding to the same label near and the ones corresponding to the opposite labels far. The latent vector is also used to train a classification artificial NN. The final results consist of $61.03 \pm 14.47\%$ accuracy for the SD paradigm, and $72.03 \pm 14.04\%$ for the SI one. Finally, a compact SI MI framework combining temporal convolution and CSP for sequential extraction of spatio-temporal features is proposed by *Nouri et al.* [84] who have made available their codes to the research community (Table 5). Using a LOSO approach, they achieve $74.41 \pm 16.75\%$ and $74.28 \pm 16.12\%$ for the SD and SI paradigms, respectively.

4.14. Other datasets

This section reports the studies resulting from the manual screening of MI datasets that were used by less than five works. These datasets are *Jeong2020* (one paper), *MBT-42* (two papers), *MED-62* (three papers), *MIDistraction* (one paper), *Nikki2021* (one paper), *SameLimb* (two papers), *UpperLimb* (two papers), *WCCI2020-Glasgow* (one paper), and *Weibo2014* (two papers).

The *Jeong2020* [43] related SD study [444] has the main aim of factorizing the EEG data. Two identical CNNs are used, one to learn class-specific features, and the other as a generator to extract common features through adversarial learning. The features are combined with an MLP. The authors consider session 1 horizontal (forward, backward, left, and right) and vertical (up, down, left, right) classes. After 10-fold CV, the model achieves $54.29 \pm 3.40\%$ for the horizontal classes and $57.29 \pm 5.30\%$ for the vertical ones.

EEGSym [54], already described in Section 4.13, has been used on the *MBT-42* dataset [29]. Considering a cross-subject TL paradigm, EEGSym obtains a $87.40 \pm 8.00\%$ accuracy. *Zhu, Forenzo, and He* [445] test multiple deep classification models, i.e., EEGNet, Shallow/Deep ConvNet, a multi-branch 3D CNN, and a parallel self-attention network on both *MBT-42* and *MED-62* [28] datasets. The first two sessions of both datasets are considered for training (20% validation), and the third session is used for testing in the SD paradigm. The SI approach is applied only on *MED-62* with a LOSO validation (subject 62 used for testing). The best average accuracy values are obtained by EEGNet for all paradigms and they are reported as around 73% on *MBT-42* in the SD paradigm, while they are reported as about 75% and 79% for the SD and SI paradigms applied on *MED-62*, respectively. *MED-62* is also employed by *Autthasan et al.* [64], who use their MIN2Net model (see Section 4.8.1) and obtain $65.90 \pm 16.50\%$ accuracy for the SD paradigm (stratified 5-fold CV), and $59.79 \pm 13.72\%$ for the SI one (LOSO-CV). Finally, *Chen et al.* [296] use an SD approach and obtain 65.70% accuracy (70/30 training/validation split) by applying their proposed filter-bank spatial filtering and temporal-spatial CNN.

Considering the *MIDistraction* dataset [45], *Cai et al.* [446] propose a graph sequence NN, based on self-attention graph convolutional networks and adversarial training. The graph constructed by node topological features of EEG signals and channels is used as the input to the network. Using 6-fold CV, the proposed model achieves a 79.56% average accuracy on the SD paradigm.

For *Nikki2021* [27], *Tibrewal, Leeuwis, and Alimardani* [447] feed the raw EEG signals directly to a CNN, composed by 2D convolution, pooling, flatten, fully connected, and softmax layers, and exploiting Adam optimizer. The data are divided into training (80%) and test (20%) sets, on which $69.42 \pm 4.97\%$ accuracy is obtained on average, considering a global paradigm.

The end-to-end DL CNN proposed by *Yang et al.* [82], reported in Table 5 as well as Sections 4.1 and 4.13, is applied to the *SameLimb* dataset [46]. The accuracy achieved using an SD paradigm is equal to 73.90%. Another decoding model for this dataset is proposed by *He et al.* [218], who developed *S-CAMLP-Net*, described in Section 4.8.2. Considering an SD paradigm and 5-fold CV, the authors obtain $78.45 \pm 0.64\%$ accuracy.

Considering the *UpperLimb* dataset [47], *Zhou, Zou, and Huang* [448] propose a wavelet neural network, where EEG signals are decomposed into sub-bands through means of wavelet packet decomposition. Afterwards, statistical features are extracted from the sub-bands and selected by principal component analysis. 10-fold CV was applied and different hyperparameter functions were considered. Notice that only the first two training sessions for each subject are used and the achieved best mean SD accuracy (with Mexican hat wavelet functions) is equal to $85.24 \pm 7.01\%$. *Wang et al.* [196] strategy, described in Section 4.8.2, consider two classification tasks: (i) a multi-class task of elbow flexion/extension, forearm supination/pronation, and hand open/close of right upper limb, and (ii) a binary task of MI vs. rest on the *UpperLimb* dataset. The authors use a variational sample-LSTM to perform multi-band decomposition and spectral discriminative analysis for MI classification. Their pipeline follows these steps: (i) channel reduction through a channel fusion operator, (ii) obtaining six band-limited intrinsic mode functions through variational mode decomposition, (iii) discriminative frequency bands selection considering the maximum sample entropy value, and (iv) classification through a LSTM model. The authors ultimately found that their model provides an efficient selection of frequency bands, and obtains a 96.60% accuracy on the binary classification task, and 76.20% for the multi-class one in the SD paradigm. The authors apply a 10-fold CV.

The only study [70] using *WCCI2020* [48] proposes the application of standard methodologies, but provides open-source codes (Table 4). The best average accuracy obtained is equal to 77% with LOSO-CV in the SI paradigm, considering the use of EEGNet in combination with different time–frequency or spatial features. The methodology used by *Xu et al.* [71] (see Section 4.10) employs the last dataset presented in this section, i.e., *Weibo2014* [49]. The best accuracy using EEGNet with Riemannian alignment and adaptive batch normalization in the within-subject (SD) scenario is around 78%. Instead, for the cross-dataset scenario, the best accuracy (74.70%) is achieved by using *Cho2017* as the source dataset, ShallowFBCSPNet with Euclidean alignment, and adaptive batch normalization. Finally, another approach applied to *Weibo2014* presents a brute-force CNN model without considering a feature extraction step and using an SD paradigm [292]. The model obtains 74.75% accuracy.

5. Discussion

This systematic review provided an overview of 394 screened studies published from January 1, 2017 to January 23, 2023 in EEG-based MI decoding using deep learning. The present section is divided into six subsections and is meant to provide a summarized description of the main findings, the implications of the reviewed methodologies, the current research gaps, shortcomings, and limitations, the significance of this study as well as some future directions in the DL-based MI decoding field.

5.1. Main findings

In this section, the main research questions that have been posed in Section 1 are discussed. Considering the first research question,

RQ1: *What are the most frequently used publicly available datasets for MI-EEG decoding, specifically DL-based decoding?*

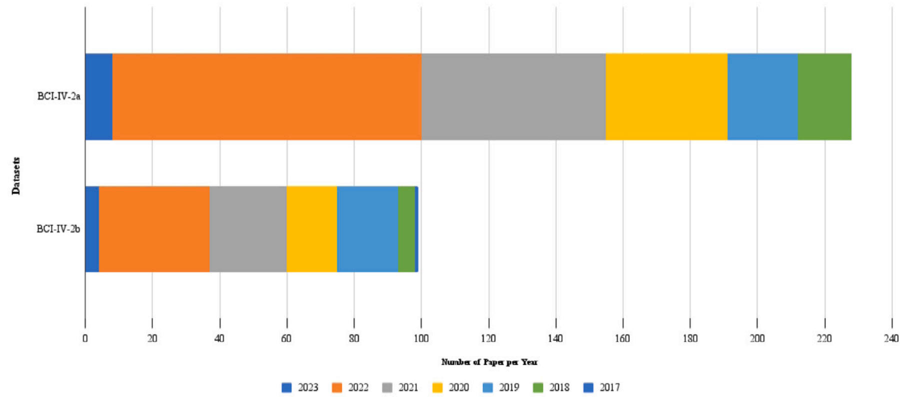
It is immediately observable from Fig. 5, and also from the length of the corresponding sections (Sections 4.4, 4.8, 4.9, and 4.11) that the most frequently used publicly available datasets for DL-based MI-EEG decoding are *BCI-IV-2a* and *BCI-IV-2b* with 228 and 99 related publications, respectively, followed by *BCI-III-IVa* and *eegmmidb* with 46 and 43 related studies, respectively.

Notice that these datasets have been recorded and made open-source in the 2000s, with *BCI-III-IVa* as the oldest (2004) and *eegmmidb* (2009) as the newest among the frequently used datasets. During the years, these datasets have become the de facto benchmarks for DL model training and testing, as well as for classical machine learning techniques and denoising strategies. The reasons behind their vast use, besides their time of publication, are manifold. Firstly, they present classical MI conditions, i.e., the movement rehearsal of left/right hand, left/right foot, both hands/feet, and tongue, that can be used for binary or multi-class classification tasks. In fact, while there are a great number of examples of successful MI decoding of large body parts, there are a few number of studies facing the problem of finer MI decoding (e.g., hand grasping, elbow flexion/extension, or forearm supination/pronation) [449,450]. Moreover, most of the reviewed studies proposing proprietary datasets besides using the publicly available ones, present MI experimental paradigms involving left/right hand movement imagination. Therefore, they usually exploit the same set of conditions present in the publicly available datasets for benchmarking purposes. See Table 2 for a summary of the characteristics of the publicly available datasets. Another reason behind the success of these datasets lies on the fact that the *BCI Competitions* have appealed to the research community as they are among the first EEG-based BCI datasets that were made publicly available. Furthermore, *eegmmidb* is one of the first datasets presenting a large number of subjects to analyze a consistent MI paradigm between subjects' recording sessions, and also used a greater number of electrodes (64) compared to previous datasets.

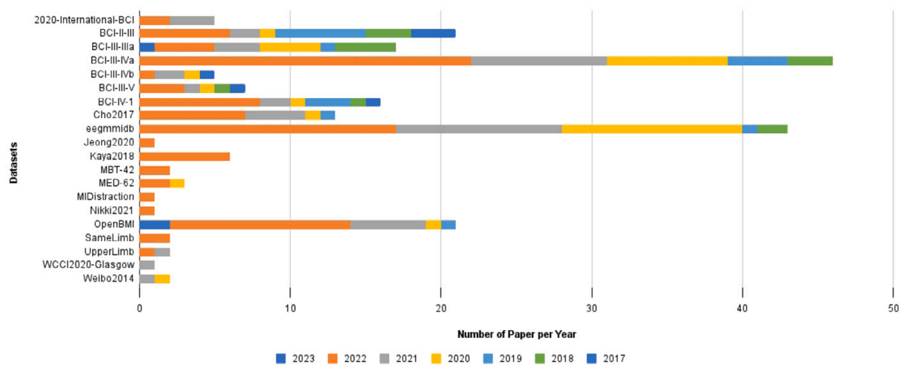
Notice that the *BCI-IV-2a* and *BCI-IV-2b* present also reference electrodes that can be exploited to reduce the ocular artifacts. While these reference data have been exploited in classical signal preprocessing techniques to provide easier-to-interpret signals before applying the classification models, they are also exploitable for new DL-based denoising strategies [451], and thus the trend of using these datasets will probably stay consistent in the future.

However, with the advancements of technologies and learning algorithms, a wider use of more recent datasets with more subjects and trials per subject can be hypothesized. For example, *OpenBMI* (2019) appears in 21 of the reviewed studies and presents some appealing characteristics that can be exploited in different fields from rehabilitation to robot control [24]. Fig. 8 presents a detailed overview of the number of papers employing specific datasets, tracking the years of publication. Notice that the 2023 label refers only to the temporal span between January 1 and 23, 2023 and that *BCI-IV-2a* and *BCI-IV-2b* are reported separately (Fig. 8(a)) to provide a better data readability. An increase in number of papers year by year in the field of DL-based MI-EEG decoding can be observed. An increase in the use of less frequently employed datasets can be also observed, especially in 2021 and 2022, suggesting an increasing interest in new experimental paradigms and recording strategies.

Another interesting observation can be made considering the papers using more than one dataset to train and test their models. Out of 394 papers, around 28% of the studies (112) employ multiple datasets.



(a) *BCI-IV-2a* and *BCI-IV-2b*.



(b) Remaining datasets.

Fig. 8. Distribution per year of the reviewed papers for each of the reported publicly available datasets.

Most of these papers use two (83) or three (25) datasets, while only 4 considers 4 different repositories. As depicted in Fig. 9, this trend involves especially the most recent years (2021–2022). Notice that the 2023 label refers only to the temporal span between January 1 and January 23, 2023, thus providing a promising datum with respect to the use of multiple datasets. Moreover, the selected multiple datasets present coherent MI conditions, containing at least two overlapping tasks (e.g., left/right hand MI) and other conditions such as tongue and feet MI.

Considering the given observations, some guidelines on the choice of datasets are proposed from the authors' perspective.

Firstly, the choice of the dataset should be based on the classification task the experimenters want to perform, i.e., either a binary classification or a multi-class recognition. Moreover, a decision should be made on the type of MI conditions to be analyzed and the dataset chosen accordingly. The tables reported in Section 3 may represent a good starting point to choose the proper datasets. Secondly, the use of multiple datasets should be considered, especially when intending to demonstrate the generalizability of the proposed DL model. Finally, we can notice the number of subjects and trials per subjects present in each dataset. In fact, a greater number of data samples is more appropriate for DL models to avoid overfitting and curse of dimensionality. This aspect is addressed in some studies by considering data augmentation strategies to be applied before executing the learning model. This observation opens the discussion related to the second research question,

RQ2: *What are the current trends, strategies, and architectures used in DL-based MI-EEG decoding?*

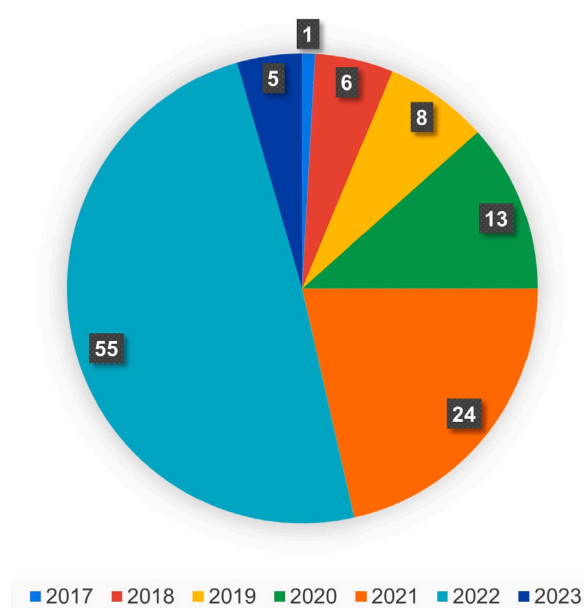


Fig. 9. Number of papers using multiple publicly available datasets to train and test their models, divided by year.

As described in Section 1, RQ2 is not only concerned with commonly applied DL-techniques, but also with the application of specific pre-processing strategies and decoding paradigms. Browsing the different

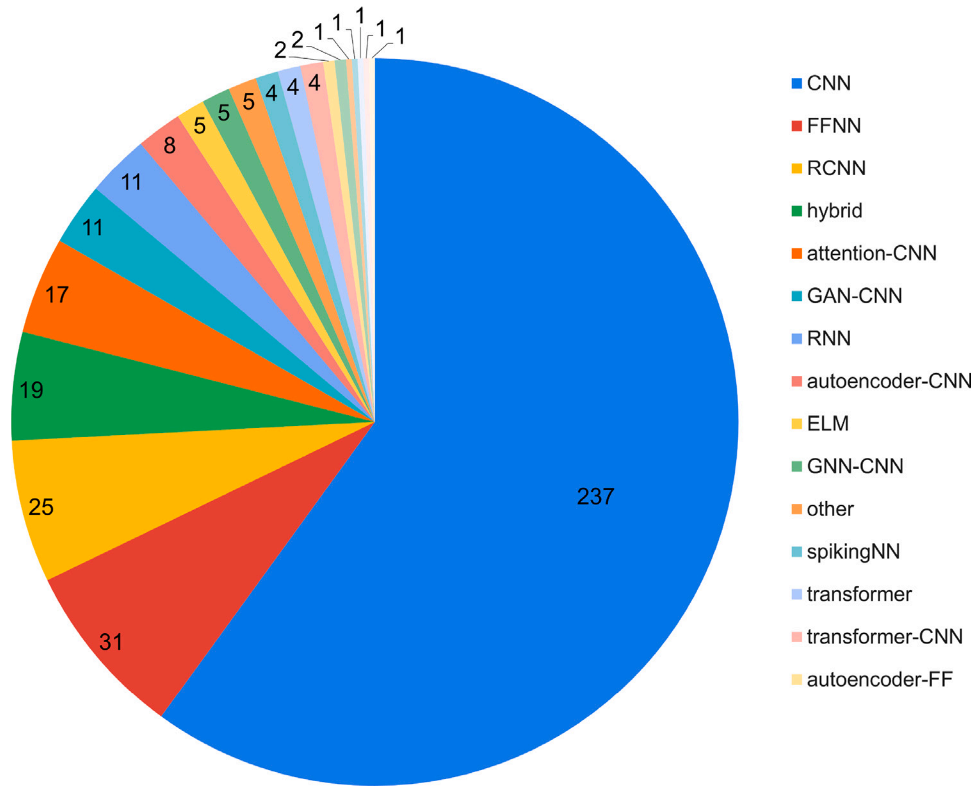


Fig. 10. DL architecture types used in the reviewed papers. Additional acronyms present in the legend: graph neural network (GNN), feed-forward (FF), recurrent convolutional neural network (RCNN), extreme learning machine (ELM). *Hybrid* models combine methods such as evolutionary optimization or Riemannian geometry with DL. *Multiple* means many architectures are evaluated in the study.

techniques presented in Section 4, the vast use of CNN-based architectures is easily observable. CNN-based models can exploit well-known architectures (possibly pre-trained) such as EEGNet, Shallow/Deep ConvNet, InceptionV3, ResNet, and VGGNet. Fig. 10 depicts the DL models used in the reviewed papers.

Moreover, statistics have been extracted to show the trend regarding the use of different architectures over the years and also for each open-source MI dataset. These statistics are visualized in Figs. 11 and 12 by considering the entries from 2017 to 2023, and providing visual references only for the datasets presenting more than 10 related papers to provide a better readability of the figure, respectively.

Although CNN remains the most prevalent type of architecture used over the years (Fig. 11(a)), a greater number of diverse approaches have been employed as a consequence of the advancements of deep learning methodologies and of emerging new architectures (e.g., based on attention [452] or transformers [453]).

This tendency to use emergent architectures can be observed in Fig. 12(b) as hybrid methods and the ones employing attention and transformers have reached the popularity of recurrent NNs and feed-forward ANNs for the most used open-source datasets, i.e., *BCI-IV-2a*, *BCI-IV-2b*, and *eegmmdib*.

Considering the different CNN models proposed by the literature works, more than modifying the architectures, the authors propose different input formats. This usually involves t[a-z]he application of techniques that allow the generation of time–frequency inputs, e.g., Morlet, continuous wavelet or fast Fourier transform, considering one or more electrodes at a time, combined in a single image or not. Mostly, CSP or some of its variations are employed, when proposing the use of handcrafted features.

As previously mentioned, another interesting trend is represented by the use of attention mechanisms, which appear in more recent papers published between 2020 and 2023. Attention is proving to be a very important concept in deep learning, having that it can improve the extraction of relevant and semantic information present in the data [454]. In the EEG domain, researchers have shown that time, frequency, or spatial information can be leveraged differently depending on the applied attention mechanism and on the location of the attention layer in the considered model (CNN or LSTM) [455]. The promising efficacy of attention mechanisms is also reported by the reviewed papers, obtaining average accuracies above 80%, especially on *BCI-IV-2a* with an SD configuration.

Considering the decoding paradigms (Fig. 6), SD is applied in around 75% of the reviewed studies, while SI in 15% of them. Both paradigms are considered by 6% of the reviewed papers, while global results are reported by around 4.5% of them. Finally, transfer learning strategies appear in around 6% of the reported studies. Therefore, the SD configuration remains the most used even in DL-based research. While the application of such paradigm may result in a more personalized codification of the EEG, thus providing a subject-centered BCI, the use of SI techniques may provide weaker performances, but more generalizable models able to learn common patterns between subjects. Surely, the TL paradigm can also represent a good starting point to provide more reliable models that can leverage both data coming from the same subject but different experimental sessions and from different subjects. TL application may boost DL-based analyses, especially due to the fact that it is able to manage small-scale data, while maintaining the learning ability with different individuals [456].

The problem of small datasets is also considered by authors that do not rely on TL mechanisms, but prefer the application of data augmentation. Of the reviewed papers, 25 clearly report the application of such strategies and prefer the use of techniques based on a sliding-window approach. Some generative methods based on GANs are also considered, but no information on the possible bias introduced when augmenting the data is reported in many studies.

Another important topic is related to the reproducibility. While the present review paper is focusing on studies using at least one

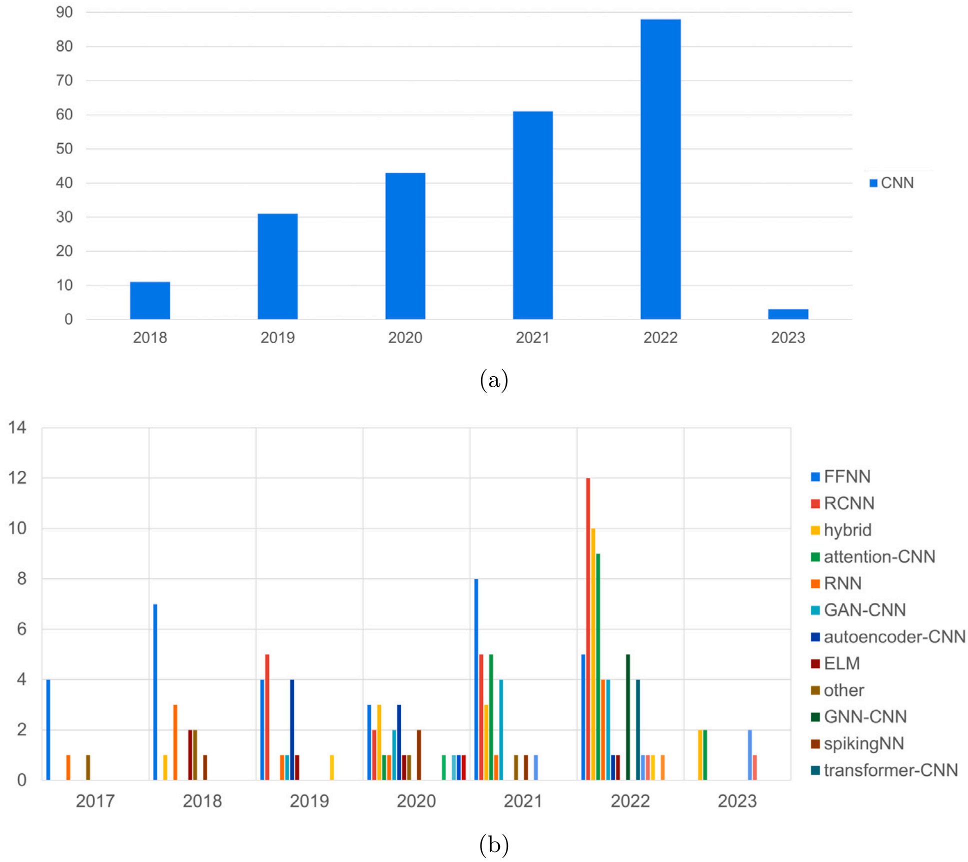


Fig. 11. Number of different DL architectures used per year for EEG MI decoding. The chart is divided in two parts to have a better figure readability. The acronyms are compliant with the legend of Fig. 10.

publicly available dataset and describing their architecture visually and/or textually in a sufficiently precise manner, no restrictions have been made considering code availability. As reported in Section 4, out of 394 reviewed studies only 30 share publicly the codes of the devised model and used DL strategies. Notice that three of these studies landed to unavailable web pages and presented codes in the pre-print version that were not available in the peer-reviewed one, and thus were excluded from this review. Fortunately, the resources reported in Tables 4 and 5 point to very different models, spanning from EEGNet-based architectures to LSTM models, from NNs based on temporally adaptive CSP to graph-based neural networks, and from transfer learning methods to data augmentation techniques. Therefore, these models can be easily used as benchmarks for new DL approaches, which in turn should present enough information to allow their reproducibility.

Summarizing these observations, the majority of the studies present CNN-based architectures, providing a larger model pool for performance comparison with respect to other DL techniques. Other models may be considered in the future, especially to learn more representative information. Further hybrid networks may be devised to exploit different information domains (e.g., spatial, temporal, and frequency domains) as well as new attention mechanisms. Afterwards, one or more decoding paradigms can be considered. When intending to provide a more personalized BCI, an SD paradigm may represent the best solution, while the SI and global paradigms are particularly adept at learning common patterns between subjects. TL can be an in-between solution, having that it can exploit data from a single subject or from multiple subjects to avoid the problems deriving from small data dimensionality.

Furthermore, the reproducibility of studies is of fundamental importance as also identified by other reviews such as *Roy et al.* [18]. Besides the use of publicly available datasets, the papers should provide enough details on the architectures to be reproduced when the related codes are not made publicly available. Moreover, the results should always be compared with other baselines and methodologies. We will discuss more about reproducibility in the following sub-sections.

Before moving to the next research questions, the summary Tables 6 and 7 are provided as a quick reference to access the best performing DL methods in terms of decoding performance for each dataset and decoding paradigm. Note that due to the different methods of evaluations employed by different studies and a lack of explanation regarding evaluation in many articles, a fair, balanced, and statistically meaningful comparison between methods for each dataset/paradigm pair is not possible. Therefore, the most significant methods for each pair are mentioned by considering not only their reported performance, but also taking into account the level of reproducibility and novelty of the method.

Notice that the terms convolution/convolutional, F1-score, and information are abbreviated to "conv"., "F1", and "info".

Focusing on the next research question

RQ3: *What are the DL architectures most frequently used as baseline models or as inspirations for new decoding methodologies?*

some observations can be made.

As introduced for RQ2, a large number of the reviewed studies develop DL architectures based on previous state-of-the-art models. In particular, this review paper starts its analysis from 2017 to detect studies based on the pioneering works of *Lawhern et al.* [38] and *Schirrmeister et al.* [37].

In fact, *Lawhern et al.*'s main aim was to propose a compact CNN that could be used for different BCI-related experimental paradigms, while being able to visualize the learned features. This brought to

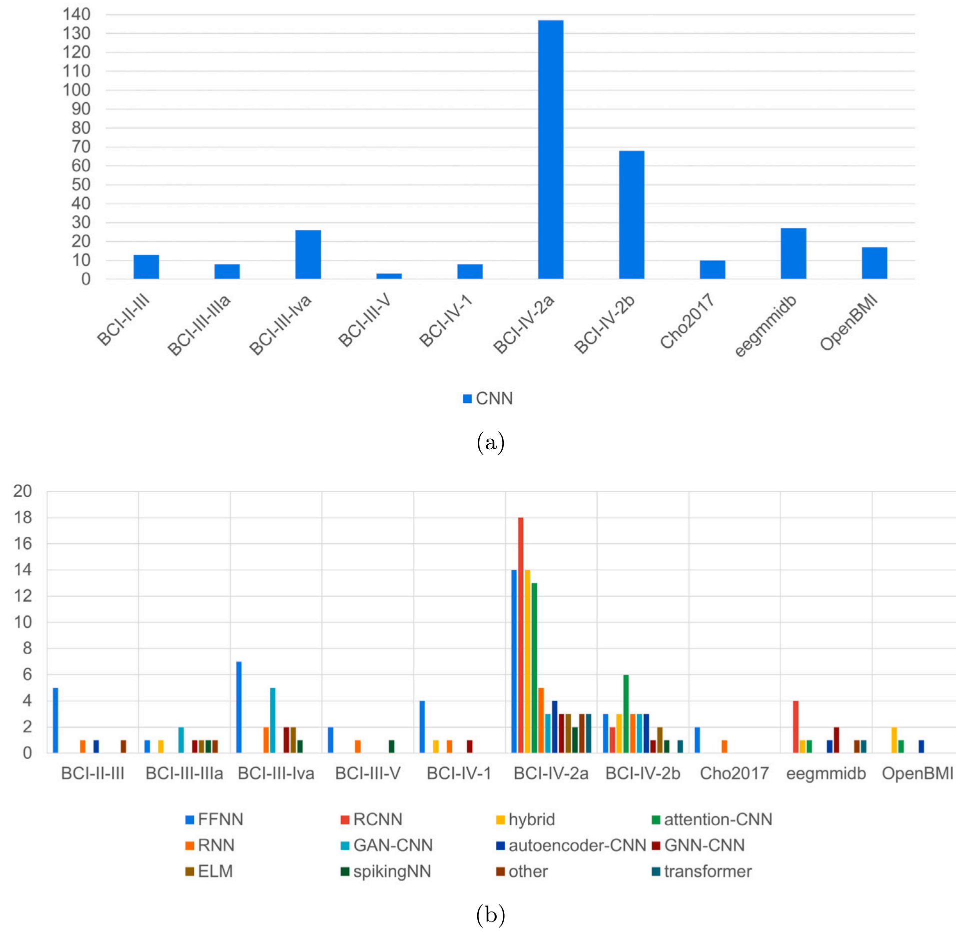


Fig. 12. Number of different DL architectures used for EEG MI decoding in the most popular datasets. The chart is divided in two parts to have a better figure readability. The acronyms are compliant with the legend of Fig. 10.

the development of the widely used *EEGNet*. Soon after the first appearance of this model on arXiv (2016, journal publication on 2018), *Schirrmeister et al.* proposed the Shallow and Deep ConvNet models. The novelty introduced by the authors was not only related to learned feature visualization, but also by the fact that the authors wanted to better understand how a CNN should be designed to exploit raw EEG signals. Moreover, they achieved better performance results with respect to the classical FBCSP-based techniques, that represented the de facto baselines for EEG-based BCI decoding.

The trend of using these models as baselines for DL-based research has not become outdated yet. In fact, searching on *Scopus* for these two publications (accessed on August 28, 2023) and exporting the number of citations per year, it can be noticed (Fig. 13) how it constantly increases along time.

The authors of the present review paper suggest that a DL-based MI BCI project should consider these benchmarks and consider the publicly available codes listed in Tables 4 and 5 to provide clear comparisons with such reliable and reproducible models.

Regarding reproducibility, we can turn to the final research question,

RQ4: *How should methodologies and results be reported in a DL MI decoding study to ensure the reproducibility of results?*

Diverse observations can be made on the reviewed papers. Numerous reviewed studies have missing information on the validation process and/or data splitting into training, test and validation sets. Moreover, the reported performance measures are usually limited to accuracy only. Cohen's kappa is present in a few papers, especially the ones referring to the *BCI Competitions* datasets. There is also a lack of consistency when reporting the performance measures, having that accuracy/κ values appear in percentages or not, and with different decimal precision. In addition, some subject-dependent studies provide only the obtained average accuracy, instead of presenting the results for each subject. Standard deviation is usually not specified when reporting average results. Given these observations, the authors would like to point out that researchers should make more effort in terms of sharing code and reporting results to ensure reproducibility of their research. Reports should consider further evaluation metrics besides accuracy only, having that this performance measure provides information only on correct and incorrect predictions and it is not reliable in presence of unbalanced datasets in terms of classes. Therefore, precision, recall, specificity, and F1-score [457] may be also considered to provide a better representation of the obtained results. Moreover, standard deviation should be reported when providing an average performance to better understand the reliability of the model with respect to performance variability.

Another issue is represented by the lack of comments on different bias types, e.g., algorithmic bias or bias due to data augmentation [458], and on computational costs. Considering that DL models can require a high inference time and that BCI systems are devised to work in real-time environments, the missing information on computational costs may represent a limit of the proposed studies. Therefore, the authors of the present review suggest that researchers should provide appropriate information to allow a clear assessment on data quality, performance reliability and efficacy of the proposed DL models in the context of MI-EEG BCI-based systems. A set of guidelines on this topic is presented in Section 5.5.

Table 6
Best performing decoding methods for different datasets and paradigms, considering the level of reproducibility and novelty of methods (**Part 1**).

Dataset	SD	SI	TL	Global
2020-International-BCI	Temporal ConvNet combined with EEGNet, inter-session avg. acc.: 56.73% [82]	Temporal ConvNet combined with EEGNet, avg. acc.: 55.24% [82]	–	–
BCI-II-III	time–frequency extracted features fed to an FFNN, avg. acc: 91.96% [94]; Temporal spatial ConvNet followed by stacked autoencoders, avg. acc.: 90.60% [103]	–	–	–
BCI-III-IIIa	Spatial-frequency-temporal features to a 3D ConvNet [116], avg. acc.: 91.85%; Class-conditional GAN based on Deep ConvNet, avg. acc.: 91.85 ± 9% [78]	Extract features from an autoencoder latent space and train a multi-class SVM, avg. acc.: 95.33% [111]	–	–
BCI-III-IVa	Class-conditional GAN based on Shallow ConvNet, avg. acc.: 77 ± 9% [78]; Constant-Q transform features to a ConvNet, avg. acc.: 76.80 ± 4.80% [137]	Extract features from an autoencoder latent space and train a random forest, avg. acc.: 99.92 ± 0.08% [136]; Energy and entropy features fed to an MLP, avg. acc.: 97 ± 2.70% [135]	Conditional DCGAN for adaptive TL and fine-tuning, avg. acc.: 71.14% [127]; EEGNet-based domain adaptation TL, avg. acc.: 82.61% [128]	Constant-Q transform features to a ConvNet, avg. acc.: 77.30 ± 1.60% [137]
BCI-III-IVb	Fourier decomposition before an FFNN, avg. acc.: 93.33% [94]; EWT features fed to an MLP, avg. acc.: 100% [132]	EWT features fed to an MLP, avg. acc.: 92.90%, sensitivity: 93%, specificity: 93%, F1: 96.40% [132]	–	–
BCI-III-V	time–frequency features extracted via successive decomposition index and fed to an FFNN, avg. acc.: 99.33% [160]	–	–	–
BCI-IV-1	Parallel CNN branches fed by FBCSP features, avg. acc.: 84.50%, κ: 0.6900 [175]	–	Target subject adapted CNN, avg. acc.: 79.35 ± 12.01% [130]	–
BCI-IV-2a	ConvNet followed by a multi-head attention block and a temporal conv. block, avg. acc.: 85.38% [73]; Multi-channel attention applied to time–frequency features followed by conv. layers, avg. acc.: 81.48%, κ: 0.7530 [214]	MIN2Net (end-to-end training): latent vectors of the autoencoder used for deep metric learning and to train a linear classifier, avg. acc.: 60.03 ± 9.24%, F1: 49.09 ± 23.28% [64]; ConvNet followed by a multi-head attention block and a temporal conv. block, avg. acc.: 70.97% [73]	ConvNet-based domain adaptation via an optimal transport solver and mixup data augmentation, avg. acc.: 60.69% [55]	Twin-cascaded softmax ConvNet, avg. acc. 80.03% [354]; Shallow ConvNet followed by a BiLSTM and an attention layer, avg. acc.: 82.70 ± 5.57% [352]
BCI-IV-2b	Self-attention applied to extracted features and fed to a ConvNet for classification, avg. acc.: 87.04 ± 11.95% [265]; A graph for each signal based on mutual info. and distance between electrodes and classes before ego-conv. layers to extract embeddings for classification, avg. acc.: 96.50% [141]	Tokenize spatio-temporal features using conv. to feed a transformer encoder and finally use a classification head, avg. acc.: 88.39% [340]	Augment BCI illiterate users' data using a conv. autoencoder combined with attention and comparing with a "golden subject"'s data and then use a ConvNet for classification, avg. acc.: 80.70 ± 6.10% [398]	1D conv. to extract features from filter banks followed by an attention layer. A relation score is calculated to determine the label with each class having a few samples as support and the query data, avg. acc.: 74.60 ± 10.20% [353]

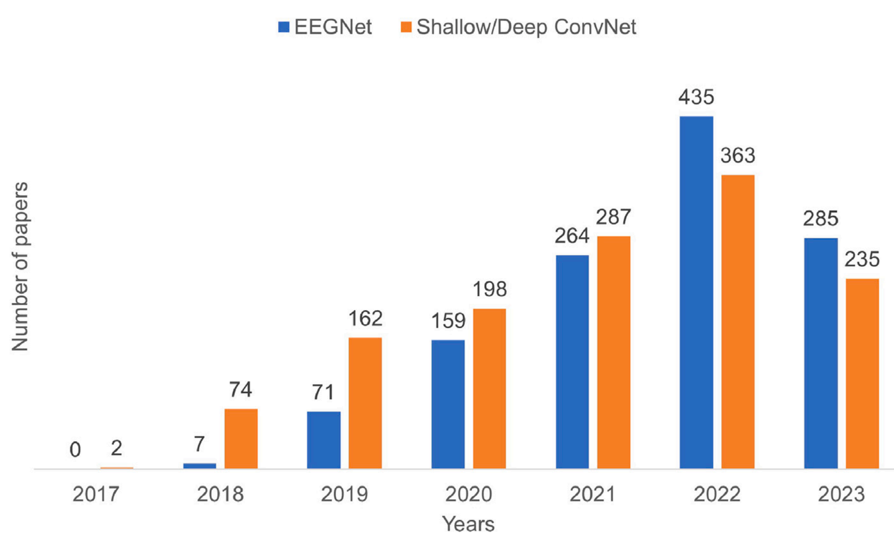


Fig. 13. Number of papers citing the identified benchmarking architectures EEGNet and Shallow/Deep ConvNet from their first journal publication, i.e., 2018 and 2017, respectively. Self-citations are included. Statistics taken from *Scopus*, accessed on August 28, 2023.

5.2. Practical and theoretical implications of decoding methodologies

Section 5.1 presents the main findings of the review study in the form of responses to the identified research questions. The provided responses contain observations, implications and guidelines on different topics characterizing the investigation made on DL-based MI-EEG decoding.

Considering the large number of reviewed studies and the inherent similarities and differences among the proposed methodologies reported in Section 4, an overview of the common practical and theoretical implications of the analyzed methods is provided below.

Generally, the use of different datasets to train and test a model provides a better understanding of the in and out-of-distribution generalization of a DL model and its ability to adapt to a real-life scenario. This surely implicates the necessity of having a clearer understanding of the EEG signals used to train and test the model, the MI conditions, and the time required by the DL approach for computation. The latter issue depends not only on the data dimensionality but also on the available technologies, on the complexity of the DL architecture, and on the applied decoding strategy. While the use of SI and global approaches allow for better generalization, depending on the dataset, the training cost and computation is higher compared to SD

Table 7
Best performing decoding methods for different datasets and paradigms, considering the level of reproducibility and novelty of methods (**Part 2**).

Dataset	SD	SI	TL	Global
Cho2017	Channel selection based on performance using masked channeling and a compact ConvNet for classification, avg. acc.: 70.80% [286]	Estimate the mutual info. between feature representations by decomposing them to class-relevant and irrelevant ones and EEGNet for classification, avg. acc.: 73.73 ± 13.75% [400]	Subject domain adaptation with CWT topograms to a ConvNet and MLP for classification, avg. acc.: 82.60 ± 8.40% [65]	–
eegmmidb	EEGNet with Euclidean alignment, avg. acc.: 74% [71]	ConvNet followed by transformer encoder, avg. acc.: 64.22% [421]	Reptile meta learning with EEGNet, avg. acc.: 68.49 ± 13.91% (left/right hand), 68.60 ± 14.36% (fists/feet) [419]	EEG2Image based on denoised-ConvNets to learn the colorspace and spatial variations of the images obtained considering only the alpha and beta rhythms, avg. acc.: 99.06% [404]
Jeong2020	Two CNN branches to extract class-specific and common features with adversarial learning [444]	–	–	–
Kaya2018	*EEGSym* symmetric input to an inception-style network, avg. acc. $85.10 \pm 9.50\%$ [54]	–	Bi-GRU, avg. acc. (cross-sub TL) $80.01 \pm 10.00\%$ and Deep ConvNet, avg. acc. (within-sub TL): $76.17 \pm 14.62\%$ [52]	–
MBT-42	Vanilla EEGNet, avg. acc.: 73% [445]	–	*EEGSym*, avg. acc.: 87.40 ± 8% [54]	–
MED-62	MIN2Net, avg. acc.: 65.90 ± 16.50%, F1: 64.13 ± 17.66% [64]	MIN2Net, avg. acc.: 59.79 ± 13.72%, F1: 61.10 ± 23.64% [64]	–	–
MIDistraction	Graph NN with conv., self-attention, and node domain adversarial training, avg. acc.: 79.56% [446]	–	–	–
Nikki2021	–	–	–	End-to-end 2D ConvNet with final fully connected layer, avg. acc.: 69.42 ± 4.97% [447]
OpenBMI	Tensor-CSPNet, avg. acc.: 75.92 ± 14.63% [221]	Compact framework combining temporal conv. and CSP for sequential extraction of spatio-temporal features, avg. acc.: 74.28 ± 16.12% [84]	*EEGSym*, avg. acc. $83.30 \pm 9.30\%$ [54]	–
SameLimb	1D-CNN encoder features fed to LSTM for self-supervised next slice prediction. Pre-trained encoder features used to train a supervised channel-attention MLP-Mixer for classification, avg. acc.: 78.45 ± 0.64%, F1: 78.39 ± 0.67% [218]	–	–	–
UpperLimb	Variational mode decomposition samples fed to an LSTM, avg. acc. (multi-class): 76.20% [196]	–	–	–
WCCI2020-Glasgow	Multiple time–frequency methods to extract features and classify using EEGNet, avg. acc.: 77.00% [70]	–	–	–
Weibo2014	End-to-end 3D ConvNet, avg. acc.: 74.75% [292]	–	–	–

strategies. On the other hand, the application of generally faster SD strategies can sometimes result in overfitting of the model and biased classification results towards specific subjects with high-quality EEG recordings.

Another important implication is related to the input of DL models. Numerous works leverage features derived from CSP-based algorithms or time–frequency images which could again increase the computation time. However, using these methods, helpful patterns can be learned to better characterize the intended MI condition.

Finally, the general comparisons with benchmark models and their exploitation to develop novel algorithms provides a better understanding of the results of each approach. Furthermore, code sharing allows reproducibility of experiments and the future development of methodologies that could be refined and used in a real-time setting.

5.3. Limitations of the reviewed studies

Considering the observations reported to answer the devised research questions and the detailed descriptions provided in Section 4, this subsection aims to report the main limitations found in the reviewed studies in a compact manner.

Starting from the most frequently used datasets, while it is possible to compare different studies working with *BCI-IV-2a* and *BCI-IV-2b* datasets, it is important to point out that these datasets have a major drawback for DL decoding, i.e., they have a small sample size.

Considering *BCI-IV-2a* and *BCI-IV-2b*, as reported in Table 2, both datasets present the recordings from only 9 subjects. Moreover, consulting the dataset descriptions in Table 1 it can be noticed that for each session and subject in *BCI-IV-2a*, only 288 trials are present for the four MI conditions, i.e., left/right hand, tongue, feet MI. For *BCI-IV-2b*, the training data comprised 240 and 160 balanced trials for each MI condition, i.e., left/right hand MI, with and without visual feedback, respectively, while the evaluation data presented 320 balanced trials.

The limited number of data to be used for EEG-MI DL-decoding is also true for the other identified most frequently used datasets, i.e., *BCI-III-IVa* and *eegmmidb*. In *BCI-III-IVa*, only five healthy subjects were involved in the experiments, and they were asked to perform the MI of left/right hand and right foot. Notice that the cue data were provided only for the right hand and foot conditions. Moreover, the experimental conditions varied for two subjects as well as the number of executed trials. In fact, 112 ± 81.63 (mean ± standard deviation) training and 168 ± 81.63 testing trials were provided by the experimenters. Concerning *eegmmidb*, while a larger number of subjects (109) is present, a limited number of repetitions (45) is performed for left/right hand or both fists/feet MI tasks.

In general, most of the reported datasets are recorded on a maximum of 10 participants and no more than 50 trials are executed per each MI task.

The absence of a sufficiently large data pool is particularly problematic for DL works employing an SD paradigm which considers only an individual subjects' data. In general, DL methods require a larger number of samples compared to traditional ML to achieve a competitive performance or outperform ML methods. To remedy this issue, data augmentation techniques can be used as in many studies discussed in Section 4. However, data augmentation can introduce biases if not performed with care. As seen in Section 4, this problem is almost never discussed in the reviewed papers. In fact, many studies do not even report the complete procedure of the proposed data augmentation. Moreover, as reported by *Song et al.* [81], some of the data augmentation methodologies relying on Gaussian noise or cropping can have a negative impact on the data quality, further decreasing the signal-to-noise ratio or the coherence of the original dataset.

Therefore, not providing the specifics on data quality and quantity is lacking in many studies in the literature, even though the reliability and the efficacy of DL models are strictly dependent on them.

In addition to data augmentation, many studies also lack information on the exact procedure of data pre-processing that was used in their study. Many works consider the use of raw EEG data or simply bandpass-filter signals to retain only the frequency bands of interest. However, in many cases, there is no information on how the signal is managed prior to classification.

Another important information that is usually lacking in the description of different DL-based EEG decoding models is their computational cost and efficiency. In fact, BCI systems are ultimately intended for real-time use, and therefore, they should have a low inference time in test time to be deployed in the real world. Unfortunately, only a few papers report the compute cost and/or inference time required by their models.

5.4. Future perspectives and developments

Future perspectives and developments in the EEG-MI DL decoding for BCI application research can be manifold. Considering data availability, researchers are beginning to consider larger datasets to validate their DL models. Particular attention should be given in the future to some of the datasets identified in this review paper such as (i) *Cho2017* collected on 52 subjects, during five to six runs of 100 to 120 trials per each MI condition, i.e., left and right hand MI; (ii) *Nikki 2021* presenting the recording of 55 subjects during a left/right hand MI experiment consisting of one session, four runs (one calibration and three feedback runs) of 40 trials balanced between the MI conditions; and (iii) *OpenBMI* collecting left and right hand grasping on 54 subjects during two sessions of 100 trials each.

Moreover, further attention should be given on how new datasets are collected, especially considering that MI-BCIs are usually restricted to laboratory environments and are extremely bounded to the individual ability of their users, which may require numerous training sessions [459]. This topic involves the necessity of moving away from subject-specificity, trying to resolve the non-easy problem of BCI illiteracy, and moving from controlled to real-life environments.

Surely, the direction taken by novel DL approaches is to decrease the number of training sessions, and thus user fatigue, while improving the classification performance. Especially during the model training phase, new methodologies for data augmentation will probably arise as the use of GANs and diffusion models are more frequently used to augment biomedical signal datasets. Furthermore, we predict that more studies will focus on the use of subject-independent and transfer learning decoding paradigms to provide a better model generalization and exploit dataset and subject transfer. Considering subject transfer, BCI-literate subjects [24] will surely be identified and their data will be exploited to enhance the model performance of BCI-illiterate users [398].

Moreover, novel methods are beginning to exploit attention mechanisms and transformers [453], which should be able to extract more relevant spatio-temporal features from EEG data that represent the neural patterns in time, frequency, and spatial domains.

In Section 4, some works proposing graph convolutional neural networks have been reviewed. New developments of these architectures can also be hypothesized, having that graph CNNs are able to capture structural relations between data [460] and may prove effective in enhancing the classification performance [461] as well as in improving EEG channel selection [462] and thus brain topology modeling [463].

Future developments can be also hypothesized regarding the use of hybrid DL architectures that are able to consider the aforementioned techniques. However, these new possible developments should be thought of with reproducibility in mind as well as model explainability [464,465] to achieve a clearer understanding of the proposed models, the features they learn, and their outcome.

Other future development can also go towards managing and fixing non-ideal EEG data [466] instead of solutions for real-life BCI applications. For example, *Chu et al.* [466] propose a solution to detect the noise-affected or missing EEG signal segments, to build clearer EEG datasets.

As a final remark, future studies working with BCIs and neurotechnologies cannot ignore the importance of ethical and legal issues arising from their use [467]. While usually the main requirement for performing and reporting BCI-related experiments is to comply with the 1964 declaration of Helsinki [468], let the participants read and sign an informed consent, and (recently) ask for the approval of an Ethical Committee, studies rarely report clear information on such topics.

In the future, researchers will have to face with more care the concepts of *neuroethics* and *neurosecurity* [469,470]. This will involve a further understanding of physical, psychological, and social factors related to these systems such as BCI users' safety, humanity and personhood, and perceived autonomy. Moreover, ethical aspects, responsibility and regulations [469] will have to be covered by multidisciplinary teams when preparing BCI-related projects.

Technical issues should be also dealt with, especially when considering the diffusion of consumer-grade devices, concerning the security of neurotechnology-based systems. Security measures should be taken to avoid BCI input and measurement manipulation, changes to the decoding activity of the system, and the possibly unexpected feedback to the users at the hand of a malicious party [470].

5.5. Guidelines for reproducibility

Throughout this review, the importance of reproducibility in research has been emphasized. Therefore, in this sub-section, guidelines and best practices are provided for sharing code, data, and experimental protocols to facilitate reproducibility in the area of DL-based MI EEG decoding. Specifically, the authors recommend the following practices for code sharing:

- Use version control systems (e.g., Git) and host code repositories on open platforms such as GitHub.
- Include clear documentation and README files explaining how to set up and run the code.
- Specify all dependencies required to run the code and their versions in a requirements file.
- Consider containerization (e.g., Docker) to ensure consistent environments across different systems.

This review has focused on publicly available MI EEG datasets. In addition to the availability of the raw EEG dataset alongside its experimental protocols, if a complex data pre-processing pipeline is used on the data, researchers should include data preprocessing scripts and report any data cleaning or transformation steps.

Finally, considering result reporting and analysis, the following best practices are encouraged:

- Report average performance and standard deviation to ensure statistical significance.
- Include scripts or notebooks used for data analysis and visualization.
- Clearly state any statistical methods used for post-processing and analysis and justify their selection.
- Clearly state any limitations or potential sources of bias in the study.

5.6. Summary of the practical significance of this review

This review mainly aimed at providing a comprehensive overview of the most popular DL-based decoding techniques used to interpret EEG signals collected while performing different MI conditions, especially in BCI-related studies.

The review can be practically used in different ways:

- Researchers working on one or more of the reported datasets can have quick access to the related literature.
- Researchers can evaluate and apply the best decoding approach according to their necessities and find some examples that could guide the preparation of their own project.
- Researchers considering specific DL architectures, such as CNNs or LSTM models, can find a broad number of papers exploiting these architectures to define their own baselines.
- Researchers may consult the provided guidelines in terms of research sharing, data and input preparation, DL and decoding approach usage and documentation, and presentation of the results.

The significance of this study is then related to knowledge sharing in the shape of a manual covering the latest trends of the field of interest.

6. Conclusions

The use of DL strategies in MI-EEG based BCI systems is increasing thanks to the evolution of BCI-related technologies, to the proliferation of graphical processing units, as well as to the presence of new datasets presenting a larger number of samples compared to the widely used but small *BCI Competition* datasets. The aim of this review paper was to provide a comprehensive overview of the BCI literature, focusing on deep learning approaches for motor imagery EEG signal decoding. From the provided analyses, the vast use of datasets presenting left/right MI experimental paradigms for binary classification tasks has been detected. However, new MI experimental paradigms have risen in the last few years and research is being done to provide finer and multi-functional BCI systems.

Moreover, four main motor imagery decoding paradigms have been identified and the reviewed studies related to each publicly available MI-EEG datasets are reported accordingly. The subject-dependent decoding paradigm is widely used but a shift to the other three paradigms which are driven by the aim of increasing the generalizablity of the decoding models is noticed, especially when studies aim to develop or pave the path for real-world and reliable BCI systems. The trend in DL-based MI-EEG decoding shows that in the future, emerging DL strategies that rely on transfer learning and subject-independent decoding will be more prevalent. Finally, DL trends concerning novel architectures and baseline models used for performance comparison were identified with an emphasis on reproducible approaches in MI-EEG decoding.

Funding

The authors did not receive support from any organization for the submitted work.

CRediT authorship contribution statement

Aurora Saibene: Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Hafez Ghaemi:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Conceptualization. **Eda Dagdevir:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] J.R. Wolpaw, Brain–Computer Interfaces, vol. 110, Elsevier, 2013, pp. 67–74.
[2] S. Saha, K.A. Mamun, K. Ahmed, R. Mostafa, G.R. Naik, S. Darvishi, A.H. Khandoker, M. Baumert, Progress in brain computer interface: Challenges and opportunities, Front. Syst. Neurosci. 15 (2021) 578875.
[3] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, BCI2000: a general-purpose brain-computer interface (BCI) system, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1034–1043.
[4] A. Saibene, M. Caglioni, S. Corchs, F. Gasparini, EEG-based BCIs on motor imagery paradigm using wearable technologies: A systematic review, Sensors 23 (5) (2023) 2798.
[5] H. Altaheri, G. Muhammad, M. Alsulaiman, S.U. Amin, G.A. Altuwaijri, W. Abdul, M.A. Bencherif, M. Faisal, Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review, Neural Comput. Appl. 35 (20) (2023) 14681–14722.

[6] M. Aldayel, M. Ykhlef, A. Al-Nafjan, Deep learning for EEG-based preference classification in neuromarketing, Appl. Sci. 10 (4) (2020) 1525.
[7] J. Kögel, J.R. Schmid, R.J. Jox, O. Friedrich, Using brain-computer interfaces: A scoping review of studies employing social research methods, BMC Med. Ethics 20 (2019) 1–17.
[8] R. Abiri, S. Borhani, E.W. Sellers, Y. Jiang, X. Zhao, A comprehensive review of EEG-based brain–computer interface paradigms, J. Neural Eng. 16 (1) (2019) 011001.
[9] A. Al-Saegh, S.A. Dawwd, J.M. Abdul-Jabbar, Deep learning for motor imagery EEG-based classification: A review, Biomed. Signal Process. Control 63 (2021) 102172.
[10] Y. Jeon, C.S. Nam, Y.-J. Kim, M.C. Whang, Event-related (de) synchronization (ERD/ERS) during motor imagery tasks: Implications for brain–computer interfaces, Int. J. Ind. Ergon. 41 (5) (2011) 428–436.
[11] D.J. McFarland, L.A. Miner, T.M. Vaughan, J.R. Wolpaw, Mu and beta rhythm topographies during motor imagery and actual movements, Brain Topogr. 12 (2000) 177–186.
[12] A.M. Ladda, F. Lebon, M. Lotze, Using motor imagery practice for improving motor performance–a review, Brain Cognit. 150 (2021) 105705.
[13] F. Malouin, C.L. Richards, Clinical applications of motor imagery in rehabilitation, in: Multisensory Imagery, Springer, 2013, pp. 397–419.
[14] D. Bright, A. Nair, D. Salvekar, S. Bhisikar, EEG-based brain controlled prosthetic arm, in: 2016 Conference on Advances in Signal Processing, CASP, IEEE, 2016, pp. 479–483.
[15] S. Jacob, V.G. Menon, F. Al-Turjman, P. Vinoj, L. Mostarda, Artificial muscle intelligence system with deep learning for post-stroke assistance and rehabilitation, IEEE Access 7 (2019) 133463–133473.
[16] N. Padfield, J. Zabalza, H. Zhao, V. Masero, J. Ren, EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges, Sensors 19 (6) (2019) 1423.
[17] J. Zhang, Z. Yin, P. Chen, S. Nichele, Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review, Inf. Fusion 59 (2020) 103–126.
[18] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T.H. Falk, J. Faubert, Deep learning-based electroencephalography analysis: A systematic review, J. Neural Eng. 16 (5) (2019) 051001.
[19] H. Lee, S. Choi, Group nonnegative matrix factorization for EEG classification, in: Artificial Intelligence and Statistics, PMLR, 2009, pp. 320–327.
[20] L. Qin, L. Ding, B. He, Motor imagery classification by means of source analysis for brain–computer interface applications, J. Neural Eng. 1 (3) (2004) 135.
[21] Y. Wang, S. Gao, X. Gao, Common spatial pattern method for channel selelction in motor imagery based brain-computer interface, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, IEEE, 2006, pp. 5392–5395.
[22] K.K. Ang, Z.Y. Chin, H. Zhang, C. Guan, Filter bank common spatial pattern (FBCSP) in brain-computer interface, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), IEEE, 2008, pp. 2390–2397.
[23] E. Dong, C. Li, L. Li, S. Du, A.N. Belkacem, C. Chen, Classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain–computer interfaces, Med. Biol. Eng. Comput. 55 (2017) 1809–1818.
[24] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, S.-W. Lee, EEG dataset and openbmi toolbox for three BCI paradigms: An investigation into BCI illiteracy, GigaScience 8 (5) (2019) giz002.
[25] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) e215–e220.
[26] H. Cho, M. Ahn, S. Ahn, M. Kwon, S.C. Jun, EEG datasets for motor imagery brain–computer interface, GigaScience 6 (7) (2017) gix034.
[27] N. Leeuwis, A. Paas, M. Alimardani, Vividness of visual imagery and personality impact motor-imagery brain computer interfaces, Front. Hum. Neurosci. 15 (2021) 634748.
[28] J.R. Stieger, S. Engel, H. Jiang, C.C. Cline, M.J. Kreitzer, B. He, Mindfulness improves brain–computer interface performance by increasing control over neural activity in the alpha band, Cerebral Cortex 31 (1) (2021) 426–438.
[29] J. Meng, B. He, Exploring training effect in 42 human subjects using a non-invasive sensorimotor rhythm based online BCI, Front. Hum. Neurosci. 13 (2019) 128.
[30] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[31] A. Craik, Y. He, J.L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review, J. Neural Eng. 16 (3) (2019) 031001.
[32] P. Arpaia, A. Esposito, A. Natalizio, M. Parvis, How to successfully classify EEG in motor imagery BCI: A metrological analysis of the state of the art, J. Neural Eng. 19 (3) (2022) 031002.
[33] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, Int. J. Surg. 88 (2021) 105906.
[34] Z. Khademi, F. Ebrahimi, H.M. Kordy, A review of critical challenges in MI-BCI: From conventional to deep learning methods, J. Neurosci. Methods 383 (2023) 109736.
[35] X. Wang, V. Liesaputra, Z. Liu, Y. Wang, Z. Huang, An in-depth survey on deep learning-based motor imagery electroencephalogram (EEG) classification, Artif. Intell. Med. (2023) 102738.
[36] E. Lionakis, K. Karampidis, G. Papadourakis, Current trends, challenges, and future research directions of hybrid and deep learning techniques for motor imagery brain–computer interface, Multimodal Technol. Interact. 7 (10) (2023) 95.
[37] R.T. Schirrmeister, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for EEG decoding and visualization, Hum. Brain Map. 38 (11) (2017) 5391–5420.
[38] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung, B.J. Lance, EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces, J. Neural Eng. 15 (5) (2018) 056013.
[39] J.-H. Jeong, J.-H. Cho, Y.-E. Lee, S.-H. Lee, G.-H. Shin, Y.-S. Kweon, J.d.R. Millán, K.-R. Müller, S.-W. Lee, 2020 International brain–computer interface competition: A review, Front. Hum. Neurosci. 16 (2022) 898300.
[40] S. Lemm, C. Schafer, G. Curio, BCI competition 2003-data set III: probabilistic modeling of sensorimotor/spl mu/rhythms for classification of imaginary hand movements, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1077–1080.
[41] B. Blankertz, BBCI Competition III, 2005, https://www.bbci.de/competition/iii/.
[42] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K.J. Miller, G. Mueller-Putz, et al., Review of the BCI competition IV, Front. Neurosci. (2012) 55.
[43] J.-H. Jeong, J.-H. Cho, K.-H. Shim, B.-H. Kwon, B.-H. Lee, D.-Y. Lee, D.-H. Lee, S.-W. Lee, Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions, GigaScience 9 (10) (2020) giaa098.
[44] M. Kaya, M.K. Binli, E. Ozbay, H. Yanar, Y. Mishchenko, A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces, Sci. Data 5 (1) (2018) 1–16.
[45] S. Brandl, B. Blankertz, Motor imagery under distraction—An open access BCI dataset, Front. Neurosci. 14 (2020) 566147.
[46] X. Ma, S. Qiu, W. Wei, S. Wang, H. He, Deep channel-correlation network for motor imagery decoding from the same limb, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (1) (2019) 297–306.
[47] X. Ma, S. Qiu, W. Wei, S. Wang, H. He, Deep channel-correlation network for motor imagery decoding from the same limb, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (1) (2019) 297–306.
[48] A. Chowdhury, J. Andreu-Perez, Clinical brain–computer interface challenge 2020 (CBCIC at WCCI2020): Overview, methods and results, IEEE Trans. Med. Robot. Bionics 3 (3) (2021) 661–670.
[49] W. Yi, S. Qiu, K. Wang, H. Qi, L. Zhang, P. Zhou, F. He, D. Ming, Evaluation of EEG oscillatory patterns and cognitive process during simple and compound limb motor imagery, PLoS One 9 (12) (2014) e114853.
[50] O. George, R. Smith, P. Madiraju, N. Yahyasoltani, S.I. Ahamed, Data augmentation strategies for EEG-based motor imagery decoding, Heliyon 8 (8) (2022).
[51] O. George, S. Dabas, A. Sikder, R.O. Smith, P. Madiraju, N. Yahyasoltani, S.I. Ahamed, State-of-the-art versus deep learning: A comparative study of motor imagery decoding techniques, IEEE Access 10 (2022) 45605–45619.
[52] O. George, S. Dabas, A. Sikder, R. Smith, P. Madiraju, N. Yahyasoltani, S.I. Ahamed, Enhancing motor imagery decoding via transfer learning, Smart Health 26 (2022) 100339.
[53] H.N. Zahra, H. Zakaria, B.R. Hermanto, Exploration of pattern recognition methods for motor imagery EEG signal with convolutional neural network approach, in: Journal of Physics: Conference Series, Vol. 2312, IOP Publishing, 2022, 012064.
[54] S. Pérez-Velasco, E. Santamaría-Vázquez, V. Martínez-Cagigal, D. Marcos-Martínez, R. Hornero, EEGSym: Overcoming inter-subject variability in motor imagery based BCIs with deep learning, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 1766–1775.
[55] P. Chen, H. Wang, X. Sun, H. Li, C. Grebogi, Z. Gao, Transfer learning with optimal transportation and frequency mixup for EEG-based motor imagery recognition, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 2866–2875.
[56] K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning, J. Big data 3 (1) (2016) 1–40.
[57] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He, A comprehensive survey on transfer learning, Proc. IEEE 109 (1) (2020) 43–76.
[58] H. Dose, J.S. Møller, S. Puthusserypady, H.K. Iversen, A deep learning MI-EEG classification model for BCIs, in: EUSIPCO, 2018, pp. 1676–1679.
[59] S. Kumar, A. Sharma, T. Tsunoda, Brain wave classification using long short-term memory network based OPTICAL predictor, Sci. Rep. 9 (1) (2019) 9153.
[60] M. Mousavi, V.R. de Sa, Temporally adaptive common spatial patterns with deep convolutional neural networks, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2019, pp. 4533–4536.

[61] Z. Tayeb, J. Fedjaev, N. Ghaboosi, C. Richter, L. Everding, X. Qu, Y. Wu, G. Cheng, J. Conradt, Validating deep neural networks for online decoding of motor imagery movements from EEG signals, Sensors 19 (1) (2019) 210.
[62] K. Roots, Y. Muhammad, N. Muhammad, Fusion convolutional neural network for cross-subject EEG motor imagery classification, Computers 9 (3) (2020) 72.
[63] D. Zhang, K. Chen, D. Jian, L. Yao, Motor imagery classification via temporal attention cues of graph embedded EEG signals, IEEE J. Biomed. Health Inform. 24 (9) (2020) 2570–2579.
[64] P. Autthasan, R. Chaisaen, T. Sudhawiyangkul, P. Rangpong, S. Kiatthaveephong, N. Dilokthanakul, G. Bhakdisongkhram, H. Phan, C. Guan, T. Wilaiprasitporn, MIN2Net: End-to-end multi-task learning for subject-independent motor imagery EEG classification, IEEE Trans. Biomed. Eng. 69 (6) (2021) 2105–2118.
[65] D.F. Collazos-Huertas, L.F. Velasquez-Martinez, H.D. Perez-Nastar, A.M. Alvarez-Meza, G. Castellanos-Dominguez, Deep and wide transfer learning with kernel matching for pooling data from electroencephalography and psychological questionnaires, Sensors 21 (15) (2021) 5105.
[66] Z. Jia, Y. Lin, J. Wang, K. Yang, T. Liu, X. Zhang, MMCNN: A multi-branch multi-scale convolutional neural network for motor imagery classification, in: Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part III, Springer, 2021, pp. 736–751.
[67] M. Pals, R.J.P. Belizón, N. Berberich, S.K. Ehrlich, J. Nassour, G. Cheng, Demonstrating the viability of mapping deep learning based EEG decoders to spiking networks on low-powered neuromorphic chips, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2021, pp. 6102–6105.
[68] Y. Pei, Z. Luo, Y. Yan, H. Yan, J. Jiang, W. Li, L. Xie, E. Yin, Data augmentation: Using channel-level recombination to improve classification performance for motor imagery EEG, Front. Hum. Neurosci. 15 (2021) 645952.
[69] S.A. Rammy, W. Abbas, S.S. Mahmood, H. Riaz, H.U. Rehman, R.Z.U. Abideen, M. Aqeel, W. Zhang, Sequence-to-sequence deep neural network with spatio-spectro and temporal features for motor imagery classification, Biocybern. Biomed. Eng. 41 (1) (2021) 97–110.
[70] S. Rasheed, W. Mumtaz, Classification of hand-grasp movements of stroke patients using EEG data, in: 2021 International Conference on Artificial Intelligence, ICAI, IEEE, 2021, pp. 86–90.
[71] L. Xu, M. Xu, Z. Ma, K. Wang, T.-P. Jung, D. Ming, Enhancing transfer performance across datasets for brain-computer interfaces using a combination of alignment strategies and adaptive batch normalization, J. Neural Eng. 18 (4) (2021) 0460e5.
[72] A. Zancanaro, G. Cisotto, J.R. Paulo, G. Pires, U.J. Nunes, CNN-based approaches for cross-subject classification in motor imagery: From the state-of-the-art to DynamicNet, in: 2021 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB, IEEE, 2021, pp. 1–7.
[73] H. Altaheri, G. Muhammad, M. Alsulaiman, Physics-informed attention temporal convolutional network for EEG-based motor imagery classification, IEEE Trans. Ind. Inform. 19 (2) (2022) 2249–2258.
[74] L. Chen, Z. Yu, J. Yang, SPD-CNN: A plain CNN-based model using the symmetric positive definite matrices for cross-subject EEG classification with meta-transfer-learning, Front. Neurorobot. 16 (2022) 958052.
[75] I. Dolzhikova, B. Abibullaev, R. Sameni, A. Zollanvari, Subject-independent classification of motor imagery tasks in EEG using multisubject ensemble CNN, IEEE Access 10 (2022) 81355–81363.
[76] G. Faria, G.H. de Souza, H. Bernardino, L. Motta, A. Vieira, Analyzing data augmentation methods for convolutional neural network-based brain-computer interfaces, in: 2022 International Joint Conference on Neural Networks, IJCNN, IEEE, 2022, pp. 1–8.
[77] J.-X. Huang, C.-Y. Hsieh, Y.-L. Huang, C.-S. Wei, Toward CNN-based motor-imagery EEG classification with fuzzy fusion, Int. J. Fuzzy Syst. 24 (8) (2022) 3812–3823.
[78] W. Ko, E. Jeon, J.S. Yoon, H.-I. Suk, Semi-supervised generative and discriminative adversarial learning for motor imagery-based brain–computer interface, Sci. Rep. 12 (1) (2022) 4587.
[79] F. Mattioli, C. Porcaro, G. Baldassarre, A 1D CNN for high accuracy classification and transfer learning in motor imagery EEG-based brain-computer interface, J. Neural Eng. 18 (6) (2022) 066053.
[80] A. Salami, J. Andreu-Perez, H. Gillmeister, EEG-ITNet: An explainable inception temporal convolutional network for motor imagery classification, IEEE Access 10 (2022) 36672–36685.
[81] Y. Song, Q. Zheng, B. Liu, X. Gao, EEG conformer: Convolutional transformer for EEG decoding and visualization, IEEE Trans. Neural Syst. Rehabil. Eng. 31 (2022) 710–719.
[82] Q. Yang, M. Yang, K. Liu, X. Deng, Enhancing EEG motor imagery decoding performance via deep temporal-domain information extraction, in: 2022 IEEE 11th Data Driven Control and Learning Systems Conference, DDCLS, IEEE, 2022, pp. 420–424.
[83] M. Alnaanah, M. Wahdow, M. Alrashdan, CNN models for EEG motor imagery signal classification, Signal Image Video Process. 17 (3) (2023) 825–830.
[84] M. Nouri, F. Moradi, H. Ghaemi, A.M. Nasrabadi, Towards real-world BCI: CCSPNet, a compact subject-independent motor imagery framework, Digit. Signal Process. 133 (2023) 103816.
[85] A. Partovi, S.M. Hosseini, M. Soleymani, K. Liaghat, S. Ziaee, E.H.P. Fard, S.S. Vajdi, F. Goodarzy, A deep learning algorithm for classifying grasp motions using multi-sessionEEG recordings, in: 2021 9th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2021, pp. 1–6.
[86] D.-K. Han, J.-H. Jeong, Domain generalization for session-independent brain-computer interface, in: 2021 9th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2021, pp. 1–5.
[87] B.-H. Kwon, B.-H. Lee, J.-H. Jeong, Motor imagery classification emphasizing corresponding frequency domain method based on deep learning framework, in: 2021 9th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2021, pp. 1–4.
[88] D.-K. Han, S. Musellim, D.-Y. Kim, J.-H. Jeong, Confidence-aware subject-to-subject transfer learning for brain-computer interface, in: 2022 10th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2022, pp. 1–4.
[89] S.A. El-aal, R.A. Ramadan, N.I. Ghali, EEG signals classifications of motor imagery using adaptive neuro-fuzzy inference system and interval type-2 fuzzy system, Int. J. Intell. Syst. Technol. Appl. 16 (2) (2017) 106–126.
[90] S. Pattnaik, M. Dash, S. Sabut, DWT-based feature extraction and classification for motor imaginary EEG signals, in: 2016 International Conference on Systems in Medicine and Biology, ICSMB, IEEE, 2016, pp. 186–201.
[91] H. Mirvaziri, Z.S. Mobarakeh, Improvement of EEG-based motor imagery classification using ring topology-based particle swarm optimization, Biomed. Signal Process. Control 32 (2017) 69–75.
[92] W. Huang, J. Zhao, W. Fu, A deep learning approach based on CSP for EEG analysis, in: Intelligent Information Processing IX: 10th IFIP TC 12 International Conference, IIP 2018, Nanning, China, October 19-22, 2018, Proceedings 10, Springer, 2018, pp. 62–70.
[93] R. Braga, C. Lopes, T. Becker, Round cosine transform based feature extraction of motor imagery EEG signals, in: World Congress on Medical Physics and Biomedical Engineering 2018: June 3-8, 2018, Prague, Czech Republic (Vol. 2), Springer, 2019, pp. 511–515.
[94] X. Yu, M.Z. Aziz, M.T. Sadiq, Z. Fan, G. Xiao, A new framework for automatic detection of motor and mental imagery EEG signals for robust BCI systems, IEEE Trans. Instrum. Meas. 70 (2021) 1–12.
[95] Y. Hu, L. Wang, W. Fu, EEG feature extraction of motor imagery based on WT and STFT, in: 2018 IEEE International Conference on Information and Automation, ICIA, IEEE, 2018, pp. 83–88.
[96] B. Xu, L. Zhang, A. Song, C. Wu, W. Li, D. Zhang, G. Xu, H. Li, H. Zeng, Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification, IEEE Access 7 (2018) 6084–6093.
[97] T. Wang, E. Dong, S. Du, C. Jia, A shallow convolutional neural network for classifying MI-EEG, in: 2019 Chinese Automation Congress, CAC, IEEE, 2019, pp. 5837–5841.
[98] T.H. Shovon, Z. Al Nazi, S. Dash, M.F. Hossain, Classification of motor imagery EEG signals with multi-input convolutional neural network by augmenting STFT, in: 2019 5th International Conference on Advances in Electrical Engineering, ICAEE, IEEE, 2019, pp. 398–403.
[99] H.K. Lee, Y.-S. Choi, Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface, Entropy 21 (12) (2019) 1199.
[100] J. Kim, Y. Park, W. Chung, Transform based feature construction utilizing magnitude and phase for convolutional neural network in EEG signal classification, in: 2020 8th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2020, pp. 1–4.
[101] Ž. Rohutná, R. Vargic, An approach to EEG based BCI for motor imagery using time-frequency representation and CNN, in: 2022 29th International Conference on Systems, Signals and Image Processing, IWSSIP, IEEE, 2022, pp. 1–4.
[102] Z. Zhang, F. Duan, J. Sole-Casals, J. Dinares-Ferran, A. Cichocki, Z. Yang, Z. Sun, A novel deep learning approach with data augmentation to classify motor imagery signals, IEEE Access 7 (2019) 15945–15954.
[103] X. Tang, J. Zhao, W. Fu, J. Pan, H. Zhou, A novel classification algorithm for MI-EEG based on deep learning, in: 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference, ITAIC, IEEE, 2019, pp. 606–611.
[104] M. Wei, R. Yang, M. Huang, Motor imagery EEG signal classification based on deep transfer learning, in: 2021 IEEE 34th International Symposium on Computer-Based Medical Systems, CBMS, IEEE, 2021, pp. 85–90.
[105] K. Wen, Multi-model ensemble for motor imagery classification using factorization machine, in: International Conference on Advanced Algorithms and Neural Networks, AANN 2022, Vol. 12285, SPIE, 2022, pp. 242–247.
[106] E. Huang, X. Zheng, Y. Fang, Z. Zhang, Classification of motor imagery EEG based on time-domain and frequency-domain dual-stream convolutional neural network, IRBM 43 (2) (2022) 107–113.
[107] A.A. Malibari, F.N. Al-Wesabi, M. Obayya, M.A. Alkhonaini, M.A. Hamza, A. Motwakel, I. Yaseen, A.S. Zamani, et al., Arithmetic optimization with retinanet model for motor imagery classification on brain computer interface, J. Healthc. Eng. 2022 (2022).
[108] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.

[109] S. Salimpour, H. Kalbkhani, S. Seyyedi, V. Solouk, Stockwell transform and semi-supervised feature selection from deep features for classification of BCI signals, Sci. Rep. 12 (1) (2022) 11773.
[110] D.-q. Xu, M.-a. Li, A dual alignment-based multi-source domain adaptation framework for motor imageryEEG classification, Appl. Intell. (2022) 1–23.
[111] S. Phadikar, N. Sinha, R. Ghosh, Unsupervised feature extraction with autoencoders for EEG based multiclass motor imagery BCI, Expert Syst. Appl. 213 (2023) 118901.
[112] I.T. Hettiarachchi, T. Babaei, T. Nguyen, C.P. Lim, S. Nahavandi, A fresh look at functional link neural network for motor imagery-based brain–computer interface, J. Neurosci. Methods 305 (2018) 28–35.
[113] Q. She, K. Chen, Y. Ma, T. Nguyen, Y. Zhang, et al., Sparse representation-based extreme learning machine for motor imagery EEG classification, Comput. Intell. Neurosci. 2018 (2018).
[114] J. Xue, F. Ren, X. Sun, M. Yin, J. Wu, C. Ma, Z. Gao, et al., A multifrequency brain network-based deep learning framework for motor imagery decoding, Neural Plast. 2020 (2020).
[115] X. Deng, B. Zhang, N. Yu, K. Liu, K. Sun, Advanced TSGL-EEGNet for motor imagery EEG-based brain-computer interfaces, IEEE Access 9 (2021) 25118–25130.
[116] M. Miao, W. Hu, W. Zhang, A spatial-frequency-temporal 3D convolutional neural network for motor imagery EEG signal classification, Signal Image Video Process. 15 (2021) 1797–1804.
[117] X. Tang, J. Zhao, W. Fu, Research on extraction and classification of EEG features for multi-class motor imagery, in: 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference, IAEAC, Vol. 1, IEEE, 2019, pp. 693–697.
[118] Z. Dokur, T. Olmez, Classification of motor imagery electroencephalogram signals by using a divergence based convolutional neural network, Appl. Soft Comput. 113 (2021) 107881.
[119] S. Choo, C.S. Nam, DCGAN basedEEG data augmentation in cognitive state recognition, in: IIE Annual Conference. Proceedings, Institute of Industrial and Systems Engineers (IISE), 2020, pp. 1–6.
[120] H. Choi, J. Park, Y.-M. Yang, A novel quick-response eigenface analysis scheme for brain–computer interfaces, Sensors 22 (15) (2022) 5860.
[121] K. Belwafi, S. Gannouni, H. Aboalsamh, An effective zeros-time windowing strategy to detect sensorimotor rhythms related to motor imagery EEG signals, IEEE Access 8 (2020) 152669–152679.
[122] C. Liu, J. Jin, I. Daly, H. Sun, Y. Huang, X. Wang, A. Cichocki, Bispectrum-based hybrid neural network for motor imagery classification, J. Neurosci. Methods 375 (2022) 109593.
[123] R. Salazar-Varas, R.A. Vazquez, Evaluating spiking neural models in the classification of motor imagery EEG signals using short calibration sessions, Appl. Soft Comput. 67 (2018) 232–244.
[124] S. Sethi, R. Upadhyay, H.S. Singh, Stockwell-common spatial pattern technique for motor imagery-based brain computer interface design, Comput. Electr. Eng. 71 (2018) 492–504.
[125] I. Razzak, Cooperative evolution multiclass support matrix machines, in: 2020 International Joint Conference on Neural Networks, IJCNN, IEEE, 2020, pp. 1–8.
[126] W. Liang, J. Jin, I. Daly, H. Sun, X. Wang, A. Cichocki, Novel channel selection model based on graph convolutional network for motor imagery, Cogn. Neurodyn. (2022) 1–14.
[127] F. Fahimi, S. Dosen, K.K. Ang, N. Mrachacz-Kersting, C. Guan, Generative adversarial networks-based data augmentation for brain–computer interface, IEEE Trans. Neural Netw. Learn. Syst. 32 (9) (2020) 4039–4051.
[128] Y. Chen, R. Yang, M. Huang, Z. Wang, X. Liu, Single-source to single-target cross-subject motor imagery classification based on multisubdomain adaptation network, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 1992–2002.
[129] T. Zaremba, A. Atyabi, Cross-subject & cross-dataset subject transfer in motor imagery BCI systems, in: 2022 International Joint Conference on Neural Networks, IJCNN, IEEE, 2022, pp. 1–8.
[130] S. Liu, J. Zhang, A. Wang, H. Wu, Q. Zhao, J. Long, Subject adaptation convolutional neural network for EEG-based motor imagery classification, J. Neural Eng. 19 (6) (2022) 066003.
[131] D. Theng, A. Atyabi, Implication of subject transfer in motor imagery brain computer interfacing systems, in: 2022 International Joint Conference on Neural Networks, IJCNN, IEEE, 2022, pp. 1–8.
[132] M.T. Sadiq, X. Yu, Z. Yuan, Exploiting dimensionality reduction and neural network techniques for the development of expert brain–computer interfaces, Expert Syst. Appl. 164 (2021) 114031.
[133] J. Zhang, D. Liu, W. Chen, Z. Pei, J. Wang, Deep convolutional neural network for EEG-based motor decoding, Micromachines 13 (9) (2022) 1485.
[134] C.J. Ortiz-Echeverri, S. Salazar-Colores, J. Rodríguez-Reséndiz, R.A. Gómez-Loenzo, A new approach for motor imagery classification based on sorted blind source separation, continuous wavelet transform, and convolutional neural network, Sensors 19 (20) (2019) 4541.
[135] M.T. Sadiq, X. Yu, Z. Yuan, F. Zeming, A.U. Rehman, I. Ullah, G. Li, G. Xiao, Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain–computer interfaces, IEEE Access 7 (2019) 171431–171451.
[136] K. Samanta, S. Chatterjee, R. Bose, Cross-subject motor imagery tasks EEG signal classification employing multiplex weighted visibility graph and deep feature extraction, IEEE Sens. Lett. 4 (1) (2019) 1–4.
[137] M.A. Balim, C. Hanilçi, N. Acir, Motor imagery signal classification using constant-Q transform for BCI applications, in: 2021 29th European Signal Processing Conference, EUSIPCO, IEEE, 2021, pp. 1306–1310.
[138] S. Chaudhary, S. Taran, V. Bajaj, A. Sengur, Convolutional neural network based approach towards motor imagery tasks EEG signals classification, IEEE Sens. J. 19 (12) (2019) 4494–4500.
[139] S. Stephe, K.V. Kumar, et al., Motor imagery EEG recognition using deep generative adversarial network with EMD for BCI applications, Tehn. Vjesn. 29 (1) (2022) 92–100.
[140] M.T. Sadiq, X. Yu, Z. Yuan, M.Z. Aziz, Motor imagery BCI classification based on novel two-dimensional modelling in empirical wavelet transform, Electron. Lett. 56 (25) (2020) 1367–1369.
[141] J. Jin, H. Sun, I. Daly, S. Li, C. Liu, X. Wang, A. Cichocki, A novel classification framework using the graph representations of electroencephalogram for motor imagery based brain-computer interface, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2021) 20–29.
[142] M.-a. Li, Z.-w. Ruan, Decoding motor imagery with a simplified distributed dipoles model at source level, Cogn. Neurodyn. 17 (2) (2023) 445–457.
[143] R.D. Pascual-Marqui, et al., Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details, Methods Find Exp. Clin. Pharmacol. 24 (Suppl D) (2002) 5–12.
[144] S. Taheri, M. Ezoji, S.M. Sakhaei, Convolutional neural network based features for motor imagery EEG signals classification in brain–computer interface system, SN Appl. Sci. 2 (2020) 1–12.
[145] Y. Zhang, Y. Wang, G. Zhou, J. Jin, B. Wang, X. Wang, A. Cichocki, Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces, Expert Syst. Appl. 96 (2018) 302–310.
[146] S. Huang, H. Peng, Y. Chen, K. Sun, F. Shen, T. Wang, T. Ma, Tensor discriminant analysis for MI-EEG signal classification using convolutional neural network, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2019, pp. 5971–5974.
[147] H. Sadreazami, Y. Mohammad-Taheri, M. Amini, Hierarchical spectral-temporal feature learning for motor task recognition in brain computer interfaces, in: 2022 IEEE International Instrumentation and Measurement Technology Conference, I2MTC, IEEE, 2022, pp. 1–5.
[148] M. Miao, W. Hu, H. Yin, K. Zhang, Spatial-frequency feature learning and classification of motor imagery EEG based on deep convolution neural network, Comput. Math. Methods Med. 2020 (2020).
[149] G.C. Jana, S. Shukla, D. Srivastava, A. Agrawal, Performance estimation and analysis over the supervised learning approaches for motor imagery EEG signals classification, in: Intelligent Computing and Applications: Proceedings of ICICA 2019, Springer, 2021, pp. 125–141.
[150] T. Fang, Z. Song, G. Zhan, X. Zhang, W. Mu, P. Wang, L. Zhang, X. Kang, Decoding motor imagery tasks using ESI and hybrid feature CNN, J. Neural Eng. 19 (1) (2022) 016022.
[151] R. Mahamune, S.H. Laskar, An automatic channel selection method based on the standard deviation of wavelet coefficients for motor imagery based brain–computer interfacing, Int. J. Imaging Syst. Technol. 33 (2) (2023) 714–728.
[152] P. Chaudhary, Y.V. Varshney, G. Srivastava, S. Bhatia, Motor imagery classification using sparse nonnegative matrix factorization and convolutional neural networks, Neural Comput. Appl. (2022) 1–11.
[153] C. Shi, L. Liu, C. Zhang, X. Wu, To investigate the ability of CNN in learning specific frequency band of motor imagery EEG, in: 2022 7th International Conference on Signal and Image Processing, ICSIP, IEEE, 2022, pp. 650–654.
[154] R. Arabshahi, M. Rouhani, A convolutional neural network and stacked autoencoders approach for motor imagery based brain-computer interface, in: 2020 10th International Conference on Computer and Knowledge Engineering, ICCKE, IEEE, 2020, pp. 295–300.
[155] D.M. Hermosilla, R.T. Codorniú, R.L. Baracaldo, R.S. Zamora, D.D. Rodriguez, Y.L. Albuerne, J.R.N. Álvarez, Shallow convolutional network excel for classifying motor imageryEEG in BCI applications, IEEE Access 9 (2021) 98275–98286.
[156] Y. Ou, S. Sun, H. Gan, R. Zhou, Z. Yang, An improved self-supervised learning for EEG classification, Math. Biosci. Eng. 19 (2022) 6907–6922.
[157] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012).
[158] B. Huang, H. Xu, M. Yuan, M.Z. Aziz, X. Yu, Exploiting asymmetric EEG signals with EFD in deep learning domain for robust BCI, Symmetry 14 (12) (2022) 2677.
[159] A. Khoyani, H. Kaur, M. Amini, H. Sadreazami, Motor imagery brain activity recognition through data augmentation usingdc-gans and mu-sigma, in: 2022 IEEE Sensors, IEEE, 2022, pp. 1–4.
[160] M.T. Sadiq, X. Yu, Z. Yuan, M.Z. Aziz, Identification of motor and mental imagery EEG in two and multiclass subject-dependent tasks using successive decomposition index, Sensors 20 (18) (2020) 5283.

[161] A. Nakra, M. Duhan, Motor imagery EEG signal classification using long short-term memory deep network and neighbourhood component analysis, Int. J. Inf. Technol. 14 (4) (2022) 1771–1779.
[162] A. Nakra, M. Duhan, Brain computer interfacing system using grey wolf optimizer and deep neural networks, Int. J. Inf. Technol. 14 (7) (2022) 3413–3424.
[163] A. Nakra, M. Duhan, Deep neural network with harmony search based optimal feature selection of EEG signals for motor imagery classification, Int. J. Inf. Technol. 15 (2) (2023) 611–625.
[164] Y. Shen, H. Lu, J. Jia, Classification of motor imagery EEG signals with deep learning models, in: Intelligence Science and Big Data Engineering: 7th International Conference, IScIDE 2017, Dalian, China, September 22-23, 2017, Proceedings 6, Springer, 2017, pp. 181–190.
[165] S. Tiwari, S. Goel, A. Bhardwaj, MIDNN-a classification approach for the EEG based motor imagery tasks using deep neural network, Appl. Intell. (2022) 1–20.
[166] M.T. Sadiq, M.Z. Aziz, A. Almogren, A. Yousaf, S. Siuly, A.U. Rehman, Exploiting pretrained CNN models for the development of an EEG-based robust BCI framework, Comput. Biol. Med. 143 (2022) 105242.
[167] X. Zhang, X. Zhou, M. Lin, J. Sun, Shufflenet: An extremely efficient convolutional neural network for mobile devices, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6848–6856.
[168] J. Yang, Z. Ma, J. Wang, Y. Fu, A novel deep learning scheme for motor imagery EEG decoding based on spatial representation fusion, IEEE Access 8 (2020) 202100–202110.
[169] D.R. Alonso, M.M.B. Vellasco, Spatial filter comparison for a brain bomputer interface, in: 2016 IEEE Latin American Conference on Computational Intelligence, LA-CCI, IEEE, 2016, pp. 1–6.
[170] G.C. Jana, A. Swetapadma, P.K. Pattnaik, Enhancing the performance of motor imagery classification to design a robust brain computer interface using feed forward back-propagation neural network, Ain Shams Eng. J. 9 (4) (2018) 2871–2878.
[171] X. Ma, W. Zheng, Z. Peng, J. Yang, FPGA-based rapid electroencephalography signal classification system, in: 2019 IEEE 11th International Conference on Advanced Infocomm Technology, ICAIT, IEEE, 2019, pp. 223–227.
[172] Z.-c. Tang, C. Li, J.-f. Wu, P.-c. Liu, S.-w. Cheng, Classification of EEG-based single-trial motor imagery tasks using a B-CSP method for BCI, Front. Inf. Technol. Electron. Eng. 20 (8) (2019) 1087–1098.
[173] M.R. Mosavi, A. Ayatollahi, S. Afrakhteh, An efficient method for classifying motor imagery using CPSO-trained ANFIS prediction, Evol. Syst. 12 (2021) 319–336.
[174] Z. Wu, B. Sun, X. Zhu, Coupling convolution, transformer and graph embedding for motor imagery brain-computer interfaces, in: 2022 IEEE International Symposium on Circuits and Systems, ISCAS, IEEE, 2022, pp. 404–408.
[175] Y. Han, B. Wang, J. Luo, L. Li, X. Li, A classification method for EEG motor imagery signals based on parallel convolutional neural network, Biomed. Signal Process. Control 71 (2022) 103190.
[176] V. Jayashekar, R. Pandian, Hybrid feature extraction for EEG motor imagery classification using multi-class SSVM, Int. J. Intell. Eng. Syst. 15 (4) (2022).
[177] H.K. Lee, J.-H. Lee, J.-O. Park, Y.-S. Choi, Data-driven data augmentation for motor imagery brain-computer interface, in: 2021 International Conference on Information Networking, ICOIN, IEEE, 2021, pp. 683–686.
[178] D.-K. Kim, Y.-T. Kim, H.-R. Jung, H. Kim, D.-J. Kim, Sequential transfer learning via segment after cue enhances the motor imagery-based brain-computer interface, in: 2021 9th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2021, pp. 1–5.
[179] Z. Yu, W. Chen, T. Zhang, Motor imagery EEG classification algorithm based on improved lightweight feature fusion network, Biomed. Signal Process. Control 75 (2022) 103618.
[180] S. Kamhi, S. Zhang, M. Ait Amou, M. Mouhafid, I. Javaid, I.S. Ahmad, I. Abd El Kader, U. Kulsum, Multi-classification of motor imagery EEG signals using Bayesian optimization-based average ensemble approach, Appl. Sci. 12 (12) (2022) 5807.
[181] P.K. Saha, M.A. Rahman, M.K. Alam, A. Ferdowsi, M.N. Mollah, Common spatial pattern in frequency domain for feature extraction and classification of multichannel EEG signals, SN Comput. Sci. 2 (2021) 1–11.
[182] L. Li, N. Sun, et al., Attention-based DSC-ConvLSTM for multiclass motor imagery classification, Comput. Intell. Neurosci. 2022 (2022).
[183] E. Lashgari, J. Ott, A. Connelly, P. Baldi, U. Maoz, An end-to-end CNN with attentional mechanism applied to raw EEG in a BCI classification task, J. Neural Eng. 18 (4) (2021) 0460e3.
[184] S. Ferreira, A. Silveira, A. Pereira, EEG-based motor imagery classification using multilayer perceptron neural network, in: Brazilian Congress on Biomedical Engineering, Springer, 2020, pp. 1873–1878.
[185] M.I. Chacon-Murguia, E. Rivas-Posada, A CNN-based modular classification scheme for motor imagery using a novel EEG sampling protocol suitable for IoT healthcare systems, Neural Comput. Appl. (2022) 1–22.
[186] M.M. Wankhade, S.S. Chorage, Optimized neural network with refined features for categorization of motor imaginary signals, Int. J. Image Graph. (2022) 2350053.
[187] A.M. Judith, S.B. Priya, R.K. Mahendran, T.R. Gadekallu, L.S. Ambati, Two-phase classification: ANN and A-SVM classifiers on motor imagery BCI, Asian J. Control (2022).
[188] X. Ma, D. Wang, D. Liu, J. Yang, DWT and CNN based multi-class motor imagery electroencephalographic signal recognition, J. Neural Eng. 17 (1) (2020) 016073.
[189] B. Sun, Z. Liu, Z. Wu, C. Mu, T. Li, Graph convolution neural network based end-to-end channel selection and classification for motor imagery brain-computer interfaces, IEEE Trans. Ind. Inform. (2022).
[190] M. Li, L. Wei, Motor imagery classification based on plain convolutional neural network and linear interpolation, J. Shanghai Jiaotong Univ. (Sci.) (2022) 1–9.
[191] K.R. Balmuri, S.R. Madala, P.B. Divakarachari, R.P. de Prado, J. Frnda, Enhanced grasshopper optimization algorithm with extreme learning machines for motor-imagery classification, Asian J. Control (2022).
[192] J. Liu, F. Ye, H. Xiong, Multi-class motor imagery EEG classification method with high accuracy and low individual differences based on hybrid neural network, J. Neural Eng. 18 (4) (2021) 0460f1.
[193] J. Chen, Z. Yu, Z. Gu, Semi-supervised deep learning in motor imagery-based brain-computer interfaces with stacked variational autoencoder, in: Journal of Physics: Conference Series, Vol. 1631, IOP Publishing, 2020, 012007.
[194] C. Liu, J. Jin, R. Xu, S. Li, C. Zuo, H. Sun, X. Wang, A. Cichocki, Distinguishable spatial-spectral feature learning neural network framework for motor imagery-based brain–computer interface, J. Neural Eng. 18 (4) (2021) 0460e4.
[195] W. Ko, J. Yoon, E. Kang, E. Jun, J.-S. Choi, H.-I. Suk, Deep recurrent spatio-temporal neural network for motor imagery based BCI, in: 2018 6th International Conference on Brain-Computer Interface, BCI, IEEE, 2018, pp. 1–3.
[196] P. Wang, M. Wang, Y. Zhou, Z. Xu, D. Zhang, Multiband decomposition and spectral discriminative analysis for motor imagery BCI via deep neural network, Front. Comput. Sci. 16 (2022) 1–13.
[197] G.C. Virgilio, A.J. Sossa, J.M. Antelis, L.E. Falcón, Spiking neural networks applied to the classification of motor tasks in EEG signals., Neural Netw. 122 (2019) 130–143.
[198] M. Sorkhi, M.R. Jahed-Motlagh, B. Minaei-Bidgoli, M.R. Daliri, Hybrid fuzzy deep neural network toward temporal-spatial-frequency features learning of motor imagery signals, Sci. Rep. 12 (1) (2022) 22334.
[199] M.-A. Li, J.-F. Han, L.-J. Duan, A novel MI-EEG imaging with the location information of electrodes, IEEE Access 8 (2019) 3197–3211.
[200] C. Zhang, Y.-K. Kim, A. Eskandarian, EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification, J. Neural Eng. 18 (4) (2021) 046014.
[201] A. Hassanpour, M. Moradikia, H. Adeli, S.R. Khayami, P. Shamsinejadbabaki, A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals, Expert Syst. 36 (6) (2019) e12494.
[202] J.-S. Bang, M.-H. Lee, S. Fazli, C. Guan, S.-W. Lee, Spatio-spectral feature representation for motor imagery classification using convolutional neural networks, IEEE Trans. Neural Netw. Learn. Syst. 33 (7) (2021) 3038–3049.
[203] X. Liu, R. Shi, Q. Hui, S. Xu, S. Wang, R. Na, Y. Sun, W. Ding, D. Zheng, X. Chen, TCACNet: Temporal and channel attention convolutional network for motor imagery classification of EEG-based BCI, Inf. Process. Manage. 59 (5) (2022) 103001.
[204] Z. Chang, C. Zhang, C. Li, Motor imagery EEG classification based on transfer learning and multi-scale convolution network, Micromachines 13 (6) (2022) 927.
[205] Y.K. Musallam, N.I. AlFassam, G. Muhammad, S.U. Amin, M. Alsulaiman, W. Abdul, H. Altaheri, M.A. Bencherif, M. Algabri, Electroencephalography-based motor imagery classification using temporal convolutional network fusion, Biomed. Signal Process. Control 69 (2021) 102826.
[206] G.A. Altuwaijri, G. Muhammad, Electroencephalogram-based motor imagery signals classification using a multi-branch convolutional neural network model with attention blocks, Bioengineering 9 (7) (2022) 323.
[207] X. Niu, N. Lu, J. Kang, Z. Cui, Knowledge-driven feature component interpretable network for motor imagery classification, J. Neural Eng. 19 (1) (2022) 016032.
[208] L. Yang, Y. Song, X. Jia, K. Ma, L. Xie, Two-branch 3D convolutional neural network for motor imagery EEG decoding, J. Neural Eng. 18 (4) (2021) 0460c7.
[209] X. Jia, Y. Song, L. Yang, L. Xie, Joint spatial and temporal features extraction for multi-classification of motor imagery EEG, Biomed. Signal Process. Control 71 (2022) 103247.
[210] G.A. Altuwaijri, G. Muhammad, A multibranch of convolutional neural network models for electroencephalogram-based motor imagery classification, Biosensors 12 (1) (2022) 22.
[211] M. Zheng, B. Yang, A deep neural network with subdomain adaptation for motor imagery brain-computer interface, Med. Eng. Phys. 96 (2021) 29–40.
[212] L. Yang, Y. Song, K. Ma, L. Xie, Motor imagery EEG decoding method based on a discriminative feature learning strategy, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 368–379.
[213] X. Hong, Q. Zheng, L. Liu, P. Chen, K. Ma, Z. Gao, Y. Zheng, Dynamic joint domain adaptation network for motor imagery classification, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 556–565.

[214] P. Chen, Z. Gao, M. Yin, J. Wu, K. Ma, C. Grebogi, Multiattention adaptation network for motor imagery recognition, IEEE Trans. Syst. Man Cybern. Syst. 52 (8) (2021) 5127–5139.

[215] S. Gao, J. Yang, T. Shen, W. Jiang, A parallel feature fusion network combining GRU and CNN for motor imagery EEG decoding, Brain Sci. 12 (9) (2022) 1233.

[216] D. Li, J. Wang, J. Xu, X. Fang, Densely feature fusion based on convolutional neural networks for motor imagery EEG classification, IEEE Access 7 (2019) 132720–132730.

[217] D. Li, J. Xu, J. Wang, X. Fang, Y. Ji, A multi-scale fusion convolutional neural network based on attention mechanism for the visualization analysis of EEG signals decoding, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (12) (2020) 2615–2626.

[218] Y. He, Z. Lu, J. Wang, S. Ying, J. Shi, A self-supervised learning based channel attention MLP-mixer network for motor imagery decoding, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 2406–2417.

[219] J. Huang, A novel residual shrinkage block-based convolutional neural network for improving the recognition of motor imagery EEG signals, Int. J. Intell. Comput. Cybern. (ahead-of-print) (2022).

[220] L. Wang, M. Li, The quantitative application of channel importance in movement intention decoding, Biocybern. Biomed. Eng. 42 (2) (2022) 630–645.

[221] C. Ju, C. Guan, Tensor-cspnet: A novel geometric deep learning framework for motor imagery classification, IEEE Trans. Neural Netw. Learn. Syst. (2022).

[222] Q. Ai, M. Zhao, K. Chen, X. Zhao, L. Ma, Q. Liu, Flexible coding scheme for robotic arm control driven by motor imagery decoding, J. Neural Eng. 19 (5) (2022) 056008.

[223] J.J. Liao, J.J. Luo, T. Yang, R.Q.Y. So, M.C.H. Chua, Effects of local and global spatial patterns in EEG motor-imagery classification using convolutional neural network, Brain Comput. Interfaces 7 (3–4) (2020) 47–56.

[224] C. Liu, J. Jin, I. Daly, S. Li, H. Sun, Y. Huang, X. Wang, A. Cichocki, SincNet-based hybrid neural network for motor imagery EEG decoding, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 540–549.

[225] S.H. Ling, H. Makgawinata, F.H. Monsivais, A.d.S.G. Lourenco, J. Lyu, R. Chai, Classification of EEG motor imagery tasks using convolution neural networks, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2019, pp. 758–761.

[226] J. Zou, Q. She, F. Gao, M. Meng, Multi-task motor imagery EEG classification using broad learning and common spatial pattern, in: Intelligence Science II: Third IFIP TC 12 International Conference, ICIS 2018, Beijing, China, November 2-5, 2018, Proceedings 2, Springer, 2018, pp. 3–10.

[227] J. Xu, H. Zheng, J. Wang, D. Li, X. Fang, Recognition of EEG signal motor imagery intention based on deep multi-view feature learning, Sensors 20 (12) (2020) 3496.

[228] D.-q. Xu, M.-a. Li, A dual alignment-based multi-source domain adaptation framework for motor imagery EEG classification, Appl. Intell. 53 (9) (2023) 10766–10788.

[229] Y. Li, L. Guo, Y. Liu, J. Liu, F. Meng, A temporal-spectral-based squeeze-and-excitation feature fusion network for motor imagery EEG decoding, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 1534–1545.

[230] H.-J. Ahn, D.-H. Lee, J.-H. Jeong, S.-W. Lee, Multiscale convolutional transformer for EEG classification of mental imagery in different modalities, IEEE Trans. Neural Syst. Rehabil. Eng. 31 (2022) 646–656.

[231] M. Shams, A. Sagheer, A natural evolution optimization based deep learning algorithm for neurological disorder classification, Bio-Med. Mater. Eng. 31 (2) (2020) 73–94.

[232] H. Wang, H. Yu, H. Wang, EEG_GENet: A feature-level graph embedding method for motor imagery classification based on EEG signals, Biocybern. Biomed. Eng. 42 (3) (2022) 1023–1040.

[233] Y. Ren, S. Zhang, J. Wang, R. Li, TSC-MI: A temporal spatial convolution neural network fused with mutual information for motor imagery based EEG classification, in: Brain Informatics: 14th International Conference, BI 2021, Virtual Event, September 17–19, 2021, Proceedings 14, Springer, 2021, pp. 474–485.

[234] J.-H. Jeong, J.-H. Choi, K.-T. Kim, S.-J. Lee, D.-J. Kim, H.-M. Kim, Multi-domain convolutional neural networks for lower-limb motor imagery using dry vs. wet electrodes, Sensors 21 (19) (2021) 6672.

[235] Y. Li, N. Zhong, D. Taniar, H. Zhang, MCGNet+: an improved motor imagery classification based on cosine similarity, Brain Inform. 9 (1) (2022) 3.

[236] J. Chen, W. Yi, D. Wang, Filter bank sinc-shallownet with EMD-based mixed noise adding data augmentation for motor imagery classification, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2021, pp. 5837–5841.

[237] W. Qiao, X. Bi, Deep spatial-temporal neural network for classification of EEG-based motor imagery, in: Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science, 2019, pp. 265–272.

[238] S. Sakhavi, C. Guan, S. Yan, Learning temporal information for brain-computer interface using convolutional neural networks, IEEE Trans. Neural Netw. Learn. Syst. 29 (11) (2018) 5619–5629.

[239] M. Riyad, M. Khalil, A. Adib, A novel multi-scale convolutional neural network for motor imagery classification, Biomed. Signal Process. Control 68 (2021) 102747.

[240] J. Hong, F. Shamsi, L. Najafizadeh, A deep learning framework based on dynamic channel selection for early classification of left and right hand motor imagery tasks, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2022, pp. 3550–3553.

[241] C. Phunruangsakao, D. Achanccaray, M. Hayashibe, Deep adversarial domain adaptation with few-shot learning for motor-imagery brain-computer interface, IEEE Access 10 (2022) 57255–57265.

[242] Y. Song, D. Wang, K. Yue, N. Zheng, Z.-J.M. Shen, EEG-based motor imagery classification with deep multi-task learning, in: 2019 International Joint Conference on Neural Networks, IJCNN, IEEE, 2019, pp. 1–8.

[243] N. Shajil, M. Sasikala, A. Arunnagiri, Deep learning classification of two-class motor imagery EEG signals using transfer learning, in: 2020 International Conference on E-Health and Bioengineering, EHB, IEEE, 2020, pp. 1–4.

[244] M. Hersche, P. Rupp, L. Benini, A. Rahimi, Compressing subject-specific brain-computer interface models into one model by superposition in hyperdimensional space, in: 2020 Design, Automation & Test in Europe Conference & Exhibition, DATE, IEEE, 2020, pp. 246–251.

[245] D.-H. Shin, D.-H. Ko, J.-W. Han, T.-E. Kam, Evolutionary reinforcement learning for automated hyperparameter optimization in EEG classification, in: 2022 10th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2022, pp. 1–5.

[246] C. Gao, W. Liu, X. Yang, Convolutional neural network and riemannian geometry hybrid approach for motor imagery classification, Neurocomputing 507 (2022) 180–190.

[247] S.U. Amin, H. Altaheri, G. Muhammad, W. Abdul, M. Alsulaiman, Attention-inception and long-short-term memory-based electroencephalography classification for motor imagery tasks in rehabilitation, IEEE Trans. Ind. Inform. 18 (8) (2021) 5412–5421.

[248] A.L. Alfeo, V. Catrambone, M.G. Cimino, G. Vaglini, G. Valenza, Recognizing motor imagery tasks fromEEG oscillations through a novel ensemble-based neural network architecture, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2021, pp. 5983–5986.

[249] M. Aljalal, R. Djemal, A comparative study of wavelet and CSP features classified using LDA, SVM and ANN in EEG based motor imagery, in: 2017 9th IEEE-GCC Conference and Exhibition, GCCCE, IEEE, 2017, pp. 1–9.

[250] A. Akrout, A. Echtioui, R. Khemakhem, M. Ghorbel, Artificial and convolutional neural network of EEG-based motor imagery classification: A comparative study, in: 2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering, STA, IEEE, 2020, pp. 46–50.

[251] R.D. Chintamani, P. Kumar, R. Karan, Motor imagery classification based on hybrid feature extraction and deep neural network, in: 2021 International Conference on Artificial Intelligence and Smart Systems, ICAIS, IEEE, 2021, pp. 885–893.

[252] K. Meenakshy, et al., Classification of motor imagery signals using ANFIS to control a robotic system, in: 2020 International Conference on Power, Instrumentation, Control and Computing, PICC, IEEE, 2020, pp. 1–5.

[253] V. Agarwal, Adaptive radial basis functions neural network for motor imagery task classification, in: 2019 Twelfth International Conference on Contemporary Computing, IC3, IEEE, 2019, pp. 1–6.

[254] Y. Ma, Y. Song, F. Gao, A novel hybrid CNN-transformer model for EEG motor imagery classification, in: 2022 International Joint Conference on Neural Networks, IJCNN, IEEE, 2022, pp. 1–8.

[255] P. Machingal, M. Thousif, S. Dora, S. Sundaram, Self-regulated learning algorithm for distributed coding based spiking neural classifier, in: 2020 International Joint Conference on Neural Networks, IJCNN, IEEE, 2020, pp. 1–7.

[256] M. Hersche, L. Benini, A. Rahimi, Binarization methods for motor-imagery brain–computer interface classification, IEEE J. Emerg. Sel. Top. Circuits Syst. 10 (4) (2020) 567–577.

[257] M. Riyad, M. Khalil, A. Adib, Hyperparameter Bayesian optimisation applied to ConvNets for motor imagery tasks, in: 2021 International Conference on INnovations in Intelligent Systems and Applications, INISTA, IEEE, 2021, pp. 1–6.

[258] Y.-T. Wu, T.H. Huang, C.Y. Lin, S.J. Tsai, P.-S. Wang, Classification of EEG motor imagery using support vector machine and convolutional neural network, in: 2018 International Automatic Control Conference, CACS, IEEE, 2018, pp. 1–4.

[259] X. Liu, L. Lv, Y. Shen, P. Xiong, J. Yang, J. Liu, Multiscale space-time-frequency feature-guided multitask learning CNN for motor imagery EEG classification, J. Neural Eng. 18 (2) (2021) 026003.

[260] S. Xu, L. Zhu, W. Kong, Y. Peng, H. Hu, J. Cao, A novel classification method for EEG-based motor imagery with narrow band spatial filters and deep convolutional neural network, Cogn. Neurodyn. (2022) 1–11.

[261] A. Pérez-Zapata, A.F. Cardona-Escobar, J.A. Jaramillo-Garzón, G.M. Díaz, Deep convolutional neural networks and power spectral density features for motor imagery classification of EEG signals, in: Augmented Cognition: Intelligent Technologies: 12th International Conference, AC 2018, Held As Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part I, Springer, 2018, pp. 158–169.

[262] I. Majidov, T. Whangbo, Efficient classification of motor imagery electroencephalography signals using deep learning methods, Sensors 19 (7) (2019) 1736.
[263] T.A. Izzuddin, N.M. Safri, M.A. Othman, Compact convolutional neural network (CNN) based on SincNet for end-to-end motor imagery decoding and analysis, Biocybern. Biomed. Eng. 41 (4) (2021) 1629–1645.
[264] T. Fang, J. Wang, W. Mu, Z. Song, X. Zhang, G. Zhan, P. Wang, J. Bin, L. Niu, L. Zhang, et al., Noninvasive neuroimaging and spatial filter transform enable ultra low delay motor imageryEEG decoding, J. Neural Eng. 19 (6) (2022) 066034.
[265] Y. Du, J. Huang, X. Huang, K. Shi, N. Zhou, Dual attentive fusion for EEG-based brain-computer interfaces, Front. Neurosci. 16 (2022) 1044631.
[266] Y.M. Saidutta, J. Zou, F. Fekri, Increasing the learning capacity of BCI systems via CNN-HMM models, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2018, pp. 1–4.
[267] K. Zhao, N. Gao, A dual-size convolutional kernel CNN-based approach toEEG signal classification, in: International Conference on Neural Computing for Advanced Applications, Springer, 2022, pp. 52–65.
[268] G. Nuo, Z. Kai, Research of EEG signal analysis method based on fusion of Riemannian space and convolutional neural network, in: 2022 7th International Conference on Signal and Image Processing, ICSIP, IEEE, 2022, pp. 188–194.
[269] X. Huang, N. Zhou, K.-S. Choi, A generalizable and discriminative learning method for deep EEG-based motor imagery classification, Front. Neurosci. 15 (2021) 760979.
[270] L. Radzinski, T. Kocejko, Deep learning approach on surface EEG based brain computer interface, in: 2022 15th International Conference on Human System Interaction, HSI, IEEE, 2022, pp. 1–6.
[271] T. Liu, D. Yang, A densely connected multi-branch 3D convolutional neural network for motor imagery EEG decoding, Brain Sci. 11 (2) (2021) 197.
[272] J. Shin, W. Chung, Motor imagery classification based on multi-kernel CNN with the amalgamated cross entropy loss, in: 2022 10th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2022, pp. 1–4.
[273] W. Ma, Y. Gong, G. Zhou, Y. Liu, L. Zhang, B. He, A channel-mixing convolutional neural network for motor imagery EEG decoding and feature visualization, Biomed. Signal Process. Control 70 (2021) 103021.
[274] D. Borra, S. Fantozzi, E. Magosso, Interpretable and lightweight convolutional neural network for EEG decoding: Application to movement execution and imagination, Neural Netw. 129 (2020) 55–74.
[275] R. Mahamune, S.H. Laskar, Classification of the four-class motor imagery signals using continuous wavelet transform filter bank-based two-dimensional images, Int. J. Imaging Syst. Technol. 31 (4) (2021) 2237–2248.
[276] A. Kar, S. Bera, S.K. Karri, S. Ghosh, M. Mahadevappa, D. Sheet, A deep convolutional neural network based classification of multi-class motor imagery with improved generalization, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2018, pp. 5085–5088.
[277] A. Echtioui, A. Mlaouah, W. Zouch, M. Ghorbel, C. Mhiri, H. Hamam, A novel convolutional neural network classification approach of motor-imagery EEG recording based on deep learning, Appl. Sci. 11 (21) (2021) 9948.
[278] N.S.r. Holm, S. Puthusserypady, et al., An improved five class MI based BCI scheme for drone control using filter bank CSP, in: 2019 7th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2019, pp. 1–6.
[279] L. Liu, C. Shi, X. Wu, Low quality samples detection in motor imagery EEG data by combining independent component analysis and confident learning, in: 2022 21st International Symposium on Communications and Information Technologies, ISCIT, IEEE, 2022, pp. 269–274.
[280] D.A. Pradana, B. Basari, Design of BCI motor imagery classification using WPT-CSP and CNN, in: AIP Conference Proceedings, Vol. 2537, AIP Publishing, 2022.
[281] A. Echtioui, W. Zouch, M. Ghorbel, C. Mhiri, H. Hamam, Fusion convolutional neural network for multi-class motor imagery of EEG signals classification, in: 2021 International Wireless Communications and Mobile Computing, IWCMC, IEEE, 2021, pp. 1642–1647.
[282] W. Abbas, N.A. Khan, Deepmi: Deep learning for multiclass motor imagery classification, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2018, pp. 219–222.
[283] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, Y.-Z. Guo, A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding, IEEE Trans. Neural Syst. Rehabil. Eng. 27 (6) (2019) 1170–1180.
[284] L. Shen, Y. Xia, Y. Li, M. Sun, A multiscale siamese convolutional neural network with cross-channel fusion for motor imagery decoding, J. Neurosci. Methods 367 (2022) 109426.
[285] W. Ma, Y. Gong, H. Xue, Y. Liu, X. Lin, G. Zhou, Y. Li, A lightweight and accurate double-branch neural network for four-class motor imagery classification, Biomed. Signal Process. Control 75 (2022) 103582.
[286] S.-J. Kim, D.-H. Lee, S.-W. Lee, Rethinking CNN architecture for enhancing decoding performance of motor imagery-based EEG signals, IEEE Access 10 (2022) 96984–96996.
[287] K. Liu, M. Yang, Z. Yu, G. Wang, W. Wu, FBMSNet: A filter-bank multi-scale convolutional neural network for EEG-based motor imagery decoding, IEEE Trans. Biomed. Eng. 70 (2) (2022) 436–445.
[288] C.-S. Wei, T. Koike-Akino, Y. Wang, Spatial component-wise convolutional network (SCCNet) for motor-imagery EEG classification, in: 2019 9th International IEEE/EMBS Conference on Neural Engineering, NER, IEEE, 2019, pp. 328–331.
[289] K.-W. Ha, J.-W. Jeong, Temporal pyramid pooling for decoding motor-imagery EEG signals, IEEE Access 9 (2020) 3112–3125.
[290] W. Huang, Y. Xue, L. Hu, H. Liuli, S-EEGNet: Electroencephalogram signal classification based on a separable convolution neural network with bilinear interpolation, IEEE Access 8 (2020) 131636–131646.
[291] T.M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, L. Benini, EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces, in: 2020 IEEE International Conference on Systems, Man, and Cybernetics, SMC, IEEE, 2020, pp. 2958–2965.
[292] B. Abibullaev, I. Dolzhikova, A. Zollanvari, A brute-force CNN model selection for accurate classification of sensorimotor rhythms in BCIs, IEEE Access 8 (2020) 101014–101023.
[293] M. Riyad, M. Khalil, A. Adib, MI-EEGNET: A novel convolutional neural network for motor imagery classification, J. Neurosci. Methods 353 (2021) 109037.
[294] S. Paszkiel, P. Dobrakowski, The use of multilayer convnets for the purposes of motor imagery classification, in: Conference on Automation, Springer, 2021, pp. 10–19.
[295] M. Riyad, M. Khalil, A. Adib, Incep-EEGNet: a convnet for motor imagery decoding, in: Image and Signal Processing: 9th International Conference, ICISP 2020, Marrakesh, Morocco, June 4–6, 2020, Proceedings 9, Springer, 2020, pp. 103–111.
[296] J. Chen, Z. Yu, Z. Gu, Y. Li, Deep temporal-spatial feature learning for motor imagery-based brain–computer interfaces, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (11) (2020) 2356–2366.
[297] T. Schneider, X. Wang, M. Hersche, L. Cavigelli, L. Benini, Q-EEGNet: An energy-efficient 8-bit quantized parallel EEGNet implementation for edge motor-imagery brain-machine interfaces, in: 2020 IEEE International Conference on Smart Computing, SMARTCOMP, IEEE, 2020, pp. 284–289.
[298] H. Raza, A. Chowdhury, S. Bhattacharyya, S. Samothrakis, Single-trial EEG classification with EEGNet and neural structured learning for improving BCI performance, in: 2020 International Joint Conference on Neural Networks, IJCNN, IEEE, 2020, pp. 1–8.
[299] L. Mousapour, F. Agah, S. Salari, M. Zare, A novel approach to classify motor-imagery EEG with convolutional neural network using network measures, in: 2018 4th Iranian Conference on Signal Processing and Intelligent Systems, ICSPIS, IEEE, 2018, pp. 43–47.
[300] J.-S. Bang, S.-W. Lee, Interpretable convolutional neural networks for subject-independent motor imagery classification, in: 2022 10th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2022, pp. 1–5.
[301] M.-a. Li, Z.-w. Ruan, A novel decoding method for motor imagery tasks with 4D data representation and 3D convolutional neural networks, J. Neural Eng. 18 (4) (2021) 046029.
[302] M. Xu, J. Yao, Z. Zhang, R. Li, B. Yang, C. Li, J. Li, J. Zhang, Learning EEG topographical representation for classification via convolutional neural network, Pattern Recognit. 105 (2020) 107390.
[303] G. Ghorbanzadeh, Z. Nabizadeh, N. Karimi, P. Khadivi, A. Emami, S. Samavi, DGAFF: Deep genetic algorithm fitness formation for EEG bio-signal channel selection, Biomed. Signal Process. Control 79 (2023) 104119.
[304] Y. Yan, H. Zhou, L. Huang, X. Cheng, S. Kuang, A novel two-stage refine filtering method for EEG-based motor imagery classification, Front. Neurosci. 15 (2021) 657540.
[305] J.-S. Bang, S.-W. Lee, Motor imagery classification based on CNN-GRU network with spatio-temporal feature representation, in: Asian Conference on Pattern Recognition, Springer, 2021, pp. 104–115.
[306] A. Al-Saegh, S.A. Dawwd, J.M. Abdul-Jabbar, CutCat: An augmentation method for EEG classification, Neural Netw. 141 (2021) 433–443.
[307] S.A. Rammy, M. Abrar, S.J. Anwar, W. Zhang, Recurrent deep learning for EEG-based motor imagination recognition, in: 2020 3rd International Conference on Advancements in Computational Sciences, ICACS, IEEE, 2020, pp. 1–6.
[308] Z. Sheng, X. Yao, R. Chen, W. Diao, W. Li, Y. Li, Analysis motion imagination EEG signal in spatiotemporal-energy domain, in: Proceedings of the 2019 3rd International Conference on Computer Science and Artificial Intelligence, 2019, pp. 268–272.
[309] T.-j. Luo, C.-l. Zhou, F. Chao, Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network, BMC Bioinform. 19 (1) (2018) 1–18.
[310] H. Li, M. Ding, R. Zhang, C. Xiu, Motor imagery EEG classification algorithm based onCNN-LSTM feature fusion network, Biomed. Signal Process. Control 72 (2022) 103342.
[311] R. Zhang, Q. Zong, L. Dou, X. Zhao, A novel hybrid deep learning scheme for four-class motor imagery classification, J. Neural Eng. 16 (6) (2019) 066004.
[312] N. Trinh, D.M. Duong, B. Tran, H. Su, Hybrid neural networks for electroencephalography motor imaging classification, in: 2022 16th International Conference on Ubiquitous Information Management and Communication, IMCOM, IEEE, 2022, pp. 1–6.
[313] D.-H. Ko, D.-H. Shin, T.-E. Kam, Attention-based spatio-temporal-spectral feature learning for subject-specific EEG classification, in: 2021 9th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2021, pp. 1–4.

[314] S.U. Amin, M. Alsulaiman, G. Muhammad, M.A. Mekhtiche, M.S. Hossain, Deep learning for EEG motor imagery classification based on multi-layer CNNs feature fusion, Future Gener. Comput. Syst. 101 (2019) 542–554.

[315] H. Li, H. Chen, Z. Jia, R. Zhang, F. Yin, A parallel multi-scale time-frequency block convolutional neural network based on channel attention module for motor imagery classification, Biomed. Signal Process. Control 79 (2023) 104066.

[316] J. Chen, W. Yi, D. Wang, J. Du, L. Fu, T. Li, FB-CGANet: Filter bank channel group attention network for multi-class motor imagery classification, J. Neural Eng. 19 (1) (2022) 016011.

[317] R. Zhang, N. Zhang, C. Chen, D. Lv, G. Liu, F. Peng, M. Su, W. Zhou, Motor imagery EEG classification with self-attention-based convolutional neural network, in: 2022 7th International Conference on Intelligent Informatics and Biomedical Science, ICIIBMS, 7, IEEE, 2022, pp. 195–199.

[318] X. Rao, H. Fang, Multi-task recognition motor imagery based on deep learning, in: 2020 7th International Conference on Information Science and Control Engineering, ICISCE, IEEE, 2020, pp. 884–888.

[319] H. Li, D. Zhang, J. Xie, MI-DABAN: A dual-attention-based adversarial network for motor imagery classification, Comput. Biol. Med. 152 (2023) 106420.

[320] T. Lotey, P. Keserwani, G. Wasnik, P.P. Roy, Cross-session motor imagery EEG classification using self-supervised contrastive learning, in: 2022 26th International Conference on Pattern Recognition, ICPR, IEEE, 2022, pp. 975–981.

[321] Q. She, B. Hu, Z. Luo, T. Nguyen, Y. Zhang, A hierarchical semi-supervised extreme learning machine method for EEG recognition, Med. Biol. Eng. Comput. 57 (2019) 147–157.

[322] A. Jafarifarmand, M.A. Badamchizadeh, S. Khanmohammadi, M.A. Nazari, B.M. Tazehkand, A new self-regulated neuro-fuzzy framework for classification of EEG signals in motor imagery BCI, IEEE Trans. Fuzzy Syst. 26 (3) (2017) 1485–1497.

[323] H.K. Lee, J.-S. Myoung, Y.-S. Choi, A lightweight end-to-end neural networks for decoding of motor imagery brain signal, in: 2022 Thirteenth International Conference on Ubiquitous and Future Networks, ICUFN, IEEE, 2022, pp. 411–413.

[324] D. Kostas, F. Rudzicz, Thinker invariance: Enabling deep neural networks for BCI across more people, J. Neural Eng. 17 (5) (2020) 056008.

[325] J. Han, X. Gu, B. Lo, Semi-supervised contrastive learning for generalizable motor imagery EEG classification, in: 2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks, BSN, IEEE, 2021, pp. 1–4.

[326] K. Machida, I. Nambu, Y. Wada, Neural network including alternative pre-processing for electroencephalogram by transposed convolution, in: Neural Information Processing: 27th International Conference, ICONIP 2020, Bangkok, Thailand, November 18–22, 2020, Proceedings, Part V 27, Springer, 2020, pp. 139–146.

[327] Y. Li, Y. Qi, Y. Wang, Avoiding subject-specific model selection via highway networks in EEG signals, in: BIBE 2019; the Third International Conference on Biological Information and Biomedical Engineering, VDE, 2019, pp. 1–5.

[328] M. Strahnen, P. Kessler, Investigation of a deep-learning based brain–computer interface with respect to a continuous control application, IEEE Access 10 (2022) 131090–131100.

[329] J. Xie, S. Chen, Y. Zhang, D. Gao, T. Liu, Combining generative adversarial networks and multi-output CNN for motor imagery classification, J. Neural Eng. 18 (4) (2021) 046026.

[330] X. Liu, Q. Hui, S. Xu, S. Wang, R. Na, Y. Sun, X. Chen, D. Zheng, Tacnet: task-aware electroencephalogram classification for brain-computer interface through a novel temporal attention convolutional network, in: Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, 2021, pp. 660–665.

[331] Z. Khademi, F. Ebrahimi, H.M. Kordy, A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals, Comput. Biol. Med. 143 (2022) 105288.

[332] I. Dolzhikova, B. Abibullaev, R. Sameni, A. Zollanvari, An ensemble cnn for subject-independent classification of motor imagery-based EEG, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2021, pp. 319–324.

[333] G. Zoumpourlis, I. Patras, Covmix: Covariance mixing regularization for motor imagery decoding, in: 2022 10th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2022, pp. 1–7.

[334] S. Saha, K.I.U. Ahmed, R. Mostafa, L. Hadjileontiadis, A. Khandoker, Evidence of variabilities in EEG dynamics during motor imagery-based multiclass brain–computer interface, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (2) (2017) 371–382.

[335] N. Ayoobi, E.B. Sadeghian, A subject-independent brain-computer interface framework based on supervised autoencoder, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2022, pp. 218–221.

[336] X. Liu, S. Xiong, X. Wang, T. Liang, H. Wang, X. Liu, A compact multi-branch 1D convolutional neural network for EEG-based motor imagery classification, Biomed. Signal Process. Control 81 (2023) 104456.

[337] S.U. Amin, G. Muhammad, W. Abdul, M. Bencherif, M. Alsulaiman, Multi-CNN feature fusion for efficient EEG classification, in: 2020 IEEE International Conference on Multimedia & Expo Workshops, ICMEW, IEEE, 2020, pp. 1–6.

[338] X. Liu, Y. Shen, J. Liu, J. Yang, P. Xiong, F. Lin, Parallel spatial–temporal self-attention CNN-based motor imagery classification for BCI, Front. Neurosci. 14 (2020) 587520.

[339] H. Wu, Y. Niu, F. Li, Y. Li, B. Fu, G. Shi, M. Dong, A parallel multiscale filter bank convolutional neural networks for motor imagery EEG classification, Front. Neurosci. 13 (2019) 1275.

[340] X. Jia, Y. Song, L. Xie, Excellent fine-tuning: From specific-subject classification to cross-task classification for motor imagery, Biomed. Signal Process. Control 79 (2023) 104051.

[341] F. Yacine, H. Salah, K. Amar, K. Ahmad, A novel ANN adaptive Riemannian-based kernel classification for motor imagery, Biomed. Phys. Eng. Express 9 (1) (2022) 015010.

[342] X. Zhao, D. Liu, L. Ma, Q. Liu, K. Chen, S. Xie, Q. Ai, Deep CNN model based on serial-parallel structure optimization for four-class motor imagery EEG classification, Biomed. Signal Process. Control 72 (2022) 103338.

[343] D. Milanés-Hermosilla, R. Trujillo Codorniú, R. López-Baracaldo, R. Sagaró-Zamora, D. Delisle-Rodriguez, J.J. Villarejo-Mayor, J.R. Núñez-Álvarez, Monte carlo dropout for uncertainty estimation and motor imagery classification, Sensors 21 (21) (2021) 7241.

[344] A. Bria, C. Marrocco, F. Tortorella, Sinc-based convolutional neural networks for EEG-BCI-based motor imagery classification, in: International Conference on Pattern Recognition, Springer, 2021, pp. 526–535.

[345] Y. Wen, W. He, Y. Zhang, A new attention-based 3D densely connected cross-stage-partial network for motor imagery classification in BCI, J. Neural Eng. 19 (5) (2022) 056026.

[346] P. Wang, A. Jiang, X. Liu, J. Shang, L. Zhang, LSTM-based EEG classification in motor imagery tasks, IEEE Trans. Neural Syst. Rehabil. Eng. 26 (11) (2018) 2086–2095.

[347] Q. He, L. Feng, G. Jiang, P. Xie, Multimodal multitask neural network for motor imagery classification with EEG and fNIRS signals, IEEE Sens. J. 22 (21) (2022) 20695–20706.

[348] X. Liu, K. Wang, F. Liu, W. Zhao, J. Liu, 3D convolution neural network with multiscale spatial and temporal cues for motor imagery EEG classification, Cogn. Neurodyn. (2022) 1–24.

[349] R. Zhang, Q. Zong, L. Dou, X. Zhao, Y. Tang, Z. Li, Hybrid deep neural network using transfer learning for EEG motor imagery decoding, Biomed. Signal Process. Control 63 (2021) 102144.

[350] H. Zhao, Q. Zheng, K. Ma, H. Li, Y. Zheng, Deep representation-based domain adaptation for nonstationary EEG classification, IEEE Trans. Neural Netw. Learn. Syst. 32 (2) (2020) 535–545.

[351] X. Wu, R.H. Chan, Does meta-learning improve EEG motor imagery classification? in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2022, pp. 4048–4051.

[352] S. Lian, J. Xu, G. Zuo, X. Wei, H. Zhou, A novel time-incremental end-to-end shared neural network with attention-based feature fusion for multiclass motor imagery recognition, Comput. Intell. Neurosci. 2021 (2021) 1–16.

[353] S. An, S. Kim, P. Chikontwe, S.H. Park, Few-shot relation learning with attention for EEG-based motor imagery classification, in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, IEEE, 2020, pp. 10933–10938.

[354] J. Luo, W. Shi, N. Lu, J. Wang, H. Chen, Y. Wang, X. Lu, X. Wang, X. Hei, Improving the performance of multisubject motor imagery-based BCIs using twin cascaded softmax CNNs, J. Neural Eng. 18 (3) (2021) 036024.

[355] J.-H. Jeong, K.-T. Kim, D.-J. Kim, S.J. Lee, H. Kim, Subject-transfer decoding using the convolutional neural network for motor imagery-based brain-computer interface, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2022, pp. 48–51.

[356] H. Raza, S. Samothrakis, Bagging adversarial neural networks for domain adaptation in non-stationary eeg, in: 2019 International Joint Conference on Neural Networks, IJCNN, IEEE, 2019, pp. 1–7.

[357] X. Li, M. Peng, S. Chen, W. Zheng, Y. Zhang, D. Gao, M. Wang, EEG motor imagery classification based on multi-spatial convolutional neural network, in: 2022 5th International Conference on Artificial Intelligence and Big Data, ICAIBD, IEEE, 2022, pp. 433–437.

[358] Y. Zhang, S. Qiu, W. Wei, X. Ma, H. He, Filter bank adversarial domain adaptation for motor imagery brain computer interface, in: 2021 International Joint Conference on Neural Networks, IJCNN, IEEE, 2021, pp. 1–7.

[359] G. Tian, Y. Liu, Study on classification of left-right hands motor imagery EEG signals based on CNN, in: 2018 IEEE 17th International Conference on Cognitive Informatics & Cognitive Computing, ICCI* CC, IEEE, 2018, pp. 324–329.

[360] G. Tian, Y. Liu, Simple convolutional neural network for left-right hands motor imageryEEG signals classification, Int. J. Cognit. Inform. Natural Intell. (IJCINI) 13 (3) (2019) 36–49.

[361] J.C. Gomes, M.C.A. Rodrigues, W.P. dos Santos, ASTERI: Image-based representation of EEG signals for motor imagery classification, Res. Biomed. Eng. 38 (2) (2022) 661–681.

[362] J. Cai, C. Wei, X.-L. Tang, C. Xue, Q. Chang, The motor imagination EEG recognition combined with convolution neural network and gated recurrent unit, in: 2018 37th Chinese Control Conference, CCC, IEEE, 2018, pp. 9598–9602.
[363] X. Tang, T. Wang, Y. Du, Y. Dai, Motor imagery EEG recognition with KNN-based smooth auto-encoder, Artif. Intell. Med. 101 (2019) 101747.
[364] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, W. Che, A deep transfer convolutional neural network framework for EEG signal classification, IEEE Access 7 (2019) 112767–112776.
[365] T. Xie, D. Kong, Q. Liu, Z. Yan, X. Tang, EEG recognition based on parallel stacked denoise autoencoder and convolutional neural network, in: Software Engineering Perspectives in Intelligent Systems: Proceedings of 4th Computational Methods in Systems and Software 2020, Vol. 2 4, Springer, 2020, pp. 700–713.
[366] B. Sun, X. Zhao, H. Zhang, R. Bai, T. Li, EEG motor imagery classification with sparse spectrotemporal decomposition and deep learning, IEEE Trans. Autom. Sci. Eng. 18 (2) (2020) 541–551.
[367] Z. Chen, Y. Wang, Z. Song, Classification of motor imagery electroencephalography signals based on image processing method, Sensors 21 (14) (2021) 4646.
[368] T. Tao, Z. Han, G. Xu, K. Zhang, A convolution neural network with mixed-size kernels for time-frequency characteristics of motor imagery, in: 2021 7th International Conference on Computing and Artificial Intelligence, 2021, pp. 69–75.
[369] C. Wang, Y. Wu, C. Wang, Y. Zhu, C. Wang, Y. Niu, Z. Shao, X. Gao, Z. Zhao, Y. Yu, MI-EEG classification using Shannon complex wavelet and convolutional neural networks, Appl. Soft Comput. 130 (2022) 109685.
[370] F. Li, F. He, F. Wang, D. Zhang, Y. Xia, X. Li, A novel simplified convolutional neural network classification algorithm of motor imagery EEG signals based on deep learning, Appl. Sci. 10 (5) (2020) 1605.
[371] Y. Rong, X. Wu, Y. Zhang, Classification of motor imagery electroencephalography signals using continuous small convolutional neural network, Int. J. Imaging Syst. Technol. 30 (3) (2020) 653–659.
[372] X. Tang, W. Li, X. Li, W. Ma, X. Dang, Motor imagery EEG recognition based on conditional optimization empirical mode decomposition and multi-scale convolutional neural network, Expert Syst. Appl. 149 (2020) 113285.
[373] R. Zhang, Q. Zong, X. Zhao, A new convolutional neural network for motor imagery classification, in: 2019 Chinese Control Conference, CCC, IEEE, 2019, pp. 8428–8432.
[374] W. Huang, L. Wang, Z. Yan, Y. Liu, Classify motor imagery by a novel CNN with data augmentation, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2020, pp. 192–195.
[375] J. Yang, S. Gao, T. Shen, A two-branch CNN fusing temporal and frequency features for motor imagery EEG decoding, Entropy 24 (3) (2022) 376.
[376] S. Roy, K. McCreadie, G. Prasad, Can a single model deep learning approach enhance classification accuracy of an EEG-based brain-computer interface? in: 2019 IEEE International Conference on Systems, Man and Cybernetics, SMC, IEEE, 2019, pp. 1317–1321.
[377] R. Li, L. Wang, P. Suganthan, O. Sourina, Sample-based data augmentation based on electroencephalogram intrinsic characteristics, IEEE J. Biomed. Health Inf. 26 (10) (2022) 4996–5003.
[378] H.K. Lee, Y.-S. Choi, A convolution neural networks scheme for classification of motor imagery EEG based on wavelet time-frequecy image, in: 2018 International Conference on Information Networking, ICOIN, IEEE, 2018, pp. 906–909.
[379] R. Kumar, S. Saha, A multi-task learning scheme for motor imagery signal classification, in: Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part III 28, Springer, 2021, pp. 311–322.
[380] A.A. Khan, A. Hassan, M.T. Jahangir, Subject wise motor imagery classification from EEG data using transfer learning, in: 2022 24th International Multitopic Conference, INMIC, IEEE, 2022, pp. 1–6.
[381] M. Parvan, A.R. Ghiasi, T.Y. Rezaii, A. Farzamnia, Transfer learning based motor imagery classification using convolutional neural networks, in: 2019 27th Iranian Conference on Electrical Engineering, ICEE, IEEE, 2019, pp. 1825–1828.
[382] X. Dang, G. Liu, X. Tang, S. Wang, T. Wang, M. Zou, Motor imagery EEG recognition based on generative and discriminative adversarial learning framework and hybrid scale convolutional neural network, IAENG Int. J. Appl. Math. 52 (4) (2022) 1–9.
[383] M. Dai, D. Zheng, R. Na, S. Wang, S. Zhang, EEG classification of motor imagery using a novel deep learning framework, Sensors 19 (3) (2019) 551.
[384] Z. Li, Y. Yu, Improving EEG-based motor imagery classification with conditional wasserstein GAN, in: 2020 International Conference on Image, Video Processing and Artificial Intelligence, 11584, SPIE, 2020, pp. 437–443.
[385] N. Lu, T. Li, X. Ren, H. Miao, A deep learning scheme for motor imagery classification based on restricted Boltzmann machines, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (6) (2016) 566–576.
[386] Y. Liu, S. Zhang, W. Xiao, B. Zhao, Weight-sharing network structure based on multi-channel EEG time-frequency map, Biomed. Signal Process. Control 76 (2022) 103698.
[387] R.H. Elessawy, S. Eldawlatly, H.M. Abbas, A long short-term memory autoencoder approach for EEG motor imagery classification, in: 2020 International Conference on Computation, Automation and Knowledge Management, ICCAKM, IEEE, 2020, pp. 79–84.
[388] A. Wang, Z. Gong, Distribution based learning network for motor imagery electroencephalogram classification, in: 2021 IEEE 6th International Conference on Computer and Communication Systems, ICCCS, IEEE, 2021, pp. 357–362.
[389] Z. Jin, G. Zhou, D. Gao, Y. Zhang, EEG classification using sparse Bayesian extreme learning machine for brain–computer interface, Neural Comput. Appl. 32 (2020) 6601–6609.
[390] K.-W. Ha, J.-W. Jeong, Motor imagery EEG classification using capsule networks, Sensors 19 (13) (2019) 2854.
[391] K.-W. Ha, J.-W. Jeong, Decoding two-class motor imagery EEG with capsule networks, in: 2019 IEEE International Conference on Big Data and Smart Computing, BigComp, IEEE, 2019, pp. 1–4.
[392] A. Usman, H. Li, R. Yao, Q. Wang, W. Hussain, S.B. ud Duja, M. Amjad, B. Ahmed, EEG emotion signal of artificial neural network by using capsule network, Int. J. Adv. Comput. Sci. Appl. 11 (1) (2020).
[393] K. Zhang, G. Xu, L. Chen, P. Tian, C. Han, S. Zhang, N. Duan, Instance transfer subject-dependent strategy for motor imagery signal classification using deep convolutional neural networks, Comput. Math. Methods Med. 2020 (2020).
[394] A.M. Roy, An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces, Biomed. Signal Process. Control 74 (2022) 103496.
[395] B. Yang, C. Fan, C. Guan, X. Gu, M. Zheng, A framework on optimization strategy for EEG motor imagery recognition, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2019, pp. 774–777.
[396] S. Roy, A. Chowdhury, K. McCreadie, G. Prasad, Deep learning based inter-subject continuous decoding of motor imagery for practical brain-computer interfaces, Front. Neurosci. 14 (2020) 918.
[397] X. Zhu, P. Li, C. Li, D. Yao, R. Zhang, P. Xu, Separated channel convolutional neural network to realize the training free motor imagery BCI systems, Biomed. Signal Process. Control 49 (2019) 396–403.
[398] B. Sun, Z. Wu, Y. Hu, T. Li, Golden subject is everyone: A subject transfer neural network for motor imagery-based brain computer interfaces, Neural Netw. 151 (2022) 111–120.
[399] O. Özdenizci, Y. Wang, T. Koike-Akino, D. Erdoğmuş, Learning invariant representations from EEG via adversarial inference, IEEE Access 8 (2020) 27074–27085.
[400] E. Jeon, W. Ko, J.S. Yoon, H.-I. Suk, Mutual information-driven subject-invariant and class-relevant deep representation learning in BCI, IEEE Trans. Neural Netw. Learn. Syst. (2021).
[401] E.M. Santos, R. San-Martin, F.J. Fraga, Comparison of LORETA and CSP for brain-computer interface applications, in: 2021 18th International Multi-Conference on Systems, Signals & Devices, SSD, IEEE, 2021, pp. 817–822.
[402] H.-S. Kim, M.-H. Ahn, B.-K. Min, Deep-learning-based automatic selection of fewest channels for brain–machine interfaces, IEEE Trans. Cybern. 52 (9) (2021) 8668–8680.
[403] D.F. Collazos-Huertas, A.M. Álvarez-Meza, G. Castellanos-Dominguez, Image-based learning using gradient class activation maps for enhanced physiological interpretability of motor imagery skills, Appl. Sci. 12 (3) (2022) 1695.
[404] L. Yue, H. Shen, S. Wang, R. Boots, G. Long, W. Chen, X. Zhao, Exploring BCI control in smart environments: intention recognition via EEG representation enhancement learning, ACM Trans. Knowl. Discov. Data (TKDD) 15 (5) (2021) 1–20.
[405] P. Alfeld, A trivariate Clough—Tocher scheme for tetrahedral data, Comput. Aided Geom. Design 1 (2) (1984) 169–181.
[406] W. Fadel, C. Kollod, M. Wahdow, Y. Ibrahim, I. Ulbert, Multi-class classification of motor imagery EEG signals using image-based deep recurrent convolutional neural network, in: 2020 8th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2020, pp. 1–4.
[407] M. Sorkhi, M.R. Jahed-Motlagh, B. Minaei-Bidgoli, M. Reza Daliri, Learning temporal-frequency features of physionet EEG signals using deep convolutional neural network, Internat. J. Modern Phys. C 34 (04) (2023) 2350047.
[408] V.A. Lomelin-Ibarra, A.E. Gutierrez-Rodriguez, J.A. Cantoral-Ceballos, Motor imagery analysis from extensive EEG data representations using convolutional neural networks, Sensors 22 (16) (2022) 6093.
[409] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[410] T. Wang, J. Mao, R. Xiao, W. Wang, G. Ding, Z. Zhang, Residual learning attention cnn for motion intention recognition based on EEG data, in: 2021 IEEE Biomedical Circuits and Systems Conference, BioCAS, IEEE, 2021, pp. 1–6.
[411] H. Alwasiti, M.Z. Yusoff, K. Raza, Motor imagery classification for brain computer interface using deep metric learning, IEEE Access 8 (2020) 109949–109963.
[412] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, K. Keutzer, Densenet: Implementing efficient convnet descriptor pyramids, 2014, arXiv preprint arXiv:1404.1869.

[413] H. Alwasiti, M.Z. Yusoff, Motor imagery classification for brain computer interface using deep convolutional neural networks and mixup augmentation, IEEE Open J. Eng. Med. Biol. 3 (2022) 171–177.

[414] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[415] J. Nekrasova, M. Kanarskii, A. Sudareva, Comparison of classification algorithms for motor imagery brain-computer interface, in: 2020 International Conference on Electrotechnical Complexes and Systems, ICOECS, IEEE, 2020, pp. 1–6.

[416] M.A. Awais, M.Z. Yusoff, D.M. Khan, N. Yahya, N. Kamel, M. Ebrahim, Effective connectivity for decoding electroencephalographic motor imagery using a probabilistic neural network, Sensors 21 (19) (2021) 6570.

[417] J.-H. Kang, J. Youn, S.-H. Kim, J. Kim, Effects of frontal theta rhythms in a prior resting state on the subsequent motor imagery brain-computer interface performance, Front. Neurosci. 15 (2021) 663101.

[418] C. Ju, D. Gao, R. Mane, B. Tan, Y. Liu, C. Guan, Federated transfer learning for EEG signal classification, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2020, pp. 3040–3045.

[419] X. Wang, M. Hersche, B. Tömekce, B. Kaya, M. Magno, L. Benini, An accurate EEGNet-based motor-imagery brain–computer interface for low-power edge computing, in: 2020 IEEE International Symposium on Medical Measurements and Applications, MeMeA, IEEE, 2020, pp. 1–6.

[420] B. Vivek, A. Adarsh, J. Gubbi, K. Muralidharan, R.K. Ramakrishnan, A. Pal, ST-GNN for EEG motor imagery classification, in: 2022 IEEE-EMBS International Conference on Biomedical and Health Informatics, BHI, IEEE, 2022, pp. 01–04.

[421] J. Xie, J. Zhang, J. Sun, Z. Ma, L. Qin, G. Li, H. Zhou, Y. Zhan, A transformer-based approach combining deep learning network and spatial-temporal information for raw EEG classification, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 2126–2136.

[422] O. Özdenizci, Y. Wang, T. Koike-Akino, D. Erdoğmuş, Transfer learning in brain-computer interfaces with adversarial variational autoencoders, in: 2019 9th International IEEE/EMBS Conference on Neural Engineering, NER, IEEE, 2019, pp. 207–210.

[423] H. Chu, C. Zhang, Motor imagery EEG classification method: based on a novel biLSTM-attention-CNN hybrid neural network, in: 2nd International Conference on Signal Image Processing and Communication, ICSIPC 2022, Vol. 12246, SPIE, 2022, pp. 291–297.

[424] S.J.H. Shah, A.A. Albishri, Y. Lee, Deep learning framework for internet of things for people with disabilities, in: 2021 IEEE International Conference on Big Data, Big Data, IEEE, 2021, pp. 3609–3614.

[425] J. Sun, J. Xie, H. Zhou, EEG classification with transformer-based models, in: 2021 Ieee 3rd Global Conference on Life Sciences and Technologies, Lifetech, IEEE, 2021, pp. 92–93.

[426] H. Dose, J.S. Møller, H.K. Iversen, S. Puthusserypady, An end-to-end deep learning approach to MI-EEG signal classification for BCIs, Expert Syst. Appl. 114 (2018) 532–542.

[427] A.C. Hernandez-Ruiz, D. Enériz, N. Medrano, B. Calvo, Motor-imagery EEGNet-based processing on a low-spec soc hardware, in: 2021 IEEE Sensors, IEEE, 2021, pp. 1–4.

[428] L. Wang, M. Li, A novel DCNN based MI-EEG classification method using spatio-frequency information, in: 2021 China Automation Congress, CAC, IEEE, 2021, pp. 532–537.

[429] B. Du, Y. Liu, G. Tian, Improving motor imagery EEG classification by CNN with data augmentation, in: 2020 IEEE 19th International Conference on Cognitive Informatics & Cognitive Computing, ICCI* CC, IEEE, 2020, pp. 111–118.

[430] J. O'Neill, R. Detyens, R. Integlia, S. Oprisan, 3D convolutional neural networks on an EEG-based motor imagery dataset, in: SoutheastCon 2022, IEEE, 2022, pp. 245–249.

[431] J. Li, Z. Shi, Y. Li, Research on EEG-based motor imagery tasks recognition using deep learning approach, in: Chinese Intelligent Systems Conference, Springer, 2022, pp. 416–425.

[432] J. Li, Y. Li, H. Yang, M. Du, Research on intention recognition for EEG signals based on DSCNN and GRU, in: Proceedings of 2021 Chinese Intelligent Systems Conference: Volume I, Springer, 2022, pp. 604–612.

[433] P. Khetrapal, S. Kadambari, Classification of motor imagery waves using hybrid-convolutional neural network, in: 2020 IEEE 17th India Council International Conference, INDICON, IEEE, 2020, pp. 1–7.

[434] Y. Li, H. Yang, J. Li, D. Chen, M. Du, EEG-based intention recognition with deep recurrent-convolution neural network: Performance and channel selection by grad-CAM, Neurocomputing 415 (2020) 225–233.

[435] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.

[436] N. Robinson, S.-W. Lee, C. Guan, EEG representation in deep convolutional neural networks for classification of motor imagery, in: 2019 IEEE International Conference on Systems, Man and Cybernetics, SMC, IEEE, 2019, pp. 1322–1326.

[437] R. Mane, N. Robinson, A.P. Vinod, S.-W. Lee, C. Guan, A multi-view CNN with novel variance layer for motor imagery brain computer interface, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2020, pp. 2950–2953.

[438] A. Nagarajan, N. Robinson, C. Guan, Relevance-based channel selection in motor imagery brain–computer interface, J. Neural Eng. 20 (1) (2023) 016024.

[439] L. Nie, H. Cai, Y. Wu, Y. Zhang, An end-to-end hemisphere discrepancy network for subject-independent motor imagery classification, in: Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part III 28, Springer, 2021, pp. 323–333.

[440] J. Zhang, K. Li, A pruned deep learning approach for classification of motor imagery electroencephalography signals, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2022, pp. 4072–4075.

[441] R. Vishnupriya, N. Robinson, R. Reddy, C. Guan, Performance evaluation of compressed deep CNN for motor imagery classification using EEG, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2021, pp. 795–799.

[442] M. Xu, J. Yao, H. Ni, Dual head and dual attention in deep learning for end-to-end EEG motor imagery classification, Appl. Sci. 11 (22) (2021) 10906.

[443] W. Ko, E. Jeon, H.-I. Suk, A novel RL-assisted deep learning framework for task-informative signals selection and classification for spontaneous BCIs, IEEE Trans. Ind. Inform. 18 (3) (2020) 1873–1882.

[444] B.-H. Lee, J.-H. Cho, B.-H. Kwon, A factorization approach for motor imagery classification, in: 2022 10th International Winter Conference on Brain-Computer Interface, BCI, IEEE, 2022, pp. 1–5.

[445] H. Zhu, D. Forenzo, B. He, On the deep learning models for EEG-based brain-computer interface using motor imagery, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 2283–2291.

[446] S. Cai, H. Li, Q. Wu, J. Liu, Y. Zhang, Motor imagery decoding in the presence of distraction using graph sequence neural networks, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 1716–1726.

[447] N. Tibrewal, N. Leeuwis, M. Alimardani, Classification of motor imagery EEG using deep learning increases performance in inefficient BCI users, PLoS One 17 (7) (2022) e0268880.

[448] X. Zhou, R. Zou, X. Huang, Single upper limb functional movements decoding from motor imagery EEG signals using wavelet neural network, Biomed. Signal Process. Control 70 (2021) 102965.

[449] J.-H. Cho, J.-H. Jeong, S.-W. Lee, Neurograsp: Real-time EEG classification of high-level motor imagery tasks using a dual-stage deep learning framework, IEEE Trans. Cybern. 52 (12) (2021) 13279–13292.

[450] P. Ofner, A. Schwarz, J. Pereira, G.R. Müller-Putz, Upper limb movements can be decoded from the time-domain of low-frequency EEG, PLoS One 12 (8) (2017) e0182578.

[451] H. Zhang, M. Zhao, C. Wei, D. Mantini, Z. Li, Q. Liu, EEGdenoisenet: a benchmark dataset for deep learning solutions of EEG denoising, J. Neural Eng. 18 (5) (2021) 056057.

[452] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, 2014, arXiv preprint arXiv:1409.0473.

[453] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[454] Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, Neurocomputing 452 (2021) 48–62.

[455] G. Cisotto, A. Zanga, J. Chlebus, I. Zoppis, S. Manzoni, U. Markowska-Kaczmar, 2020, arXiv preprint arXiv:2012.01074.

[456] Z. Wan, R. Yang, M. Huang, N. Zeng, X. Liu, A review on transfer learning in EEG signal analysis, Neurocomputing 421 (2021) 1–14.

[457] M. Hossin, M.N. Sulaiman, A review on evaluation metrics for data classification evaluations, Int. J. Data Min. Knowl. Manage. Process 5 (2) (2015) 1.

[458] Y. Zheng, Y. Luo, H. Shao, L. Zhang, L. Li, DABaCLT: A data augmentation bias-aware contrastive learning framework for time series representation, Appl. Sci. 13 (13) (2023) 7908.

[459] C. Jeunet, B. N'Kaoua, F. Lotte, Advances in user-training for mental-imagery-based BCI control: Psychological and cognitive factors and their neural correlates, Prog. Brain Res. 228 (2016) 3–35.

[460] S. Zhang, H. Tong, J. Xu, R. Maciejewski, Graph convolutional networks: a comprehensive review, Comput. Soc. Netw. 6 (1) (2019) 1–23.

[461] B. Sun, H. Zhang, Z. Wu, Y. Zhang, T. Li, Adaptive spatiotemporal graph convolutional networks for motor imagery classification, IEEE Signal Process. Lett. 28 (2021) 219–223.

[462] W. Liang, J. Jin, I. Daly, H. Sun, X. Wang, A. Cichocki, Novel channel selection model based on graph convolutional network for motor imagery, Cogn. Neurodyn. 17 (5) (2023) 1283–1296.

[463] S. Cai, T. Schultz, H. Li, Brain topology modeling with EEG-graphs for auditory spatial attention detection, IEEE Trans. Biomed. Eng. (2023).

[464] W. Ding, M. Abdel-Basset, H. Hawash, A.M. Ali, Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey, Inform. Sci. (2022).

[465] C. Ieracitano, N. Mammone, A. Hussain, F.C. Morabito, A novel explainable machine learning approach for EEG-based brain-computer interface systems, Neural Comput. Appl. (2021) 1–14.
[466] Y. Chu, X. Zhao, Y. Zou, W. Xu, J. Han, Y. Zhao, A decoding scheme for incomplete motor imagery EEG with deep belief network, Front. Neurosci. 12 (2018) 680.
[467] S. Salardi, et al., Neurotecnologie tra Potere e Libertà. Medicina, Etica, Discriminazioni di Genere, vol. 6, CEDAM-Wolters Kluwer, 2024.
[468] R.E. Ashcroft, et al., The declaration of helsinki, Oxford Textbook Clin. Res. Ethics (2008) 141–148.
[469] A. Coin, M. Mulder, V. Dubljević, Ethical aspects of BCI technology: what is the state of the art? Philosophies 5 (4) (2020) 31.
[470] T. Denning, Y. Matsuoka, T. Kohno, Neurosecurity: security and privacy for neural devices, Neurosurg. Focus 27 (1) (2009) E7.

Aurora Saibene is a Postdoctoral Research Fellow at the Department of Informatics, Systems and Communication of the University of Milano-Bicocca, Italy (2022) and member of MultiMedia Signal Processing laboratory (2018). She has been a member of the Milan Center for Neuroscience (NeuroMI) since 2019. She obtained her Ph.D. (2022) in Computer Science with the additional Doctor Europaeus certification at the University of Milano-Bicocca. Her current research interests include brain–computer interfacing, electroencephalography, human–machine interaction, and signal processing.

Hafez Ghaemi is a Ph.D. student at University of Montreal and Mila, Quebec AI Institute, Canada. Previously, he completed his M.Sc. in Computer Engineering: AI and Robotics at University of Tehran, Iran (2023). He holds another M.Sc. in Data Science and Engineering from Politecnico di Torino, Italy (2022). His current research interests include self-supervised learning, deep learning, theoretical neuroscience, and reinforcement learning.

Eda Dagdevir received her M.Sc. and Ph.D. degrees in Biomedical Engineering at Erciyes University, Turkey, in 2017 and 2021, respectively. She was a research assistant at Erciyes University between 2015–2021. She is working as an assistant professor at the Vocational School of Technical Sciences, Kayseri University, Turkey, since 2021. Her main research areas are biomedical signal processing, brain–computer interface systems, and neuromarketing.