Neural Computing and Applications (2023) 35:14681–14722
https://doi.org/10.1007/s00521-021-06352-5

S.I.: MACHINE LEARNING FOR BIG DATA ANALYTICS IN SMART HEALTHCARE SYSTEMS

Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: a review

Hamdi Altaheri[1,4] · Ghulam Muhammad[1,4] · Mansour Alsulaiman[1,4] · Syed Umar Amin[1,4] · Ghadir Ali Altuwaijri[1,2,4] · Wadood Abdul[1,4] · Mohamed A. Bencherif[1,4] · Mohammed Faisal[3,4]

Received: 23 February 2021 / Accepted: 18 July 2021 / Published online: 25 August 2021

Abstract
The brain–computer interface (BCI) is an emerging technology that has the potential to revolutionize the world, with numerous applications ranging from healthcare to human augmentation. Electroencephalogram (EEG) motor imagery (MI) is among the most common BCI paradigms that have been used extensively in smart healthcare applications such as post-stroke rehabilitation and mobile assistive robots. In recent years, the contribution of deep learning (DL) has had a phenomenal impact on MI-EEG-based BCI. In this work, we systematically review the DL-based research for MI-EEG classification from the past ten years. This article first explains the procedure for selecting the studies and then gives an overview of BCI, EEG, and MI systems. The DL-based techniques applied in MI classification are then analyzed and discussed from four main perspectives: preprocessing, input formulation, deep learning architecture, and performance evaluation. In the discussion section, three major questions about DL-based MI classification are addressed: (1) Is preprocessing required for DL-based techniques? (2) What input formulations are best for DL-based techniques? (3) What are the current trends in DL-based techniques? Moreover, this work summarizes MI-EEG-based applications, extensively explores public MI-EEG datasets, and gives an overall visualization of the performance attained for each dataset based on the reviewed articles. Finally, current challenges and future directions are discussed.

Keywords Deep learning · Electroencephalogram (EEG) · Motor imagery (MI) · Brain–computer interface (BCI) · Classification · Survey

1 Introduction

Recent advancements in smart healthcare sensors and communication technologies have transformed the healthcare domain in terms of new services, accuracy, availability, and response time and are also producing enormous amounts of medical data [1, 2]. Electroencephalogram (EEG) sensors measure biometric data from the human brain, which can be decoded to understand underlying physical and psychological status and then utilized to further enhance quality of life. Unlike general smart healthcare sensors, EEG brain signals are utilized by smart healthcare systems in two ways: to import healthcare-related information, i.e., sensing, and to interact with the physical world, i.e., control, using smart equipment such as a wheelchair or an exoskeleton, etc. [3]. This sensing and control interaction using brain signals, known as the brain–computer interface (BCI), is relevant to critical healthcare applications such as post-stroke rehabilitation. EEG-based motor imagery (MI) signals have been employed in several healthcare applications, such as neurological rehabilitation [4, 5], restoring lost or impaired limb function by controlling a prosthesis or exoskeleton [6, 7], replacing walking function with a robotic wheelchair for people who cannot walk [8–10], and speller and cursor control [11, 12]. MI-EEG signals, however, are complex and have a high-dimensional structure. Therefore, advanced machine learning and deep learning (DL) algorithms are required to process and decode such complex brain data.

Extended author information available on the last page of the article

Traditional machine learning approaches have been widely used to classify MI-EEG data. Traditional methods typically process the MI-EEG signal in three main steps: preprocessing, feature extraction, and classification. The preprocessing step consists of several operations, such as channel selection (selecting the most valuable EEG channels for MI tasks), signal filtering (selecting the most valuable frequency range for MI tasks), signal normalization (normalizing each EEG channel around the time axis), and artifact removal (removing noise from MI-EEG signals). The most-used method for artifact removal is independent component analysis (ICA) [13, 14]. In feature extraction, various techniques for extracting task-related MI features from high-dimensional EEG signals have been proposed. The MI features fall into three categories, depending on the domain in which the data are processed: temporal features, spectral features, and spatial features. Temporal features are extracted in the time domain at different time points or during different time segments, such as mean, variance, Hjorth parameters, and skewness [15]. Spectral features include either frequency-domain features, such as power spectral density (PSD) and fast Fourier transform (FFT) [16], or time–frequency features such as short-time Fourier transform (STFT) and wavelet transform (WT) [17, 18]. Spatial features aim to identify features from specific electrode locations on the scalp, such as common spatial patterns (CSPs) [19]. CSP and its derivatives are the most common feature extraction methods for MI-EEG data [20–24]. Several researchers have attempted to expand and improve the CSP method. Sparse CSP [25] uses a regularization feature to add sparsity to CSP values. Stationary CSP [26], divergence CSP [27], and probabilistic CSP [28] are some of the other techniques that have been attempted to enhance CSP functionality. Filter bank CSP (FBCSP) [23] is another extended version of the CSP approach that uses the spatial information in EEG channels, as well as the frequency data in MI-EEG signals. FBCSP showed the best performance on MI classification out of all the other methods that rely on manual feature extraction. In the classification phase, several classifiers were used to classify the derived MI features into separate MI tasks, such as the naive Bayesian classifier [23], linear discriminant analysis (LDA) [24], support vector machine (SVM) [17], and extreme learning machine (ELM) [20, 21].

Although there has been considerable improvement in conventional MI-EEG signal classification methods, these methods are still plagued by major difficulties. First, the EEG signals are easily affected by many sources of noise including biological artifacts (e.g., heartbeats, eye blinking, tongue and muscle movements, concentration level, respiration, and fatigue), electronic equipment (e.g., wireless devices, cell phones, and computers), and environmental noise (e.g., sounds and lighting). These artifacts, combined with channel correlation, subject dependency, and the high dimensionality of EEG signals make interpretation and classification of brain signals a difficult task [29]. Therefore, it is critical to develop a more stable and generic MI-EEG BCI framework that can operate in a variety of scenarios and automatically derive distinctive features from challenging MI-EEG data. Second, the EEG signals have very low signal-to-noise ratios (SNRs), time-dependent covariates, and are nonstationary. Due to the time complexity of conventional preprocessing and feature extraction approaches, and the possibility of information loss, low SNR cannot be easily solved by conventional MI-EEG classification methods [30]. Third, the extraction of features relies strongly on human experience in a particular area. For instance, basic biological expertise is essential for analyzing the state of MI tasks through EEG signals. Although human experience can assist in some respects, in more general circumstances, it is insufficient. Therefore, an automated approach to feature extraction is required.

In the past five years, DL methods have been used to address the difficulties involved in classifying MI-EEG signals. Unlike conventional machine learning approaches, DL can automatically learn high-level and latent complex features from raw MI-EEG data using deep architecture, while eliminating the need for preprocessing and time-consuming feature extraction. Deep learning also achieves excellent results and scales well with the size of the training data. Several DL models have been successfully employed in different fields, such as computer vision [31], speech analysis [32], and medical diagnosis [33–35], and they have achieved outstanding performance. Motivated by the great success of DL techniques in other fields, many researchers have used DL approaches to classify MI-EEG data.

Several review articles of varying scope have explored the field of brain signal classification. The article in [30] reviewed the broader field of noninvasive brain signal classification using deep learning techniques. Other reviews have focused on EEG brain signals. Lotte et al. [36] and Rashid et al. [29] explored EEG-based brain signal classification using machine learning techniques, while Craik et al. [37] specifically investigated the DL-based techniques. In studies of narrower scope, Padfield et al. [38] and Aggarwal et al. [39] reviewed the machine learning techniques for MI-EEG classification. Apart from classification-based techniques, other articles have reviewed specific machine learning strategies and techniques for EEG brain signals, such as feature extraction [15], transfer learning [40], and data augmentation [41].

In this survey, we present a systematic review of the DL-based research on MI-EEG classification that has been carried out over the past ten years. We analyzed the

preprocessing strategy, input formulation, deep learning strategy, network structure, and performance evaluation of the collected peer-reviewed publications. Three main questions about DL-based MI classification are discussed in this review: (1) Is preprocessing required for DL-based techniques? (2) What input formulations are best for DL-based techniques? (3) What are the current trends in DL-based techniques? In addition, the MI-EEG-based applications and public MI-EEG datasets have been extensively reviewed. Finally, the paper ends with a discussion of current challenges and highlights some future directions.

2 Method

The review in this paper uses the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) procedure [42] to select studies and narrow down the search space, as shown in Fig. 1. Using this procedure, three steps were performed sequentially. First, the studies of the past 10 years were searched on two databases (Web of Science and PubMed) using the following keywords: ("Deep Learning" OR "Deep Machine Learning" OR "Deep Neural Network*" OR "Deep Belief Network*" OR "Convolutional" OR "CNN" OR "Recurrent" OR "LSTM" OR "Boltzmann Machine") AND ("EEG" OR "Electroencephalography") AND ("MI" OR "Motor Imagery"). The search was performed on 3 October 2020 (and was updated with some recent papers during the revision process). Duplicates between databases and unrelated studies were then screened out. After the papers were screened for relevance, full-text papers were assessed for eligibility according to the following constraints, which define the scope of the survey:

1. Electroencephalography only—Research with only EEG signals (without combining them with other signals, e.g., functional near-infrared spectroscopy (fNIRS), electrooculography (EOG), or electromyography (EMG)).
2. Motor imaginary only—classification of motor imaginary tasks only.
3. Deep learning—only research that using deep learning-based approaches, i.e., neural networks with at least two hidden layers.
4. Time—this survey focuses on studies from the past 10 years.

After applying the PRISMA procedure, 89 studies were selected for this survey. The temporal distribution of these studies is shown in Fig. 2. Although the search scope in this review includes papers published in the past ten years, studies that explored MI-EEG classification using deep learning (following PRISMA procedure) only started in the past five years as shown in Fig. 2. This figure also shows that the number of studies increased rapidly starting in 2017.

The MI-EEG datasets were collected from BCI-specialized websites and the specialized dataset search engine by Google using the following keywords: ("dataset"* OR "database"*) AND ("EEG" OR "Electroencephalography") AND ("MI" OR "Motor Imagery"). Additionally, we included public datasets mentioned in the articles reviewed in this survey.

2.1 Extracted data

We collected the following data from the articles and dataset sources:

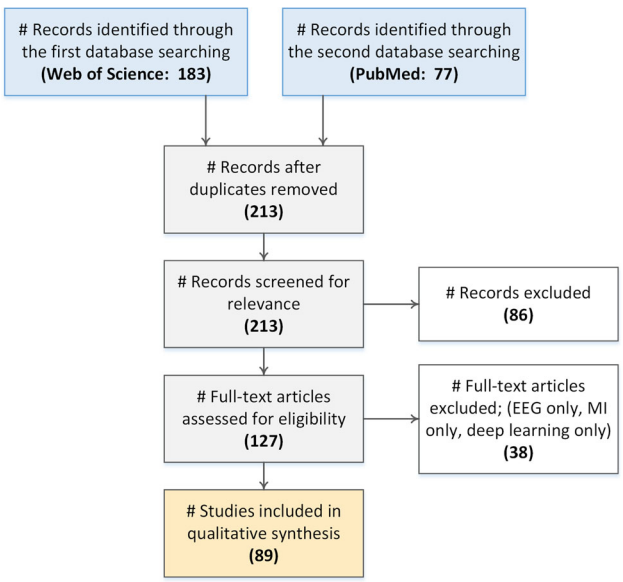

Fig. 1 Diagram of article selection based on the PRISMA procedure

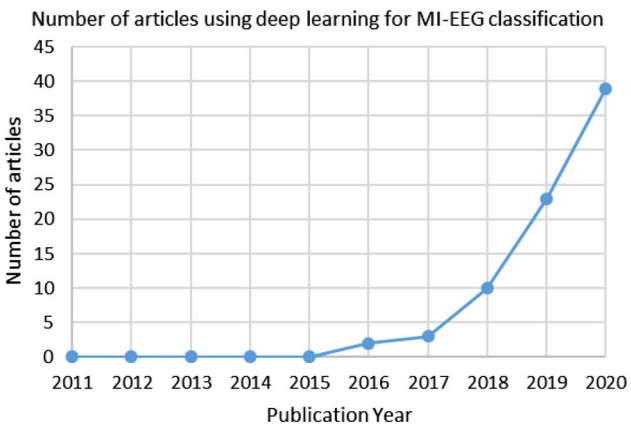

Fig. 2 Number of articles using deep learning to classify MI-EEG signals in the past ten years

1. Preprocessing strategy
   a. Artifact removal approach
      - Automatic removal
      - Manual removal
      - Without removal
   b. Analyzed frequency band
   c. EEG channel selection
2. Input formulation
   a. Extracted features
   b. Spectral images
   c. Raw signal values
   d. Topological maps
3. Deep learning approach
   a. General strategy
      - Discriminative models
        - Convolutional neural network (CNN)
        - Recurrent neural network (RNN)
        - Multi-layer perceptron (MLP)
      - Representative models
        - Auto-encoder (AE)
        - Restricted Boltzmann machines (RBMs)
        - Deep belief network (DBN)
      - Generative models
        - Generative adversarial network (GAN)
        - Variational autoencoder (VAE).
      - Hybrid models
   b. Architectures: number of hidden layers, type of hidden layers, activation.
4. Performance Evaluation
   a. Training approach:
      - Within-subject, cross-subject
   b. Evaluation approach:
      - Subject/session—dependent/independent,
   c. Evaluation strategy:
      - Hold-out, cross-validation.
   d. Performance metrics:
      - Accuracy, kappa, others.
5. MI Datasets, the variables listed below are defined:
   a. *General* name, year, key features, documentation link, download URL, and citation reference.
   b. *Tasks* number and type of EEG (MI/motor/non-motor) and non-EEG classes, number and type of non-task-related EEG data (Rest/Noise/Other).
   c. *Data* #subjects (M and F), #trials (total/per subject/per class), # sessions, duration of the session, rest between sessions, #runs per session, duration of runs, rest between runs, #trials per run, #MI in a trial (one-MI per trial, or several/long MIs per trial), and trial duration [Before | MI | After].
   d. *Software/Equipment* recording software, equipment name, #electrodes (type), sampling rate, frequency band, and voltage resolution.
   e. *Validation strategy* signal quality validation (during recording), and data validation (EEG signal analysis).

3 MI-EEG-based BCI

3.1 Brain–computer interface (BCI)

BCI, also known as neural control interface (NCI) or brain–machine interface (BMI), is a system that interprets brain activities and converts them into commands to control smart equipment, such as a wheelchair, drone, robotic arm, and virtual reality device. A BCI system includes three basic components, as shown in Fig. 3: recording equipment that measures magnetic, electrical, or metabolic brain activities; a processing unit that interprets brain signals, extracts key features, classifies them into specific brain tasks, and outputs the relevant commands; a smart equipment, an external device, or a computer running the produced commands. In this section, the first component of the BCI system, including the recording equipment and the characteristics of EEG and MI signals, will be discussed. The MI BCI processing unit is discussed from a deep learning perspective in Sect. 4. Later in this paper, in Sect. 6, the smart equipment used in the literature with MI-EEG signals is investigated.

3.2 Brain signal recording techniques

Mental activity in the central nervous system (CNS) produces continuous patterns that vary over time, known as neural oscillations or brain waves. During a mental activity, neurons in the brain communicate with each other, causing a change in the electrical current and blood flow in the brain, which can be measured using various techniques. Brain current can be measured using electrical and magnetic fields, while cerebral blood flow can be measured using optical and magnetic properties, as shown in Fig. 4.

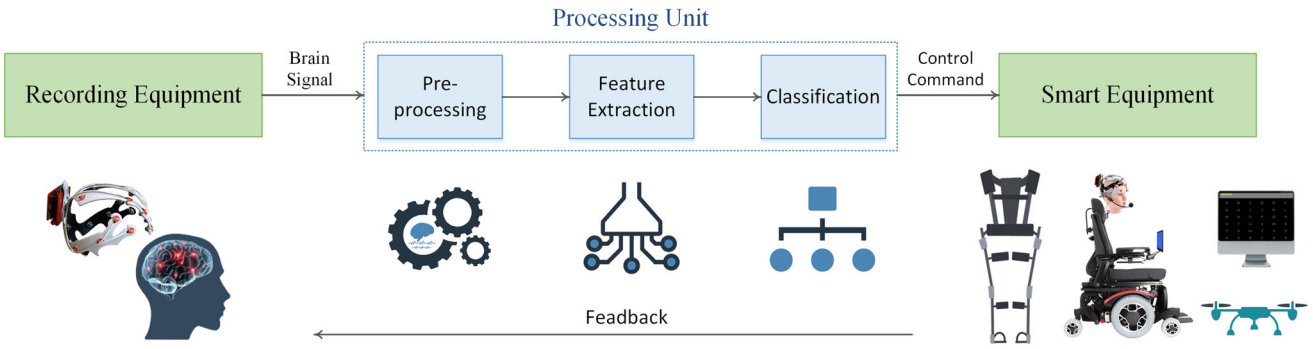

Fig. 3 Basic components of a BCI system

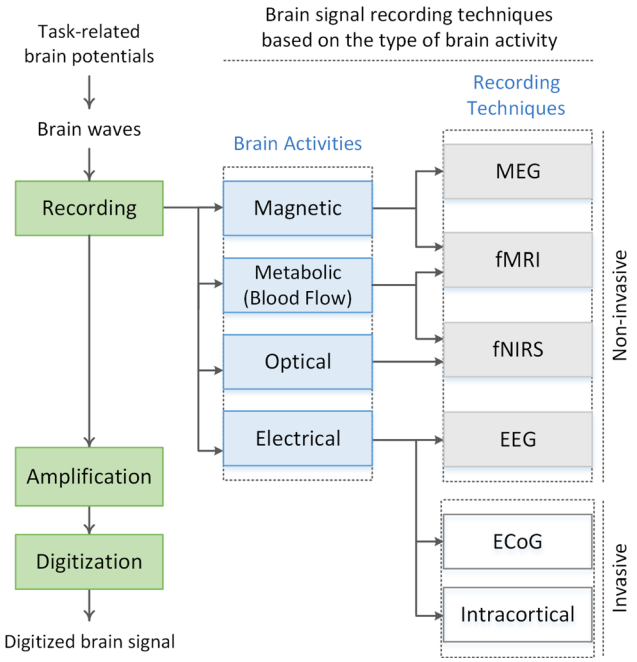

Fig. 4 A taxonomy of brain signal acquisition showing different brain signal recording techniques based on the type of brain activity

The choice of recording equipment to acquire the brain signal is determined by many factors, such as the field of application, the cost, and the community that will use it. Based on the recording equipment, BCI systems can be broadly divided into invasive and noninvasive BCIs. Invasive recording measures the electrical activity of a brain signal using implanted electrodes (sensors) placed under the skull (e.g., electrocorticography (ECoG)) or placed directly into the cortex (e.g., intracortical signals). Invasive BCI has very high spatial and temporal resolution, providing more precise information about brain activities. However, it requires a complicated setup (which includes surgery) and expensive equipment. On the other hand, a noninvasive recording method is usually carried out by placing multiple electrodes on the human scalp. With this method, the brain signal can be measured using magnetic activity (e.g., magnetoencephalography (MEG)), magnetic-based metabolic activity (e.g., functional magnetic resonance imaging (fMRI)), optical-based metabolic activity (e.g., fNIRS), or electrical activity (e.g., EEG), as shown in Fig. 4. In general, for noninvasive BCIs, EEG is used extensively and is preferred over other recording techniques due to its ease of use, portability, low cost, and high temporal resolution [43].

3.3 EEG signals

EEG is a technique for recording electrical brain activities using a noninvasive electrophysiological method that measures voltage fluctuations induced by the ionic current within brain neurons [44]. Because the ionic current produced inside the brain is recorded on the scalp, obstructions (such as the skull) significantly reduce the quality of the signal. The recorded EEG signal is only about 5% of the actual brain signal [45]. Therefore, to improve signal quality, raw EEG signals are normally preprocessed before feature extraction and classification.

EEG signals typically consist of a 2D matrix of real values (channel and time) that represent task-related brain potentials [46]. These two dimensions represent the spatial and temporal information of the EEG signal. The spatial resolution refers to the spatial positions of the electrodes on the scalp (number of electrodes), while the temporal resolution represents the number of time points per second (i.e., sampling rate). The spatial resolution ranges from 1 to 256 electrodes; however, for research or clinical purposes, a range of 21 to 64 electrodes is typically used. The sampling rate of EEG signals typically ranges from 128 to 1000 Hz. A sample of a 23-channel EEG signal sampled at 256 Hz is shown in Fig. 5. The electrodes are placed in fixed locations on the scalp, as shown in Fig. 6.

3.4 Motor imagery (MI) paradigm

Motor imagery is the process of imagining the movement of a part of the human body (e.g., a limb) without

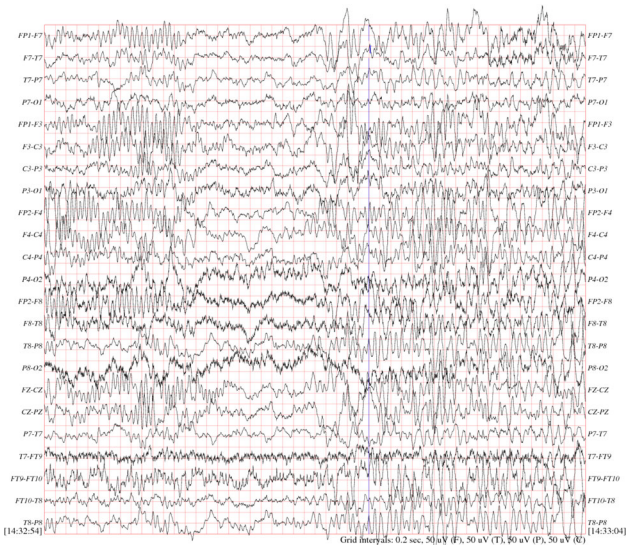

Fig. 5 Sample of 23-channel EEG signal recorded for 10 ms and sampled at 256 Hz with 16-bit resolution [47]

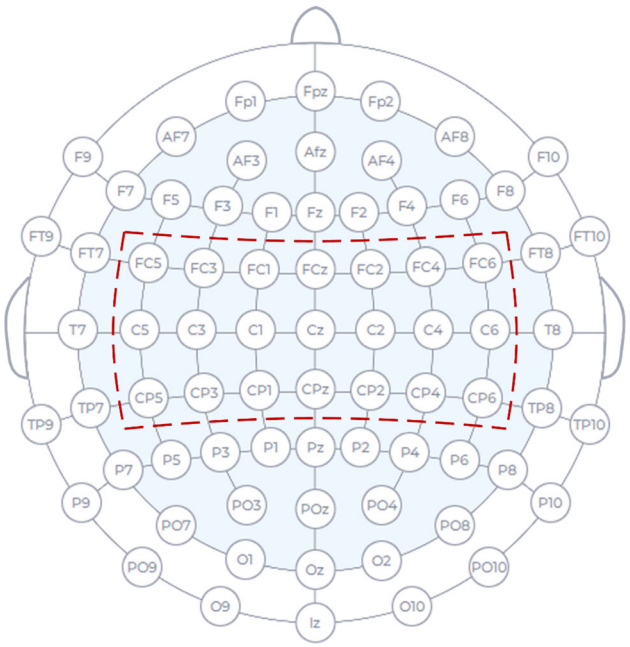

Fig. 6 Locations of 74 electrodes on the scalp using the standard 10–20 system. The names of the electrodes are based on their location: C (central), T (temporal), F (frontal), Fp (prefrontal), P (parietal), and O (occipital). AF, FC, FT, CP, TP, and PO are intermediate electrodes between (C, T, F, Fp, P, and O). Dashed lines mark the electrodes in the sensorimotor cortex

physically moving it [48]. BCI systems are mainly used in three paradigms: motor imagery and two of the event-related potential (ERP) (P300 and steady-state evoked potential (SSEP), specifically visually evoked potential (SSVEP)) [49, 50]. Among them, the MI paradigm is less susceptible to timing and depends on voluntary modulation rather than external stimuli, which is necessary for developing an independent BCI framework [48].

According to neurophysiological research of the sensorimotor cortex, mu (8–12 Hz) (the mu band is the alpha band that recorded from the sensorimotor cortex [51]) and beta (18–26 Hz) rhythms are modified by actual movement, preparation for movement, and even imagining of movement (known as motor imagery (MI)) [52]. The energy modulation of brain rhythms in a particular frequency range as a result of an event is known as event-related desynchronization (ERD)/event-related synchronization (ERS). ERD reflects a power reduction in mu/beta rhythms in the sensorimotor cortex during MI events, while ERS denotes a rise in power that normally happens after the MI event. ERD/ERS with mu/beta rhythms exhibits distinct spatial patterns during MI tasks related to various human body parts, e.g., right hand, left hand, leg, and tongue. Several experiments have shown that people can learn to use motor imagery to control (decrease or increase) the power of sensorimotor rhythms [53]. Therefore, MI provides a good paradigm for developing an independent BCI system depending on sensorimotor rhythms.

4 Deep learning-based approaches applied in MI classification

In this section, we investigate the deep learning approaches used in MI classification from four main perspectives: preprocessing method, input formulation, deep learning architecture, and performance evaluation.

4.1 Preprocessing

In order to extract valuable MI components from EEG signals, preprocessing is generally performed in three main steps: channel selection, signal frequency filtering, and artifact removal. In channel selection, a subset of MI-EEG data is selected from a set of EEG electrodes that contain the most distinct MI features, helping to reduce system complexity, computational time, equipment cost, and possibly system performance. More than 79% of the studies reviewed used all EEG channels in the datasets, while 8 studies investigated the effect of channel selection on the accuracy of MI classification with different numbers of electrodes [51, 54–60]. The analysis of these studies is detailed in the discussion section.

Signal frequency filtering was used as a preprocessing step for MI classification in the majority of the studies (91%) for two reasons: selecting the most valuable frequency bands for MI tasks and eliminating artifacts. For MI-EEG signal, the ERD/ERS of sensorimotor rhythms mainly occurs in μ (8–12 Hz) and β (18–26 Hz) frequency

bands. Therefore, almost all the studies reviewed in this survey (96%) included these two frequency bands in their analysis, as they contain the most distinctive features related to MI activities. With frequency filtering, a large portion of the noise can also be eliminated, such as low-frequency artifacts (e.g., EOG, caused by eye blinking) and high-frequency noise (e.g., EMG above 35 Hz) [61]. For this reason, 47% of the studies suggested using frequency bands in the range of 6–35 Hz. However, artifacts cannot be easily excluded using band-pass filters, as they may interfere with the effective ERD/ERS bands. Several other studies (35%) suggested using a wider frequency band than 6–35 Hz, in the range of 0–40 Hz, as shown in Fig. 7. In this review, frequency filtering was not considered as a method for artifact removal, as it was used in 91% of the reviewed studies. The best selection of frequency bands is explored in the discussion section.

Previous literature [13, 14, 61] investigated methods for identification and elimination of MI-EEG artifacts, which will not be repeated in this paper. Outside of studies that did not specify any artifact removal procedure, we identify three main strategies for removing artifacts in the reviewed papers: automatic removal (20%), manual removal (4%), without artifacts removal (40%), as shown in Fig. 8. Most of the reviewed studies classified the MI-EEG signal without any artifact removal based on the fact that deep learning is able to extract useful features from raw and unfiltered data. Other studies employed artifact removal approaches before inputting the MI data into deep learning models. The most common method used in the reviewed papers was ICA [7, 62–64] and common average reference (CAR) [60, 65–67]. Some studies [66, 68] used more advanced tools to remove MI signal artifacts, such as the automatic artifact removal (AAR) toolbox [69].

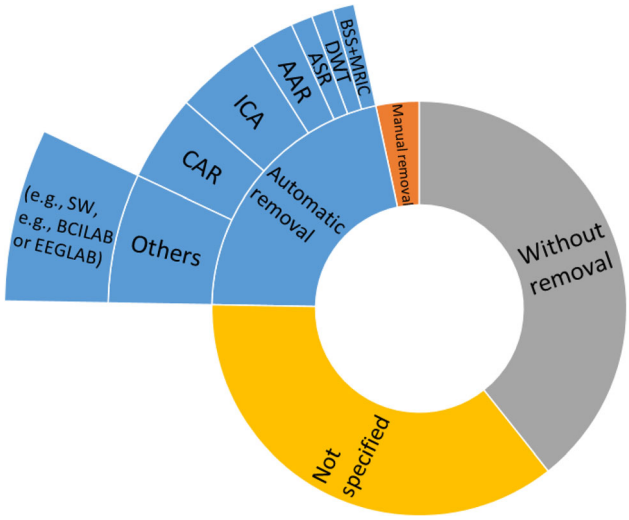

Fig. 8 Artifact removal strategies for the MI-EEG signal and their percentage in the reviewed studies

4.2 Input formulation

The input formulation of the EEG signal in the deep learning models can be categorized into four types: extracted features, spectral images, raw signal values, and topological maps. The choice of input formulation was largely dependent on the architecture of the deep learning model. Figure 9 shows the input formulations used in the reviewed articles. Figure 10 presents a taxonomy of the different input formulations used by deep learning methods for MI-EEG signal classification.

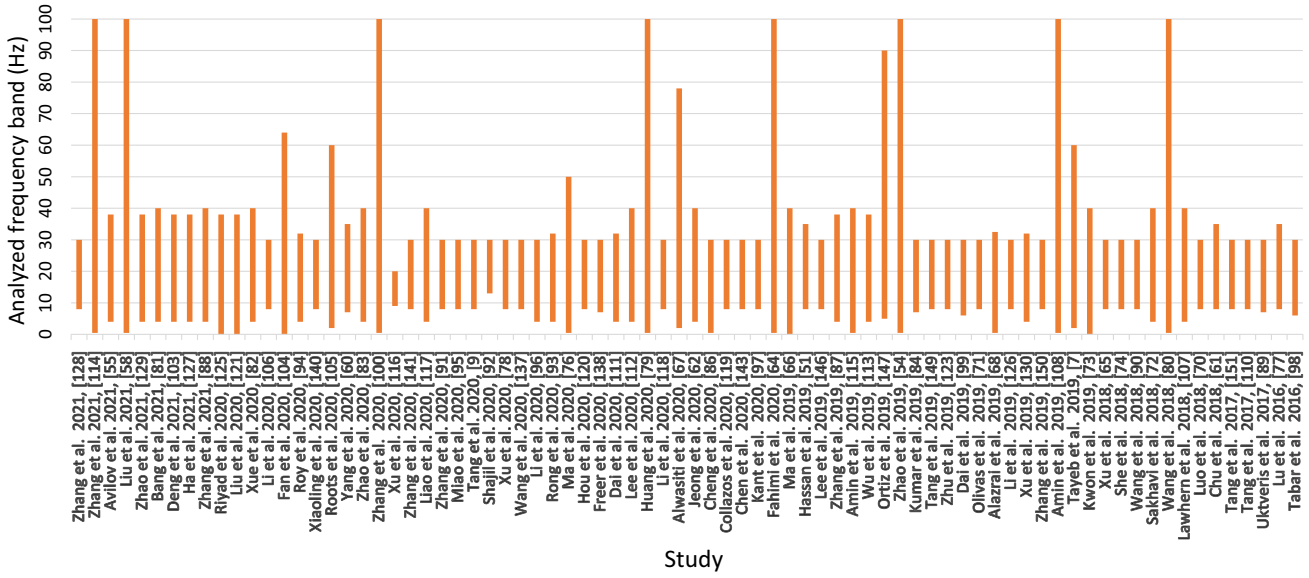

Fig. 7 Range of frequency bands analyzed in the reviewed studies

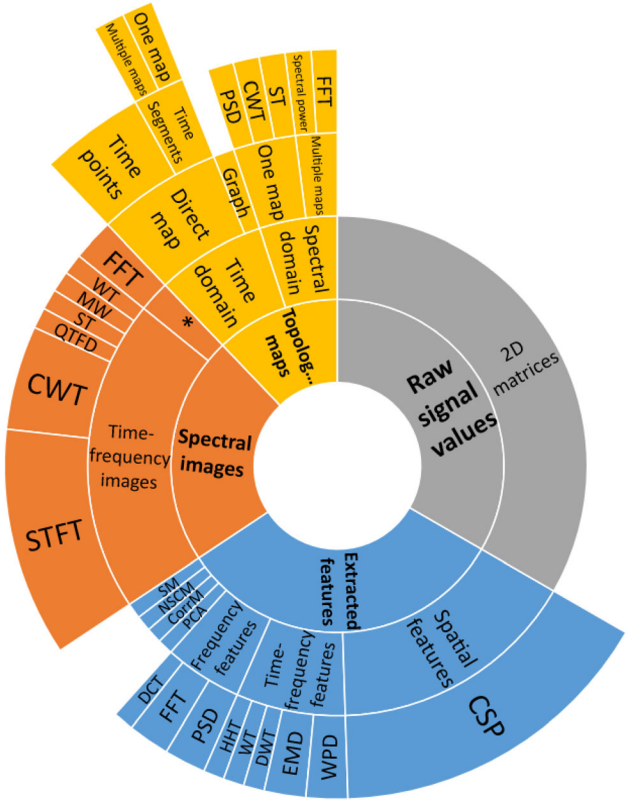

Fig. 9 Proportional representation of each input formulation in the reviewed articles. (*Spatial-frequency images)

4.2.1 Extracted features

In the feature-based input formulation, the process of MI classification is performed in two steps. First, conventional feature extraction approaches translate EEG signals into vectors. Then, the feature vectors are entered into a deep learning model that trains to classify the data associated with those features. The most popular features extracted from MI-EEG data in the previous works were the CSPs [70–74]. Luo et al. [70] utilized FBCSP to extract spatial-frequency-sequential time slices from MI-EEG signals and classified them using long short-term memory (LSTM) and gated recurrent unit (GRU) models. The proposed method achieved good results for both recurrent models. In [71], the authors also used the FBCSP approach to extract spatial features from MI data and fed them into a CNN model as a 2D matrix. Other types of features have also been used for MI-EEG classification with DL-based methods including frequency features (e.g., FFT [51], discrete cosine transform (DCT) [75], and PSD [61, 76]), time–frequency features (e.g., wavelet packet decomposition (WPD) [77, 78], discrete wavelet transform (DWT) [76], empirical mode decomposition (EMD) [9, 75], and Hilbert–Huang transform (HHT) [79]), and temporal features (e.g., statistical measures [80]). Several DL models were proposed for MI classification using handcrafted features, such as CNN [66, 81–83], LSTM, [80, 84, 85], GRU [70], ELM [74], stacked autoencoder [51], DBN [61, 77, 86], and hybrid CNN/LSTM models [87, 88].

4.2.2 Spectral images

Spectral images generated from the EEG signal, such as spectrograms, have been used as an input formulation for several neural models, especially CNNs [7, 65, 89–93]. For spectral images, the MI-EEG signal was represented as time–frequency [7, 65, 94] (i.e., spectrogram, T (time window) $\times$ F (frequency)), or spatial-frequency images [89, 95] (i.e., C (channel) $\times$ F (frequency)), as shown in Fig. 10, images (f) and (e), respectively. Time–frequency images can be generated using WT [65, 96, 97], STFT [7, 90, 98, 99], Stockwell transform (ST) [67], and quadratic time–frequency distribution (QTFD) [68]. Spatial-frequency images are usually generated using FFT [89, 95].

For the time–frequency representation, MI data from multiple channels can be fed into the DL model as separate images or combined into a single 2D or 3D image representing the MI trail. The combined 2D images were generated in three ways: by concatenating the EEG channels with frequency values in the same dimension (frequency-channel axis) [$T \times (F + C)$] [68, 90–93, 98–100]; in the time dimension (time-channel axis) [$(T + C) \times F$] [97]; or in both the time and frequency dimensions (the time-channel and the frequency-channel axis) [$(T + C) \times (F + C)$], e.g., based on the topology of the electrodes [67], as shown in Fig. 10, images (h), (i), and (j), respectively. MI-EEG data from different channels can also be represented in a third dimension to form a 3D tensor (i.e., [T: time, F: frequency, C: channel]) [65, 96], as shown in Fig. 10g.

Tayeb et al. [7] used STFT to translate MI-EEG data into time–frequency images, i.e., spectrograms. The authors used local and public [101] MI datasets with two classes and three EEG channels. Spectral images were extracted from the three EEG channels in the 2–60 Hz frequency range and fed to CNN and hybrid CNN/RNN models. Another study [98] also used STFT to transform the EEG data into time–frequency images. The authors utilized the EEG features in the mu (8–13 Hz) and beta (13–30 Hz) bands to classify MI signals using a hybrid CNN/SAE model. The study used two public MI datasets (BCI-C II-3 [102] and BCI-C IV-2b [101]) with two MI classes and three EEG channels. The three EEG channels were combined with the two frequency bands in a single dimension of the spectral image. This input formulation combines the temporal, spectral, and spatial information of MI data in a 2D image of size [$T \times (F + C)$]. In another study [52], the authors concatenated the continuous

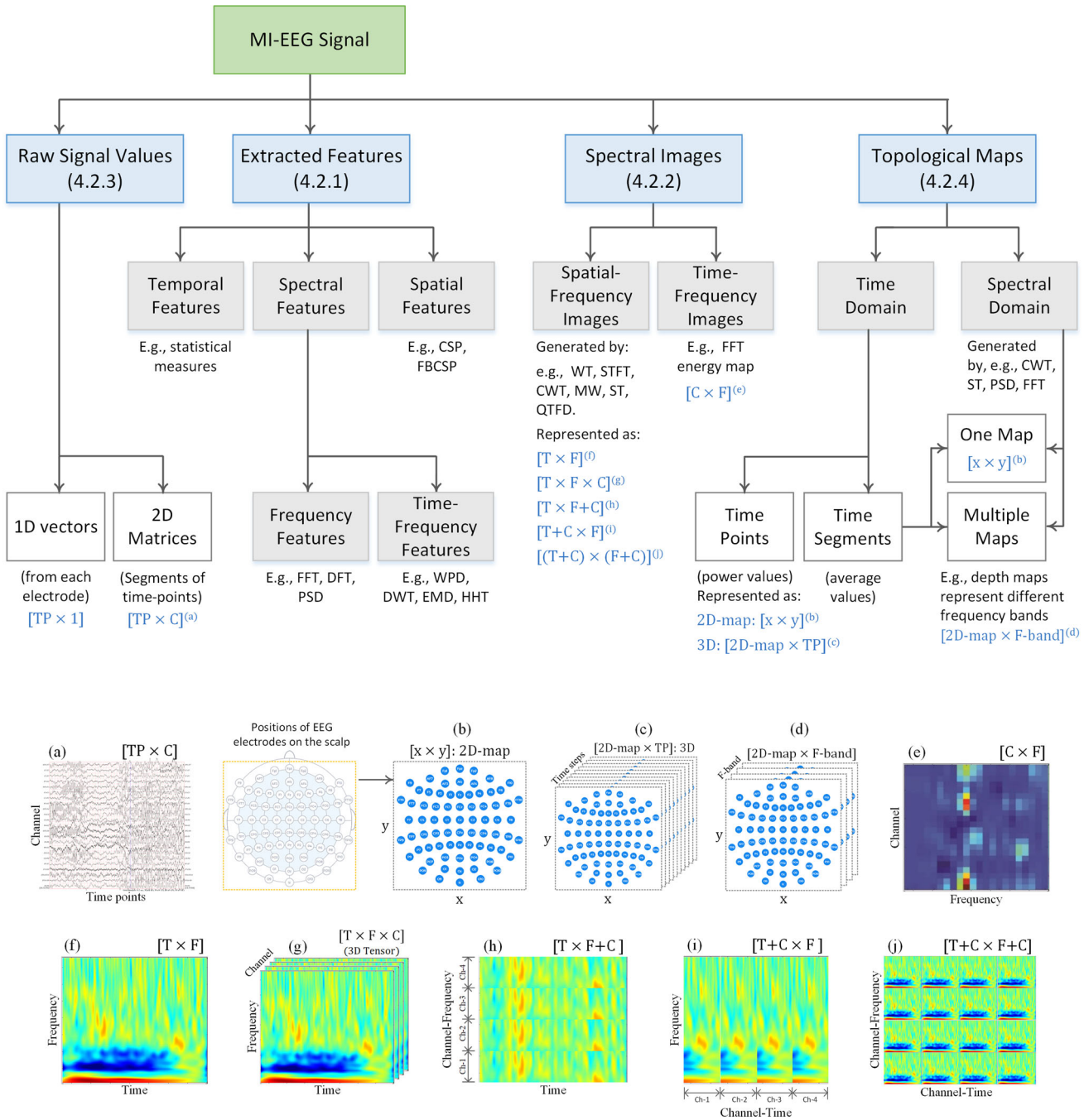

Fig. 10 Taxonomy of the different input formulations used by deep learning methods for MI-EEG signal classification. Images **a** through **j** show different image representations of the MI-EEG signal used by deep learning approaches. **a** Raw MI-EEG signal; **b–d** topological images; **e** spatial-frequency image; **f–j** time–frequency images. *T*: Time window (time segment); TP: Time point (sampling point) or time step; *F*: Frequency; F-band: frequency band; *C*: channel (electrode); *x*, *y*: the *x* and *y* coordinates on the scalp

wavelet transform (CWT) spectrum of two channels along the time axis to form 2D images of size $[(T + C) \times F]$. However, this structure may not be suitable if the processed MI signal is composed of a large number of channels, i.e., the size of the produced spectral image will be extremely large. In this case, the channels can be represented in a separate dimension rather than combined with the frequency or time values in the same dimension. The study in [65] used the WT to represent MI-EEG signals as time–frequency images. The authors selected three EEG channels for MI classification and represented them as 3D tensors (time × frequency × channel), which were used as inputs for a CNN model.

4.2.3 Raw signal values

The raw EEG signal in the time domain, i.e., [TP (time point) × *C* (channel)] matrices, as shown in Fig. 10a, was also used directly as an input to the deep neural networks. This was motivated by the ability of deep learning models to learn complex features from large amounts of data without using handcrafted features, encouraging the concept of end-to-end learning. In this concept, the deep learning model should learn both the optimized feature representation and the classification of raw EEG data in a supervised way, without (or with limited) preprocessing. EEG signals from multiple electrodes were either processed separately as 1-D vectors or combined into a 2D matrix, as shown in Fig. 10. The 2-D matrices were represented either directly by dividing the EEG signal into segments based on time points, with each segment consisting of a [TP × *C*] matrix, or indirectly by converting each time point to a 2D image based on the spatial topology of electrodes, as will be explained in the next section. Several deep learning models have been proposed to classify MI tasks using raw EEG data and have achieved competitive outcomes [60, 62, 103–114]. The raw EEG signal was used as an input to DL models having a light architecture [107, 109, 110] as well as advanced and compact architectures [108, 115] with little or no preprocessing. The authors in [108, 115] used the time series of MI signals as raw data in the form of 2D matrices without preprocessing or artifact removal. The raw MI data were fed into multi-level CNN and hybrid CNN-MLP/AE models that achieved remarkable performance. Another study [107] reported good performance in classifying MI tasks using a light CNN architecture with raw MI signals and minimal preprocessing.

4.2.4 Topological maps

In the topological map input formulation, the EEG signal is represented as 2-D or 3-D images based on the spatial topology of the electrodes, i.e., the position of electrodes on the scalp. Topological maps can be constructed from EEG signals in the time domain [54, 100, 116–118] or frequency domain [67, 119], as shown in Fig. 10. In the time domain, maps were generated either from each time point (i.e., the sampling point) [54, 58, 116, 117] or from a segment (window) of several time points [59, 118]. The studies in [116, 117] dealt with each 2-D map extracted from each time point as separate samples, similar to image (b) in Fig. 10, while in [54, 58], the 2-D maps from each time point were combined to represent a 3-D map [2-D map × Time point], as illustrated in the image (c) in Fig. 10. The researchers in [118] proposed a 3-D topological map [2-D map × *F*: frequency], similar to image (d) in Fig. 10, using the Clough–Tocher interpolation algorithm. The proposed method averages the time-domain powers of MI-EEG signal segments from three frequency bands. The data from each frequency band were represented in the frequency dimension (F) of the 3-D map. In another study [100], a graph-based structure was proposed to represent the electrode positions of the MI-EEG signal. In frequency-domain-based topological maps, the MI-EEG signal is first converted to the frequency domain and the topological maps are created by averaging the spectral values. The research in [119] used CWT and PSD to convert the MI-EEG signal to the frequency domain, and then, the spectral values were used to construct topological maps using spherical spline interpolation. Differing from all previous studies, the study in [67] suggested creating a large topological image by combining 2-D images (spectral images) extracted from each channel, rather than scalar values (single values), according to the positions of the electrodes in the scalp.

4.3 Deep learning architectures

In this section, we investigate the different DL architectures employed in MI-EEG classification research. DL models are categorized according to their function into four subcategories [30]: discriminative, representative, generative, and hybrid DL models, as shown in Fig. 11.

4.3.1 Discriminative DL models

Discriminative DL models refer to DL architectures that can learn distinct features from input signals using nonlinear transformations and classify them using probabilistic prediction into pre-defined classes. Therefore, these techniques can be used for both feature extraction and classification. Discriminative models include CNN, RNNs (and their variations, GRU and LSTM), MLP, and ELM.

A CNN is one of the most common models for deep learning that specializes in extracting local and spatial patterns. The CNN architecture consists of a group of neural networks arranged in a particular order with layers of different sizes, where each layer performs a particular task. The earlier layers learn low-level features, while the deeper layers learn high-level features. CNNs typically consist of three structure blocks: convolutional layers (for feature extraction), pooling layers (for feature dimensionality reduction), and fully connected (FC) layers (for classification). A convolutional layer is an essential component of the CNN architecture that performs feature extraction. A pooling layer provides a typical downsampling operation that reduces network computation. The output feature maps of the pooling layer are typically

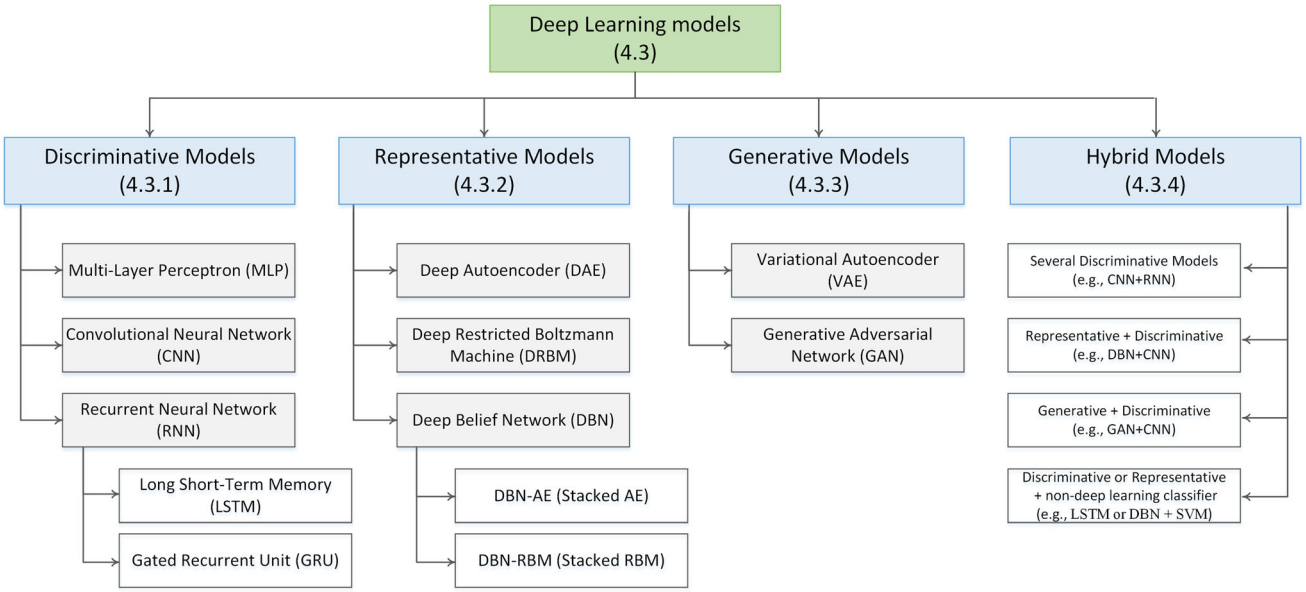

Fig. 11 Taxonomy of the deep learning models applied to MI-EEG classification

flattened and are connected to one or more fully connected layers.

CNN-based DL strategies were used in 78% of the studies reviewed (including standalone and hybrid CNNs), as shown in Fig. 12. Several studies have used CNN models to classify MI-EEG-based signals using standard CNNs with light [96, 107] and deep architectures [118, 120], as well as many other CNN varieties, including, as listed in Table 1, attention-based CNN [104, 106, 121, 122], residual-based CNN [58, 104, 123, 124], inception-based CNN [9, 114, 125],

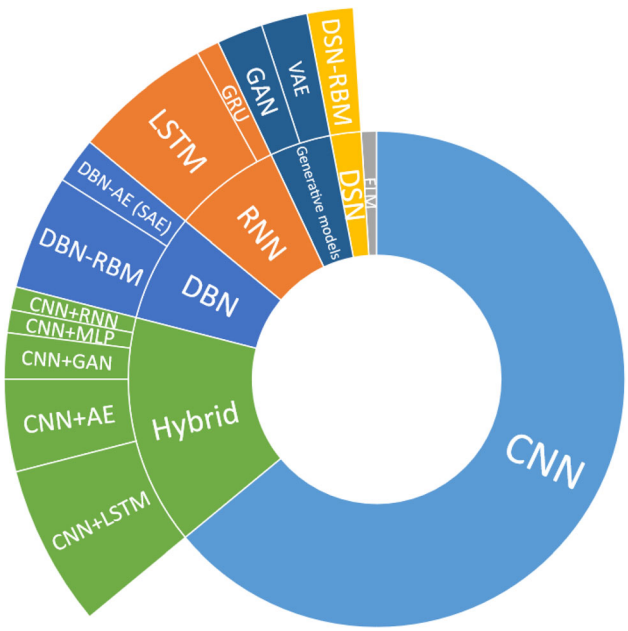

Fig. 12 Deep learning methods across all studies

Dense Net [67], 3D-CNNs [54, 58], multi-branch CNNs, i.e., based on ensemble learning, [54, 58, 66, 73, 75, 105, 112], multi-layer CNNs [62, 82, 108, 111, 126], multi-scale CNNs [9, 106], CNN with multi-level pooling [127], and CNN architectures with transfer learning ability [83, 97, 121, 128–130].

The study in [107] proposed a light CNN architecture with few parameters to classify different EEG tasks using raw signals and achieved remarkable performance. Contrarily, in [71], the authors presented a CNN model for multiple MI classification with features extracted from MI data using the FBCSP approach. In [65], a temporal-frequency image representation was proposed using a WT and coupled with a CNN model to classify MI signals, achieving an accuracy of 85.59% in a dataset with four MI classes [131]. The authors in [7] proposed a CNN model for MI-EEG classification using spectral images extracted from EEG signals. The proposed model achieved an accuracy of 84.24% and was successfully implemented in real-time robotic arm control. Amin et al. suggested a multi-layer CNN architecture with multilevel feature fusion for MI-EEG classification from raw EEG signals [108]. In this architecture, the features were extracted at different levels of convolutional layers and merged using FC layers. This approach achieved an accuracy of 74.5% using the BCI Competition IV-2a dataset [131]. Li et al. [118] proposed a very deep model based on CNN that used Fourier's EEG signal transformation and interpreted it as topological maps of the scalp. These maps were fed as images to the CNN model. In another study [67], the authors used a deep metric learning (DML) method with a triplet network architecture and DenseNet-based CNN

Table 1 List of the different CNN techniques used for MI classification

CNN architecture/technique	Study
Attention	[104, 106, 121, 122]
Inception	[9, 114, 125]
Residual	[58, 104, 123, 124]
DenseNet	[67]
Three-dimensional CNNs	[54, 58]
Multi-branch CNNs (ensemble learning)	[54, 58, 66, 73, 75, 105, 112]
Multi-layer CNNs	[62, 82, 108, 111, 126]
Multi-scale CNNs	[9, 106]
Multi-level pooling	[127]
CNN with transfer learning	[83, 97, 121, 128–130]

networks to classify MI-EEG signals. The proposed DML method yielded promising results with a limited number of training instances ($\sim$ 120 samples). In [54, 58], the authors proposed a multi-branch 3D CNN network for MI classification using a 3D topological representation of MI data. The multi-branch network consisted of three CNN blocks, each with different receptive field size, operating in parallel. The output of these blocks was then fed into a softmax layer to produce the final classification result. The researchers in [58] found that the CNN network with three branches performed better than the CNN network with one or two branches. Several other researchers have employed CNN models to classify MI-EEG signals, including Wang et al. [90], Dose et al. [109], and Tang et al. [110], and reported accuracies of 92.7, 80.4, and 86.4%, respectively.

RNN is a deep learning architecture primarily applicable to time-series data. RNN network can extract temporal features and patterns from sequential data, making it a powerful approach in video, speech, and medical signal analysis, e.g., EEG. Two RNN networks have been used extensively in the literature: LSTM and GRU.

The LSTM model is a type of RNN network that can learn long-term relationships and overcome the vanishing gradient problem of traditional RNNs. The LSTM model consists of LSTM cells (equivalent to nodes in the multilayer perceptron) that are controlled by three gates: the input, forget, and output gates. Stacked LSTM cells create an LSTM layer that can work in a forward or backward time direction to form a bidirectional or unidirectional LSTM.

LSTM models have been employed in some studies to classify MI tasks [7, 80, 84]. A deep LSTM model based on one dimension-aggregate approximation (1d-AX) approach was investigated in [80] to classify MI-EEG tasks. In another study [84], the authors proposed an LSTM model that uses CSP for feature extraction, LDA for feature reduction, and SVM as a classifier. The proposed model achieved accuracies of 68.19% and 82.52% in the public datasets GigaDB [132] and BCI-C IV-1 [133], respectively. Kumar et al. [85] also used an LSTM model combined with CSP and SVM for MI classification. The study proposed an adaptive method for frequency band selection using a genetic algorithm, achieving an average accuracy of 69.59% using a two-class dataset (Cho et al. [132]).

The GRU is a new generation of RNN, and it differs slightly from LSTM in that it has two gates (update and reset), instead of the three in LSTM, and they are connected somewhat differently. The GRU can be considered a simplified version of the LSTM with lightweight architecture. Only one study using the GRU for MI classification was found [70]. In this paper, the authors proposed GRU and LSTM networks with a sliding window cropping strategy (SWCS) to classify MI signals using spatial-frequency features extracted using the FBCSP approach. The researchers validated their models using two public datasets with four and two classes (BCI-C IV-2a [131] and BCI-C IV-2b [101]). The results showed that the GRU achieved the best results with accuracies of 73.6% and 82.8% for the first and second datasets, respectively, compared with 72.6% and 81.5% for LSTM.

An ELM is a type of discriminative feedforward neural network used for feature extraction and classification. In ELM, the learning process is almost instantaneous, as the hidden nodes are randomly assigned and do not need to be tuned or updated. ELM uses the best selection of randomly initialized neuron parameters to estimate the boundary of the required decision. This feature represents a major advantage over deep learning. However, the accuracy obtained by combining random weights is not as high as that obtained by a fine-tuned backpropagation neural network. Another advantage of ELM is that the activation function can be as complex as is desired because it does not need to be differentiable and trainable via backpropagation. ELM is usually used as a fast classifier with hand-designed features. The study in [74] proposed a semi-supervised method for multiclass MI classification using the deep architecture of the ELM. The authors presented two ELM models, one for feature extraction and one for classification

of MI tasks, using manually designed features based on the CSP approach. A hierarchical ELM (H-ELM) was first used to learn MI features, and then a semi-supervised ELM (SS-ELM) algorithm was employed to classify MI tasks in a semi-supervised manner. The results of the study showed that the proposed approach performed well in terms of accuracy and speed.

4.3.2 Representative DL models

Representative DL models refer to DL architectures that specialize in feature extraction in an unsupervised manner, which can be used for various tasks, such as clustering and classification. Representative DL models include deep AEs (D-AEs), deep RBMs (D-RBMs), and DBN.

An autoencoder (AE) is a type of representative artificial neural network used to learn features in an unsupervised manner with efficient data coding. AE composes of three main components: an encoder, code, and decoder. The encoder compresses the input into a latent-space representation known as the code, which is then used by the decoder to reconstruct the input. There are many varieties of AE, and this review identifies three general types that differ significantly in how they function: D-AE, SAE, and VAE. D-AE learns like normal AEs, where all layers in the network are trained at the same time, no matter how many layers in the network. In SAE, different stacked AE blocks are trained separately, where the representation of each block (code) is used as input for the next block. SAE is defined as an AE-based DBN (DBN-AE) and will be explored next with DBNs. VAEs, proposed in 2013 [134], differ from other AEs in that they have a layer of data means and standard deviations at their core, allowing for easy interpolation and random sampling. VAE is one of the most powerful generating methods. VAE is explored with generative DL models.

Typically, AEs, other than VAEs, are used for feature extraction; therefore, AE models are usually combined with other discriminative DL models to create hybrid models. For example, in [98], a hybrid CNN/AE model was proposed for MI task classification. Hybrid DL models will be addressed later in this section.

A DBN is a representative model composed of a series of RBM or AE networks [30]. Therefore, we separate the DBN into two parts: DBN-RBM (also known as stacked RBM), which is made up of RBM, and DBN-AE (also known as stacked AE), which is made up of AE. DBN-RBMs have been used in some research to classify MI-EEG signals [61, 77, 78, 86]. Lu et al. [77] proposed a deep DBN architecture based on stacked RBM layers for MI signal classification. They used WPD and FFT to train three RBMs, which was followed by an additional output layer to form a four-layer DBN. A decoding method for MI-EEG signals using the Lomb–Scargle periodogram (LSP) for feature extraction and an RBM-based DBN as a classifier was proposed in [61]. The LSP method was used to extract useful PSD features from incomplete (part of the data is lost) MI-EEG data with a high level of artifacts, reporting an average accuracy of 83%. In another research [78], the authors proposed a deep DBN-RBM model, enhanced by a t-distributed stochastic neighbor embedding (t-SNE), for feature extraction and SVM for classification. The study utilized WPD and CSP to extract the temporal-spectral and spatial features from MI-EEG data, respectively, achieving an accuracy of 78.51% using a dataset with four MI classes. Hassanpour et al. [51] proposed a stacked sparse AE model, defined as DBN-AE, for MI-EEG classification using FFT frequency features. The study used a sliding window augmentation approach to increase the number of training data and achieved 71% accuracy using the public BCI-C IV-2a dataset [131].

4.3.3 Generative DL models

Generative DL models are typically used to augment and improve training data. The most popular generative DL models are GAN and VAE. Several studies in this review used traditional data augmentation approaches, i.e., non-DL, to increase the size of training data, such as noise addition [114], sliding window [88, 105], and amplitude perturbation [135]. Two of the reviewed studies introduced DL-based data augmentation using GAN and VAE networks [64, 91]. The results from these studies showed that using GAN models for MI data augmentation significantly increased classification performance. Zhang et al. [91] proposed a four-layer GAN model for MI data augmentation and compared its performance with VAE and other traditional augmentation methods such as geometric transformation and noise addition. The results revealed that both GAN and VAE outperformed the traditional methods, while the GAN had the best performance. The study found that the performance of a CNN model trained on MI data augmented with GANs improved by 17% and 21% for the BCI-C IV-2a [131] and IV-2b [101] datasets, respectively, when compared to training without data augmentation. The authors in [64] introduced a GAN-based generative model with light architecture for MI data argumentation, showing that increasing the number of training samples improved the performance of the CNN model by 3.57% using the BCI-C III-4a dataset [136]. The study also demonstrated the superiority of GAN models over VAE. In another study [99], the authors proposed a hybrid DL model based on VAE and CNN. In this study, VAE was used as a classifier rather than a generative model. Hybrid models are discussed in the next section.

4.3.4 Hybrid DL models

Hybrid DL models integrate two or more DL models into a single network. In addition to the standalone deep learning models mentioned above, researchers have attempted to merge different deep learning networks, and encouraging results have been obtained for MI classification tasks [7, 63, 98–100, 115, 137, 138]. This review identifies five categories of combinations: two discriminative models (e.g., CNN/LSTM [56, 63, 88, 100, 137, 138], CNN/GRU [59], and CNN/MLP [115]), representative model combined with a discriminative model (e.g., CNN/SAE [60, 98]), generative model combined with a discriminative model (e.g., CNN/GAN [64, 91] and CNN/VAE [99]), discriminative model followed by a non-DL classifier (e.g., LSTM + SVM [85] and CNN + SVM [68, 75]), and representative model followed by a non-DL classifier (e.g., DBN + SVM [78]).

A hybrid CNN/RNN model, called recurrent convolutional neural network (RCNN), was proposed by the study in [7]. The model consisted of a single convolutional layer and four recurrent layers followed by a fully connected layer. The MI signal was converted to spectral images before being fed to the RCNN model. The performance of this model was studied using the authors' local dataset consisting of two MI classes and three channels, reporting accuracy of 77.72%. The study in [98] proposed a hybrid CNN/SAE architecture consisting of a 1-D convolutional layer followed by a six-layer SAE that received MI signals as 2D spectral images. The authors used two public datasets (BCI-C II-3 [102] and BCI-C IV-2b [101]) containing two MI classes with one and nine subjects, achieving 90.0% and 77.6% accuracy, respectively. The researchers in [115] suggested a fusion of multi-layer CNNs with AE and MLP networks. The multi-layer CNNs consisted of different CNN models that were trained on different frequency bands. The CNN features extracted from these bands were merged into a single feature representation and combined with AE and MLP models. The proposed CNN/MLP and CNN/AE models achieved 75.7% and 73.8% accuracy, respectively, on the public BCI-C IV-2a dataset [131]. In [99], a hybrid DL model based on CNN and VAE was proposed for MI classification using a temporal–spectral–spatial representation of the MI signal, reporting a mean kappa value of 0.56 in the BCI-C IV-2b dataset [101]. The study also examined the fusion of CNN and LSTM models.

Figure 12 shows a proportional representation of the deep learning methods in the reviewed studies. The key information extracted from the reviewed papers is summarized in Table 2.

4.4 Performance evaluation

For the MI EEG classification, accuracy is the most used measure of performance (95.4%). The Kappa metric is also a common performance measure (35.4%), which is useful in removing the impact of unbalanced data and random classification [99]. The confusion matrix has been used in many studies (24%) to provide additional information on classifier efficiency [62, 97, 112]. Several performance measures can be drawn from the confusion matrix such as precision (positive predictive value (PPV)) [67, 109], sensitivity (recall or true-positive rate (TPR)) [108, 110], specificity (true-negative rate (TNR)) [84, 148], and F-score [65, 68], which integrates TPR and PPV into a single metric using the harmonic mean. When a continuous parameter is used for classification, the area under the curve (AUC) [100, 107] and receiver operating characteristic (ROC) [63, 118] are often used [29]. Some researchers have also used statistical tests with performance measures such as p-value, t-test [91, 112, 137], and Wilcoxon test [127, 140].

Typically, classification performance is measured offline on a pre-recorded dataset using one of two techniques: hold-out or cross-validation. In the hold-out approach, the dataset is divided into two separate groups, one to train the MI model and the other to evaluate its performance. In the cross-validation approach, the MI model is trained and evaluated through several folds by repeating the process of dividing the dataset into training and testing sets several times in a complete round. The performance of the MI model is then averaged over all folds. The hold-out approach is preferred on large datasets, while on small datasets, cross-validation is preferred. In the reviewed papers, these two methods were used roughly equally, 53% of the studies used hold-out and 59% used cross-validation.

Since the EEG data is subject-dependent and session-dependent with large inter-subject and inter-session variabilities, the performance results for an MI model trained and tested using the same subject(s)/session(s) differ from the model tested with a new subject(s)/session(s) not seen during training. In performance evaluation, it is easy to confuse several terms that refer to subject or session dependency in the training and evaluation approaches. Here, we explain two sets of terms using subject dependency, which have the same meaning as for session dependency. The terms in the first set are within-subject, subject-specific, and subject-dependent, and for the second set, cross-subject and subject-independent. The first set of terms are often mixed in the literature and regularly used as synonyms, which refer to a training approach in which a portion of a subject's MI data is used to train a model tailored to that subject, and this process is repeated for all

Table 2 Summary of the extracted information for the reviewed papers

Study	Pre-processing			Input formulation*	Deep learning approaches		
	Selected channels	Analyzed frequency band (Hz)	Artifact removal approach		General strategy	Architecture	Activation function
Zhang et al. [128]	ALL (62)	8–30	N/A	RV: 2D matrices [TP × C]	CNN (adaptive transfer learn.)	5 CONV 1 FC 2 OUT	ELU: conv Smax: L-FC
Zhang et al. [114]	ALL (22, 3)	FB (0.5–100)	W	RV: 2D matrices [TP × C]	CNN (inception) (augment: NS)	6 × 5 CONV 2 FC 4/2 OUT	ReLU: conv N/A: FC Smax: L-FC
Avilov et al. [55]	variable (3–128)	4–38	N/A	RV: 2D matrices [TP × C]	CNN	3 CONV 1 FC 2 OUT	ELU: conv Smax: L-FC
Kumar et al. [85]	ALL (64)	Adaptive selection	W	EF: CSP	RNN-LSTM + SVM	2 LSTM-L 1 FC 2 OUT	N/A
Liu et al. [58]	ALL (22) + variable	FB (0.5–100)	W	TM: TP-3D	CNN (3D) (residual) (multi-branch)	10 CONV 3 FC 4 OUT	ELU: conv ReLU: FC Smax: L-FC
Zhao et al. [129]	ALL (22, 3)	4–38	W	RV: 2D matrices [TP × C]	CNN (domain adaptation)	2 CONV 3 FC 4/2 OUT	ReLU: conv ReLU: FC sigm: L-FC
Bang et al. [81]	DS1: 22 DS2: 3 DS3: 20	4–40	N/A	EF: NSCM	CNN (3D)	2 CONV 2 FC 2 OUT	ReLU: conv ReLU: FC N/A: L-FC
Deng et al. [103]	ALL (22, 60)	4–38	W	RV: 2D matrices [TP × C]	CNN	3 CONV 1 FC 4 OUT	ELU: conv Smax: L-FC
Ha et al. [127]	ALL (22, 3)	4–38	W	RV: 2D matrices [TP × C]	CNN (multi-level pooling)	4 CONV 2 FC 4/2 OUT	ELU: conv ELU: FC Smax: L-FC
Zhang et al. [88]	ALL (22)	4–40	N/A	EF: CSP	Hybrid: CNN/LSTM (transfer learn.)	3 CONV 1 LSTM-L 4 FC 4 OUT	ReLU: conv ReLU: FC Smax: L-FC
Riyad et al. [125]	ALL (22)	0–38 4–38	W	RV: 2D matrices [TP × C]	CNN (inception) (augment: SW)	11 CONV 1 FC 4 OUT	ELU: conv Smax: L-FC
Liu et al. [121]	ALL (22)	0–38	W	RV: 2D matrices [TP × C]	CNN (self-attention) (transfer learn.)	7 CONV 1 FC 4 OUT	N/A: conv Smax: L-FC
Xue et al. [82]	ALL (22, 60)	4–40	N/A	EF: CSP	CNN (multi-layer)	7 CONV 3 FC 4 OUT	ELU: conv ELU: FC Smax: L-FC
Li et al. [106]	ALL (22)	8–30	W	RV: 2D matrices [TP × C]	CNN (multi-scale) (attention)	10 CONV 2 FC 4 OUT	ReLU: conv ReLU: FC Smax: L-FC
Li et al. [59]	ALL (64) + variable	N/A	W	TM: TP (D)	Hybrid: CNN/GRU	3 CONV 1 FC 2 GRU-L 2 FC 4 OUT	N/A: conv N/A: FC Smax: L-FC
Fan et al. [104]	ALL (64)	0.1–64	W	RV: 2D matrices [TP × C]	CNN (attention) (residual)	14 CONV 1 FC 4 OUT	ReLU: conv N/A: L-FC
Roy et al. [94]	ALL (3)	4–32	N/A	SI: TFI (STFT) [T × F × C]	CNN	3 CONV 2 FC 2 OUT	ReLU: conv N/A: FC Smax: L-FC
Xiaoling et al. [140]	28	8–30	N/A	RV: 2D matrices [TP × C]	CNN	2 CONV 2 FC 2 OUT	tanh: conv sigm: FC sigm: L-FC
Lun et al. [57]	2 + variable	N/A	W	RV: 2D matrices [TP × C]	CNN	5 CONV 1 FC 4 OUT	LReLU: conv Smax: L-FC
Roots et al. [105]	ALL (64)	2–60	W	RV: 2D matrices [TP × C]	CNN (multi-branch) (augment: SW)	3 × 3 CONV 1 FC 2 OUT	ELU: conv Smax: L-FC

Table 2 (continued)

Study	Pre-processing			Input formulation*	Deep learning approaches		
	Selected channels	Analyzed frequency band (Hz)	Artifact removal approach		General strategy	Architecture	Activation function
Yang et al. [60]	Variable (3–25)	7–35	A: CAR	RV: 2D matrices [TP × C]	Hybrid: CNN/SAE (multi-layer-CNN)	5 CONV 2 AE (1 hid) 1 FC 2 OUT	ReLU: conv Smax: L-FC
Zhao et al. [83]	ALL (64, 22)	4–40	N/A	EF: CSP	CNN (domain adaptation)	4 CONV 1 FC 2 OUT	ReLU: conv Smax: L-FC
Zhang et al. [100]	ALL (64, 22)	FB (DS2: 0.5–100)	W	TM: TP (G)	Hybrid: CNN/LSTM (recurrent attention)	1 CONV 2 LSTM-L 1 FC 4 OUT	ELU: conv Smax: L-FC
Xu et al. [116]	ALL (22)	9–20 + variable	M	TM: TP (D)	CNN	3 CONV 2 FC 4 OUT	ReLU: conv ReLU: FC Smax: L-FC
Zhang et al. [141]	ALL (3)	8–30	W	SI: TFI (STFT) [T × F + C]	CNN	2 CONV 2 FC 2 OUT	ReLU: conv Smax: FC Smax: L-FC
Liao et al. [117]	ALL (22)	4–40	N/A	TM: TP (D)	CNN	3 CONV 1 FC 4 OUT	LReLU: conv Smax: L-FC
Zhang et al. [91]	3	8–30	W	SI: TFI (STFT) [T × F + C]	Hybrid: CNN/GAN (also VAE)	4:4 CONV CNN: 2 CONV 2 FC 2 OUT	LReLU: d-conv ReLU: conv Smax: FC Smax: L-FC
Miao et al. [95]	49	8–30	N/A	SI: SFI (Energy) [C × F-band]	CNN	2 CONV 3 FC 2 OUT	ReLU: conv ReLU: FC Smax: L-FC
Tang et al. [9]	3	8–30	A	EF: (EMD)	CNN (1D) (multi-scale) (inception)	4 CONV 2 FC 2 OUT	ReLU: conv N/A: FC Smax: L-FC
Shajil et al. [92]	5	1–100 13–30	N/A	SI: TFI (STFT) [T × F + C]	CNN	1 CONV 2 FC 4 OUT	ReLU: conv N/A: FC Smax: L-FC
Xu et al. [78]	ALL (22)	8–30	A: EEGLAB	EF: WPD, CSP	DBN-RBM (stacked RBM) + SVM	4 RBM (1 hid) 4 OUT	sigm: RBM Linear: last RBM
Taheri et al. [75]	1	N/A	N/A	EF: CSP, DCT, EMD	CNN + SVM (multi-branch)	5 CONV 2 FC 2 OUT	ReLU: conv ReLU: FC
Wang et al. [137]	ALL (22)	8–30	M	RV: 2D matrices [TP × C]	Hybrid: CNN/LSTM	3 CONV 1 FC 2 LSTM-L 1 FC 4 OUT	ELU/Linaer: conv Smax: L-FC
Li et al. [96]	ALL (3)	4–30	N/A	SI: TFI (CWT) [T × F × C]	CNN	2 CONV 2 FC 2 OUT	ReLU: conv ReLU: FC Smax: L-FC
Rong et al. [93]	ALL (3)	4–32	W	SI: TFI (STFT) [T × F + C]	CNN	3 CONV 1 FC 2 OUT	ReLU: conv Smax: L-FC
Ma et al. [76]	ALL (22)	0.5–50	A	EF: DWT + PSD	CNN	4 CONV 2 FC 4 OUT	ReLU: conv N/A: FC N/A: L-FC
Hou et al. [120]	ALL (64)	8–30	A	EF: WT	CNN	6 CONV 2 FC 4 OUT	LReLU: conv LReLU: FC Smax: L-FC
Freer et al. [138]	ALL (22)	7–30	N/A	N/A	Hybrid: CNN/LSTM (augment)	4 CONV 1 LSTM-L 1 FC 4 OUT	ELU: conv Smax: L-FC
Dai et al. [111]	3	4–32	N/A	RV: 2D matrices [TP × C]	CNN (multi-layer) (augment)	2 CONV 2 FC 4/2 OUT	ELU: conv N/A: FC N/A: L-FC

Table 2 (continued)

Study	Pre-processing			Input formulation*	Deep learning approaches		
	Selected channels	Analyzed frequency band (Hz)	Artifact removal approach		General strategy	Architecture	Activation function
Lee et al. [112]	24	4–40	N/A	RV: 2D matrices [TP × C]	CNN (multi-branch)	4 CONV 1 FC 9 OUT	ELU: conv Smax: L-FC
Huang et al. [79]	ALL (22)	FB 0.5–100	N/A	EF: HHT	CNN	5 CONV 2 FC 4 OUT	Linear/ReLU: conv N/A: FC Smax: L-FC
Li et al. [118]	ALL (64, 22, 3)	8–30	N/A	TM: TP (D)	CNN	31 CONV 1 FC 4/4/2 OUT	ReLU: conv Smax: L-FC
Alwasiti et al. [67]	ALL (64)	2–78	A: CAR	SI: ST [T + C × F + C]	CNN (DenseNet) (deep metric learning)	1 CONV 4 DB 2 FC 3 OUT	ReLU: conv ReLU: DB ReLU: FC Smax: L-FC
Jeong et al. [62]	20	4–40	A: ICA	RV: 2D matrices [TP × C]	CNN (multi-layer)	5 CONV 2 FC 3 OUT	ELU: conv ELU: FC Smax: L-FC
Cheng et al. [86]	ALL	0.5–30	N/A	EF: PCA	DBN-RBM	5 RBM (1 hid) 1 FC 2 OUT	Smax: L-FC
Collazos et al. [119]	ALL (22)	8–30	N/A	TM: SP (CWT, PSD)	CNN (multiple input CNN)	4 CONV 2 FC 3 OUT	ReLU: conv ReLU: FC Smax: L-FC
Chen et al. [143]	ALL (22, 15)	8–30	W	EF: CSP	CNN	3 CONV 2 FC 4 OUT	ReLU: conv N/A: FC Smax: L-FC
Kant et al. [97]	ALL (2)	8–30	A: DWT	SI: TFI (CWT) [T + C × F]	CNN (transfer learning)	14 CONV 4 FC 2 OUT	ReLU: conv ReLU: FC Smax: L-FC
Fahimi et al. [64]	ALL	0.5–100	A: ICA, ASR	RV: 2D matrices [TP × C]	Hybrid: CNN/ GAN	2:2 CONV 3 CONV 2 FC 2 OUT	tanh: G-conv ReLU: conv ReLU: FC sigm: L-FC
Ma et al. [66]	ALL (64)	0.1–40	A: CAR, AAR	EF: CorrM	CNN (multi-branch)	2 CONV 1 FC 3 OUT	ReLU: conv Smax: L-FC
Hassanpour et al. [51]	ALL (22) + variable	8–35	W (+ A: SWT)	EF: FFT	DBN-AE DBN-RBM (augment: SW)	5 RBM/AE (1 hid) 1 FC 4 OUT	N/A: AE Smax: L-FC
Zhu et al. [56]	variable (3–64)	N/A	N/A	RV: 2D matrices [TP × C]	Hybrid: CNN/ LSTM (also CNN)	2 CONV 1 LSTM-L 1 FC 2 OUT	N/A
Lee et al. [146]	ALL (3)	8–30	N/A	SI: TFI (CWT) [T × C + F]	CNN	1 CONV 1 FC 2 OUT	ReLU: conv N/A: L-FC
Zhang et al. [87]	ALL (22)	4–38	N/A	EF: FBCSP	Hybrid: CNN/ LSTM	3 CONV 3 LSTM-L 4 OUT	ReLU: conv Smax: out
Amin et al. [115]	ALL (22)	0.5–40	W	RV: 2D matrices [TP × C]	Hybrid: CNN/ MLP (M) CNN/AE (A) (multi-layer-CNN)	5 CONV 4 FC MLP (2 hid)/ AE (1 hid) 4 OUT	ELU: conv ELU: AE ELU: MLP N/A: FC Smax: L-FC
Wu et al. [113]	ALL (22, 3)	4–38	W	RV: 2D matrices [TP × C]	CNN	5 CONV 1 FC 4/2/3 OUT	Linear
Ortiz et al. [147]	18	0.5–90	A: BSS + MRIC	SI: CWT [T × C + F]	CNN	2 CONV 2 FC 2 OUT	ReLU: conv ReLU: FC Smax: L-FC

Table 2 (continued)

Study	Pre-processing			Input formulation*	Deep learning approaches		
	Selected channels	Analyzed frequency band (Hz)	Artifact removal approach		General strategy	Architecture	Activation function
Zhao et al. [54]	ALL (22) + variable	0.5–100 + variable	W	TM: TP-3D	CNN (3D) (multi-branch)	3 CONV 3 FC 4 OUT	ELU: conv ReLU: FC Smax: L-FC
Kumar et al. [84]	ALL (64, 59)	7–30	W	EF: CSP	RNN-LSTM	2 LSTM-L 1 FC 2 OUT	N/A
Chaudhary et al. [148]	N/A	N/A	N/A	SI: TFI (STFT/ CWT) [T × F]	CNN	5 CONV 3 FC 2 OUT	ReLU: conv Smax: FC Smax: L-FC
Li et al. [135]	ALL (22)	N/A	N/A	RV: 2D matrices [TP × C]	CNN (augment: AP)	5 CONV 2 FC 4 OUT	ELU: conv ELU: FC Smax: L-FC
Tang et al. [149]	DS1: ALL (3) DS2: 6	8–30	W	RV: 2D matrices [TP × C]	DSN-RBM (semi-supervised)	7 RBM (1 hid) 2 OUT	N/A
Zhu et al. [123]	ALL (3, 15)	8–30	W	EF: CSP	CNN (residual)	13 CONV 1 FC 2 OUT	ReLU: conv Smax: L-FC
Dai et al. [99]	ALL (3, 5)	6–30	W	SI: TFI (STFT) [T × F + C]	Hybrid: CNN/ VAE	1 CONV 5 hid. VAE 2 OUT	ReLU: conv
Olivas-Padilla et al. [71]	8	8–30	A: BCILAB	EF: FBCSP (set as a matrix)	CNN	4 CONV 1 FC 4 OUT	ReLU: conv Smax: L-FC
Alazrai et al. [68]	ALL (16)	0.5–32.5	A: AAR	SI: TFI (QTFD) [T × F + C]	CNN + SVM	2 CONV 1 FC 11 OUT	ReLU: conv Smax: L-FC
Li et al. [126]	ALL (22)	8–30	M	EF: CSP	CNN (multi-layer)	9 CONV 2 FC 4 OUT	ReLU: conv Smax: FC Smax: L-FC
Xu et al. [130]	ALL (3)	4–32	W	SI: TFI (STFT) [T × F × C]	CNN (transfer learning)	13 CONV 3 FC 2 OUT	ReLU: conv ReLU: FC Smax: L-FC
Zhang et al. [150]	ALL (3, 14)	8–30	W	SI: TFI (WT: MW) [T × F × C]	CNN (augment)	2 CONV 2 FC 2 OUT	ReLU: conv ReLU: FC Smax: L-FC
Amin et al. [108]	ALL (22)	FB (0.5–100)	W	RV: 2D matrices [TP × C]	CNN (multi-layer)	5 CONV 4 FC 1 FC 4 OUT	ELU: conv ELU: FC Smax: L-FC
Tayeb et al. [7]	ALL (3)	2–60	A: ICA (FASTER)	SI: TFI (STEF) [T × F]	CNN (also LSTM, and RCNN (CNN\RNN))	3 CONV 1 FC 2 OUT	ReLU: conv Smax: L-FC
Kwon et al. [73]	20	0–40	N/A	EF: CSP	CNN (multi-branch)	3 × 3 CONV 2 FC 2 OUT	ReLU: conv N/A: FC Smax: L-FC
Xu et al. [65]	DS1: 3 (DS2:ALL)	8–30	A: CAR	SI: TFI (WT) [T × F × C]	CNN	2 CONV 2 FC 4/2 OUT	ReLU: conv N/A: FC N/A: L-FC
She et al. [74]	ALL (22)	8–30	N/A	EF: CSP	ELM	3 hid 2 OUT	N/A
Dose et al. [109]	ALL (64)	N/A	W	RV: 2D matrices [TP × C]	CNN	2 CONV 1 FC 2/3/4 OUT	ReLU: conv Smax: L-FC
Wang et al. [90]	3	8–30	N/A	SI: TFI (STFT) [T × F + C]	CNN (also LSTM)	2 CONV 2 FC 2 OUT	SELU: conv N/A: FC Smax: L-FC

Table 2 (continued)

Study	Pre-processing			Input formulation*	Deep learning approaches		
	Selected channels	Analyzed frequency band (Hz)	Artifact removal approach		General strategy	Architecture	Activation function
Sakhavi et al. 2018, [72]	ALL (22)	4–40	N/A	EF: CSP	CNN	3 CONV 1 FC 4 OUT	ReLU: conv Smax: L-FC
Wang et al. [80]	ALL (22)	FB (0.5–100)	N/A	EF: SM	RNN-LSTM	3 LSTM-L 2 OUT	N/A
Lawhern et al. [107]	ALL (22)	4–40	W	RV: 2D matrices [TP × C]	CNN	3 CONV 1 FC 4 OUT	ELU: conv Smax: L-FC
Luo et al. [70]	ALL (22, 3)	8–30	W	EF: FBCSP (time slices)	RNN-GRU (also RNN-LSTM)	2 GRU-L/ LSTM-L 1 FC 4/2 OUT	N/A
Chu et al. [61]	ALL (64)	FB (8–35)	A	EF: PSD (LSP)	DBN-RBM	3 RBM (1 hid) 1 FC 3 OUT	Smax: L-FC
Yang et al. [63]	9	N/A	A: ICA	RV: 2D matrices [TP × C]	Hybrid: CNN/ LSTM	3 CONV 1 LSTM-L 1 FC 2 OUT	ReLU: conv Smax: L-FC
Tang et al. [151]	DS1: ALL(3) DS2: 6	8–30	W	RV: 2D matrices [TP × C]	DSN-RBM	2 RBM (1 hid) 2 OUT	N/A
Tang et al. [110]	ALL (28)	8–30	N/A	RV: 2D matrices [TP × C]	CNN	2 CONV 1 FC 2 OUT	tanh: conv sigm: FC sigm: L-FC
Uktveris et al. [89]	ALL (22)	7–30	N/A	SI: SFI (FFT) [C × F]	CNN	2 CONV 1 FC 4 OUT	ReLU: conv Smax: L-FC
Lu et al. [77]	ALL (3)	8–35	W	EF: FFT (also WPD)	DBN-RBM	3 RBM (1 hid) 1 FC 2 OUT	Smax: L-FC
Tabar et al. [98]	ALL (3)	6–30	W	SI: TFI (STEF) [T × F + C]	Hybrid: CNN/ SAE (also, CNN, SAE)	1 CONV 6 AE (1 hid) 1 FC 2 OUT	ReLU: conv sigm: AE N/A: FC

Study	Dataset	Performance evaluations			
		Strategy	Performance measures		
			Accuracy %	kappa	Others (name)
Zhang et al. [128]	Lee et al. [50]	sub-d: HO (70: 30) sub-i: CV (LOSO)	sub-d: 63.54 ± 14.25 sub-i: 84.19 ± 9.98	_	Computation time, t-test
Zhang et al. [114]	DS1: BCI-C IV-2a [131] DS2: BCI-C IV-2b [101]	HO (75: 25)	DS1: 88.4 ± 7 DS2: 88.6 ± 5	_	CM, ROC, AUC, F-score, TPR
Avilov et al. [55]	Local: 22 sub, 2 MI (presses/releases a button), 128 elec, 1144 trials/class, 2048 Hz	CV (10 folds)	83.2	_	_
Kumar et al. [85]	GigaDB [132]	CV (10 folds)	69.59	0.398	TPR, TNR
Liu et al. [58]	BCI-C IV-2a [131]	CV (10 folds)	81.22 ± 6.85	0.72 ± 0.12	p-value, test/train time
Zhao et al. [129]	DS1: BCI-C IV-2a [131] DS2: BCI-C IV-2b [101]	c-sub: HO DS1: (50:50) DS2: (56:44)	DS1: 74.75 DS2: 83.98	DS1: 0.663 DS2: 0.68	_
Bang et al. [81]	DS1: BCI-C IV-2a [131] DS2: BCI-C IV-2b [101] DS3: Lee et al. [50]	CV (10 folds)	DS1: 87.15 DS2: 75.85 DS3: 70.37	_	t-test
Deng et al. [103]	DS1: BCI-C IV-2a [131] DS2: BCI-C III-3a [136]	CV (5 folds)	DS1: 78.96 DS2: 85.30	DS1: 0.72 DS2: 0.80	t-test
Ha et al. [127]	DS1: BCI-C IV-2a [131] DS2: BCI-C IV-2b [101]	HO DS1: (50:50) DS2: (56:44)	DS1: 73.19 DS2: 82.83	_	w-test

Table 2 (continued)

Study	Dataset	Performance evaluations			
		Strategy	Performance measures		
			Accuracy %	kappa	Others (name)
Zhang et al. [88]	BCI-C IV-2b [101]	c-sub: HO (50: 50)	_	0.81	_
Riyad et al. [125]	BCI-C IV-2b [101]	CV (5 folds)	74.61	0.662	CM
Liu et al. [121]	BCI-C IV-2b [101]	c-sub: CV (10 folds) HO (50: 50)	HO: 78.51 CV: 90.15	_	CM
Xue et al. [82]	DS1: BCI-C IV-2b [101] DS2: BCI-C III-3a [136]	HO (70: 30)	DS1: 83.83 DS2: 89.45	DS1: 0.78 DS2: 0.86	_
Li et al. [106]	BCI-C IV-2b [101]	HO (50: 50)	79.9	_	CM
Li et al. [59]	EEGMMIDB [139]	HO (75: 25)	97.36	_	_
Fan et al. [104]	EEGMMIDB [139]	CV (5 folds)	65.82	_	CM
Roy et al. [94]	BCI-C IV-2b [101]	sub-d: HO (56: 44) sub-i: CV (LOSO)	sub-d: 77.5 ± 14.5 sub-i: 70.9 ± 9.9	sub-d: 0.55 ± 0.29 sub-i: 0.42 ± 0.2	_
Xiaoling et al. [140]	Local: 4 sub, 2 MI left-hand/foot, 560 trials/sub, 1000 Hz (1–40 Hz), 64 elec	HO (80: 20)	90.08 ± 2.22	_	CM, RC, PR, F-score, ROC, w-test, T-comp
Lun et al. [57]	EEGMMIDB [139]	sub-d: CV (10 folds) sub-i: HO (106: 3 subs)	sub-d: 94.80 sub-i: 72.47	_	CM, RC, PR, F-score, ROC, AUC
Roots et al. [105]	EEGMMIDB [139]	c-sub HO (80: 20)	83.8	_	CM, RC, PR, F-score, t-test
Yang et al. [60]	DS1: BCI-C IV-1 [133] DS2 (Local): 6 sub, 2 MI L/R hand, 64 elec, 300 trials/sub, 256 Hz	sub-d: CV (8 folds) sub-i: CV (LOSO)	sub-i: DS1: 86.4 DS2: 84.7	sub-i DS1: 0.45 DS2: 0.46	_
Zhao et al. [83]	DS1: GigaDB [132] DS2: BCI-C IV-2a [131]	c-sub: HO (8: 1 subs) (5: 1 subs)	N/A	_	_
Zhang et al. [100]	DS1: EEGMMIDB [139] DS2: BCI-C IV-2a [131]	sub-i: HO (subs) DS1: (95:10) DS2: (8: 1)	DS1: 74.2 DS2: 60.1	_	ROC, AUC
Xu et al. [116]	BCI-C IV-2a [131]	HO	84.57	0.801	_
Zhang et al. [141]	BCI-C IV-2b [101]	CV (10 folds)	94.7 ± 2:6	0.664	_
Liao et al. [117]	BCI-C IV-2a [131]	HO (50: 50)	74.60	0.66	_
Zhang et al. [91]	DS1: BCI-C IV-1 [133] DS2: BCI-C IV-2b [101]	CV (10 folds)	DS1: 83.2 ± 3.5 DS2: 93.2 ± 2.8	DS1: 0.468 DS2: 0.671	t-test, p-value
Miao et al. [95]	DS1: BCI-C III-4a [136] DS2 (Local): 5 sub, 2 MI finger/rest, 21 elec, 1000 Hz	CV (10 folds)	DS1: 90.0	_	Running time
Tang et al. [9]	DS1: BCI-C IV-2b [101] DS2 (Local): 5 sub, 2 MI L/R hand, 14 elec, 128 Hz, 10 s trial	N/A	DS1: 82.61 DS2: 85.83	_	p-value
Shajil et al. [92]	Local: 12 sub, 4 MI (L/R hand, both hands, feet), 16 elec, 500 Hz	N/A	87.37 ± 1.68	_	_
Xu et al. [78]	BCI-C IV-2a [131]	CV (10 folds)	78.51	0.6278	_
Taheri et al. [75]	BCI-C III-4a [136]	HO (70: 30)	96.34	_	_
Wang et al. [137]	BCI-C IV-2a [131]	HO (50: 50)	_	0.64 ± 0.14	t-test

Table 2 (continued)

Study	Dataset	Performance evaluations			
		Strategy	Performance measures		
			Accuracy %	kappa	Others (name)
Li et al. [96]	BCI-C IV-2b [101]	CV (10 folds)	83.2	0.651	_
Rong et al. [93]	BCI-C IV-2b [101]	HO (90: 10)	82.8	0.663	_
Ma et al. [76]	BCI-C IV-2a [131]	CV (8 folds)	96.21	_	Test/train time
Hou et al. [120]	EEGMMIDB [139]	CV (10 folds)	94.5	_	_
Freer et al. [138]	BCI-C IV-2a [131]	N/A	_	_	PR, RC
Dai et al. [111]	DS1: BCI-C IV-2a [131] DS2: BCI-C IV-2b [101]	HO	DS1: 91.57 DS2: 87.6	_	P-values
Lee et al. 2020, [112]	Local: 9 MI, 12 sub, 50 trials/sess, 3 sess, 1000 Hz, 64 elec	CV (5 folds)	81	_	CM
Huang et al. [79]	BCI-C IV-2a [131]	CV (4 folds)	77.9	_	_
Li et al. [118]	DS1: EEGMMIDB [139] DS2: BCI-C IV-2a [131] DS3: BCI-C IV-2b [101]	DS1,3: CV (10 folds) DS2: HO (50: 50)	DS1,CV: 89 DS2,HO: 89 DS3,CV: 97	DS1: 0.77 DS2: 0.78 DS3: 0.94	CM, ROC, AUC
Alwasiti et al. [67]	EEGMMIDB [139]	HO (80: 20)	64.7	_	CM, PR, RC
Jeong et al. [62]	DS1: ULMov [142] DS2 (Local): 10 sub, 3 MI (forearm angle), 150 trials, 100 Hz, 32 elec	HO (80: 20)	DS1: 51.0 ± 4.0 DS2: 65.0 ± 9.0	_	CM, t-test
Cheng et al. [86]	DS1: BCI-C IV-2b [101] DS2: BCI-C II-3 [102]	CV (10 folds)	DS1: 91.71 DS2: 96.25	DS1: 0.8342 DS2: 0.925	t-test, test/train time
Collazos et al. [119]	BCI-C IV-2a [131]	CV (10 folds)	71.2 ± 7.0	0.56	p-values
Chen et al. [143]	DS1: BCI-C IV-2a [131] DS2: Steyrl et al. [144]	HO (70: 30)	DS1: 72 DS2: 82.9	DS1: 0.627 DS2: 0.657	CM, t-test
Kant et al. [97]	BCI-C II-3 [102]	HO (50: 50)	95.71	0.91	CM
Fahimi et al. [64]	DS1: BCI-C III-4a [136] DS2 (Local): 14 sub, 2 MI open/close R-hand, 62 elec	HO (50: 50)	DS1: 71.14	_	_
Ma et al. [66]	MIJoint [145]	CV (5 folds)	87.03	_	_
Hassanpour et al. [51]	BCI-C IV-2a [131]	HO (50: 50)	DBN-AE: 71.0 DBN-RBM: 68.4	_	t-test, train time
Zhu et al. [56]	EEGMMIDB [139]	N/A	82.93 (CNN: 79.7)	_	_
Lee et al. [146]	DS1: BCI-C IV-2b [101] DS2: BCI-C II-3 [102]	CV (10 folds)	DS1: 83.0 ± 1.6 DS2: 92.9	_	_
Zhang et al. [87]	BCI-C IV-2a [131]	HO	84	0.81	_
Amin et al. [115]	BCI-C IV-2a [131]	sub-d: HO (50: 50) sub-i: CV (LOSO)	sub-d: M: 75, A: 73 sub-i: M: 42, A: 55	_	CM, train time
Wu et al. [113]	DS1: BCI-C IV-2a [131] DS2: BCI-C IV-2b [101]	c-sub: HO	DS1: 75.9 DS2: 84.7	_	_
Ortiz et al. [147]	BCI-C III-4a [136]	CV (10 folds)	94.66	_	_
Zhao et al. [54]	BCI-C IV-2a [131]	CV (10 folds)	75.02	0.644	t-test, test/train time

Table 2 (continued)

Study	Dataset	Performance evaluations			
		Strategy	Performance measures		
			Accuracy %	kappa	Others (name)
Kumar et al. [84]	DS1: GigaDB [132] DS2: BCI-C IV-1 [133]	CV (10 folds)	DS1: 68.19 DS2: 82.52	DS1: 0.374 DS2: 0.650	TPR, TNR
Chaudhary et al. [148]	BCI-C III-4a [136]	HO (80: 20)	99.35	0.987	TPR, TNR, F-score, CM
Li et al. [135]	BCI-C IV-2a [131]	HO (50: 50)	74.6	_	CM, PR, RC, F-score, train time
Tang et al. [149]	DS1: BCI-C IV-2b [101] DS2 (Local): 7 sub, 2 MI L/R hand, 128 Hz, 240 trials/sub, 14 elec	HO	DS1: 83.55	_	p-value, train time
Zhu et al. [123]	DS1: BCI-C IV-2b [101] DS2 (Local): 25 sub, 2 MI L/R hand, 1000 Hz, 200 trials/sub, 15 elec	sub-i CV (LOSO)	DS1: 64.0 DS2: 73.0	_	ITR
Dai et al. [99]	DS1: BCI-C IV-2b [101] DS2 (Local): 5 sub, 2 MI L/R hand, 400 trials, 3 sess, 250 Hz, 5 elec	CV (10 folds)	_	DS1: 0.564 DS2: 0.568	p-value, train time
Olivas-Padilla et al. [71]	DS1: BCI-C IV-2a [131] DS2 (Local): 8 sub, 4 MI L/R hand/foot, 5 sess, 120 trials/sess, 250 Hz, 8 elec	DS1: HO (50: 50) DS2: CV (10 folds)	DS1: 78.41 ± 5.9 DS2: 73.78 ± 4.2	DS1: 0.59 ± 0.11 DS2: 0.64 ± 0.07	_
Alazrai et al. [68]	Local: 11 MI, 22 sub, 2048 Hz, 16 elec	CV (10 folds)	73.70	_	PR, RC, F-score, train/test time
Li et al. [126]	BCI-C IV-2a [131]	HO (50: 50)	79.9	_	_
Xu et al. [130]	BCI-C IV-2b [101]	HO (80: 20)	74.2	_	train time
Zhang et al. [150]	DS1: BCI-C II-3 [102] DS2 (Local): 5 sub, 2 MI L/R hand, 256 Hz, 120 trials/sub, 14 elec	CV (5 folds)	DS1: 90.1 DS2: 90.0	_	_
Amin et al. [108]	BCI-C IV-2a [131]	sub-i: CV (LOSO)	74.5	_	CM, PR, RC, train time
Tayeb et al. [7]	DS1: BCI-C IV-2b [101] DS2 (Local-public): 20 sub, 2 MI L/R hand, 2 sess, 4 runs, total 750 trials, 256 Hz, 3 elec	CV (5 folds)	DS1: CNN: 91.63 DS2: CNN: 84.24 RCNN: 77.7	_	_
Kwon et al. [73]	Lee et al. [50]	sub-d: HO (50: 50) sub-i: CV (LOSO)	sub-d: 71.3 ± 15.8 sub-i: 74.2 ± 15.8	_	t-test
Xu et al. [65]	DS1: BCI-C IV-2a [131] DS2: BCI-C II-3 [102]	CV (5 folds)	DS1: 85.59 DS2: 89.56	DS1: 0.766	F-score, train time
She et al. [74]	BCI-C IV-2a [131]	CV (9 folds)	67.76	0.5701	_
Dose et al. [109]	EEGMMIDB [139]	CV (5 folds)	2-class: 80.4 3-class:69.8 4-class:58.6	_	CM, PR, RE, train time
Wang et al. [90]	Local: 14 sub, 2 MI L/R hand, 60 trials/sub, 256 Hz, 11 elec	CV (4 folds)	CNN: 92.73 LSTM: 80.2	_	CM, p-value
Sakhavi et al. [72]	BCI-C IV-2a [131]	CV (10 folds)	74.46	0.659	_
Wang et al. [80]	BCI-C IV-2a [131]	CV (5 folds)	79.6	_	_
Lawhern et al. [107]	BCI-C IV-2a [131] (128 samples/s)	CV (4 folds)	69	_	AUC
Luo et al. [70]	DS1: BCI-C IV-2a [131] DS2: BCI-C IV-2b [101]	HO DS1: (50:50) DS2: (56:44)	GRU: 73.6, 82.8 LSTM: 72.6, 81.5	_	train/test time, complexity, t-test
Chu et al. [61]	Local: 9 sub, 3 MI L/R hand and foot, 10 runs, 300 trials/sub, 10 s trial, 1000 Hz, 64 elec	HO (75: 25)	70.72 ± 2.68	_	_

Table 2 (continued)

Study	Dataset	Performance evaluations			
		Strategy	Performance measures		
			Accuracy %	kappa	Others (name)
Yang et al. [63]	Local: 2 MI L hand/R foot, 6 sub, 500 Hz, 9 elec	HO (70: 30)	86.7	_	ROC, train time
Tang et al. [151]	DS1: BCI-C IV-2b [101] DS2 (Local): 7 sub, 2 MI L/R hand, 128 Hz, 240 trials/sub, 14 elec	HO (50: 50)	DS1: 81.35	_	p-value
Tang et al. [110]	Local: 2 sub, 2 MI L/R hand, 460 trials/sub, 1000 Hz, 28 elec	CV (10 folds)	86.4 ± 0.77	_	CM, PR, RC, F-score
Uktveris et al. [89]	BCI-C IV-2a [131]	CV (10 folds)	68	_	_
Lu et al. [77]	BCI-C IV-2b [101]	HO (56: 44)	84	_	t-test
Tabar et al. [98]	DS1: BCI-C IV-2b [101] DS2: BCI-C II-3 [102]	DS1: CV (10 folds) DS2: HO (50: 50)	DS1: 77.6 ± 2.1 DS2: 90.0	DS1: 0.55 DS2: 0.80	_

Pre-processing, Selected channels, ALL: All dataset channels, variable: varying numbers of channels. Analyzed frequency band, FB: full-bandwidth in the dataset (0–frequency-end). Artifact removal approach, W: Without, M: Manual, A: Automatic [ICA: Independent component analysis, DWT: Discrete wavelet transform, CAR: Common average reference filter, AAR: Automatic artifact removal toolbox, ASR: Artifact subspace reconstruction, BSS: Blind source separation, MRIC: Movement related independent component, SWT: Synchrosqueezed wavelet transforms]

Input formulation, (* refer to Fig. 10), RV: Raw values, EF: Extracted features [Frequency features [FFT: Fast Fourier transform, DCT: Discrete cosine transform, PSD: Power spectral density [LSP: Lomb-Scargle periodogram]], Time–frequency features [EMD: Empirical mode decomposition, HHT: Hilbert-Huang transform, WT: Wavelet transform, DWT: Discrete wavelet transform, WPD: Wavelet packet decomposition], Spatial features [CSP: Common spatial pattern, FBCSP: Filter bank CSP], NSCM: Normalized sample covariance matrix, SM: Statistical measures, CorrM: Correlation matrix, PCA: Principal component analysis], SI: Spectral images [TFI: Time–frequency images [ST: Stockwell transform, QTFD: Quadratic time–frequency distribution, WT [CWT: Continuous wavelet transform, MW: Morlet wavelets], STFT: Short-time Fourier transform], SFI: Spatial-frequency images], TM: Topological maps [TP: Time-domain point [D: Direct map, G: Graph-based], SP: Spectral-domain power].T: Time window (time segment), TP: Time point (sampling point), F: Frequency, F-band: Frequency band, C: Channel (electrode)

Deep learning approaches, General strategy, CNN, RNN [GRU, LSTM], MLP, RBM, AE, DBN [DBN-RBM, DBN-AE], ELM, DSN: Deep stacking network [DSN-RBM], GAN, VAE, Hybrid [CNN/LSTM, CNN/GRU, CNN/MLP, CNN/AE, CNN/VAE, CNN/GAN], SVM, multi-layer: multi-layer technique (for CNNs), multi-branch: multiple branches of CNNs (Ensemble learning), augment: data augmentation, SW: Sliding window, NS: Noise addition, AP: amplitude-perturbation. Architectures: CONV: Convolutional layer, FC: Fully connected layer, DB: Dense block, LSTM-L: LSTM layer, GRU-L: GRU layer, hid: hidden layer, OUT: number of (output) classes. Activation function, ReLU: Rectified linear unit, LReLU: Leaky rectified linear unit, ELU: Exponential linear unit, SELU: Scaled exponential linear unit, tanh: hyperbolic tangent, sigm: Sigmoid, Smax: Softmax function, Linear: Linear function, L-FC: Last fully connected layer, G-conv, d-conv: Convolution layer in a (generator/discriminant generator) model

Dataset, Local: Private dataset (not available), sub: Subjects, elec: Electrode, L/R: left/right, sess: Session, "x s trial": Trial duration is x seconds

Evaluation Strategy, HO: Hold-out (train: test), CV: Cross-validation, LOSO: leave-one-subject-out, c-sub: Cross-subject, sub-d: Subject-dependent, sub-i: Subject-independent, CM: Confusion matrix, PR: Precision (PPV), RC: Recall (True negative rate (TPR)/sensitivity), TNR: True negative rate (specificity), ITR: Information transfer rate, ROC: Receiver operating characteristic curve, AUC: Area under the curve, T-comp: Time complexity, w-test: Wilcoxon test, "(x: y subs)": x subjects for training and y subjects for testing

other subjects. In contrast, the term cross-subject refers to a training approach in which a set of MI data from several subjects is used to train a model that is common to all subjects, whereas the term subject-independent refers to an evaluation approach in which separate subjects (not seen in the training data) are used to evaluate the model, as illustrated in Fig. 13.

EEG datasets are usually acquired from several sessions where some sessions are used for training and the rest for testing (hold-out approach) [127, 129]. In this case, since the MI model is evaluated using MI data from different sessions, it is called a session-independent model. In some other studies, a cross-validation approach is used, in which the MI data from all sessions are merged and then randomly divided into k equal sets, which may over-rate the accuracy [121].

The inter-subject variability is a major challenge for MI classification [128]. Most studies train and test the MI model for specific subject using MI data obtained from the same subject. Researchers prefer this training strategy since it achieves higher classification accuracy; however, it provides less generalization ability over different subjects. Some other researchers have attempted to develop more general, subject-independent, MI models that are trained and tested using separate subjects [94, 100, 108, 115, 123, 128]. The performance of subject-independent models is evaluated using hold-out or cross-validation techniques, similar to those described previously. In the hold-out technique, some subjects are used for training and the rest for testing [57, 100]. In the cross-validation technique, known as leave-one-subject-out, the number of folds equals the number of subjects, and for each fold, one subject is used for evaluation and the others for training [115, 123, 128].

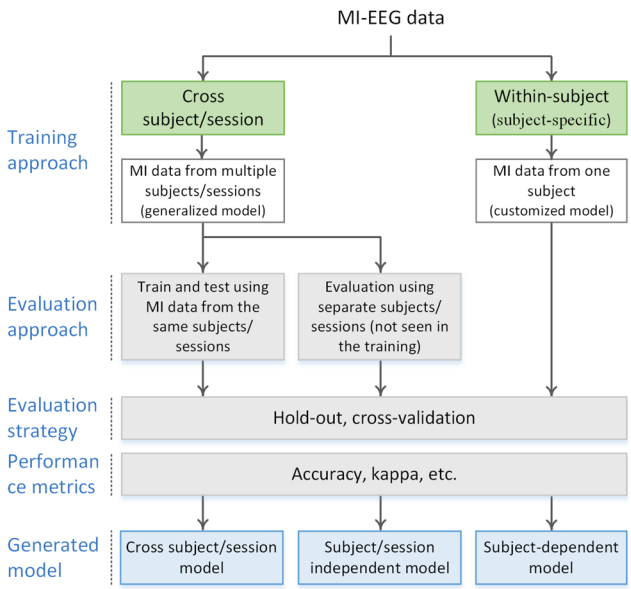

Fig. 13 Taxonomy of performance evaluation terms defined in this article

5 Public MI datasets

Publicly available datasets have been critical to the development and advancement of many areas such as computer vision [152, 153] and speech analysis [154]. Nowadays, global open science initiatives have encouraged many research institutes to publish the datasets for their articles. Therefore, it is now possible to develop BCI technology, and more specifically, MI-based BCI using advanced machine learning or deep learning techniques without the provision of expensive EEG recording equipment.

In this section, we extensively review open MI-EEG datasets, which can be used with deep learning techniques. The datasets were collected from two BCI-specialized websites (BNCI Horizon 2020 [155] and the website of Prof. Fabien Lotte [156]) and from the specialized dataset search engine by Google. Table 3 describes 18 publicly available datasets for EEG-based MI signals. In addition, Table 2 briefly mentions the key features of 21 private MI datasets that were used by the studies reviewed in this survey.

For the public MI datasets, the following characteristics were extracted: number and type of EEG (MI/motor/non-motor) and non-EEG classes, number and type of non-task-related EEG data (Rest/Noise/Other), #subjects (M and F), #trials (total/per-subject/per-class), #sessions, duration of the session, rest between sessions, #runs per session, duration of runs, rest between runs, #trials per run, #MI in a trial (one-MI per trial or several/long MIs per trial), trial duration [Before | MI | After], recording software, equipment name, #electrodes, type of electrodes, sampling rate, frequency band, voltage resolution, signal quality validation (during recording), and data validation (EEG signal analysis).

Table 3 describes all publicly available MI-EEG datasets released between 2002 and 2020, sorted from newest to oldest. The BCI-Competition II-3 [102] was the first public MI-EEG dataset, created by the Institute for Biomedical Engineering, Graz University of Technology in 2002. The dataset contains only one subject, with two classes for left- and right-hand imagery movements collected from three EEG electrodes at a sampling rate of 128 Hz.

The BCI Competition (BCI-C) datasets, II (II-3) [102], III (III-3a, III-3b, III-4a, III-4b, and III-4c) [136], and IV (IV-1 [133], IV-2a [131], and IV-2b [101])), can be described together as the most popular MI datasets. Among them, BCI-C IV 2a and 2b are the benchmarks in MI-EEG classification. On the other hand, BCI-C III-4 datasets (a, b, and c) have a high temporal and spatial resolution (i.e., a high sampling rate and a large number of electrodes) with a

Table 3 Public MI-EEG datasets

References	Key features	# MI classes (Type)	Non-MI	Non-EEG	# Non-TR EEG (Rest/Noise) (Type)	# Subjects	#Trials			# Sessions (duration) (time between sessions)
							Total	Per subject	Per class	
MIJoint[7] [145]	Different joints (same limb), Raw data, pre-processed, and direct trials	3 RH, RE, RS	_	EMG EOG	1 RS	25 19 M 6 F	7500 300 × 25	300 100 × 3	2500 100 × 25	7 MI:5, RS:2 (MI: 320 s, RS: 400 s) (5–10 min)
Lee et al. [50]	Three BCI paradigms: MI, ERP, and SSVEP	2 LH, RH	ERP, and SSVEP paradigms	EMG	_	54 25 F	21,600 TD: 10,800 ED: 10,800	400 TD: 100 × 2 ED: 100 × 2	10,800 200 × 54	2 (51 min) (DD)
MISCP[3] [158]	Many MI classes and a large number of trials	10 LH, RH, LL, RL,T, 5 FN	Press two keys on keyboard	_	1 RS	13 8 M 5 F	∼ 60,000	∼ 4600	_	75 (50–55 min) (_)
GigaDB[1] [132]	Many subjects, different signals: MI/Non-MI, EEG/Non-EEG/ Non-TR EEG	2 LH, RH	AM (1-run, same as MI tasks)	EMG	6 EB, EUD, ELR, HM, JC, RS	52 33 M 19 F	∼ 5200–6240 (AM: ∼ 1040)	100–120 (AM: 20)	∼ 2600–3120 (AM: ∼ 520)	1 (50 min) (_)
ULMov[17] [142]	Upper limb MI movements	6 EF, EE, FS, FP, HO, HC	AM (1-run, same as MI tasks)	EOG, GS, ES	1 RS	15 9 F	6300 420 × 15 (AM: 6300)	420 10 × 42 (AM: 420)	900 60 × 15 (AM: 900)	2 (AM:1, MI:1) (_) (DD)
IndImag[16] [157]	Several mental tasks performed by disabled peoples	2 RH, FT	Mental activities: SN, MWS, and MS	_	_	9 (disabled) 7 F	1440 160 × 9	160 80 × 2	720 80 × 9	2 (_) (DD)
Steyrl et al.[18] [144]	Two class motor imagery	2 RH, FT	_	_	_	14	2240 TD: 1400 ED: 840	160 TD: 50 × 2 ED: 30 × 2	1120 TD: 50 × 14 ED: 30 × 14	1 (_) (same session)
OpenViBE[14] [159]	Individual imagery	2 LH, RH	_	_	_	1	560	560	280	3 (_) (DD)
EEGMMIDB[2] [139]	Many subjects	4 LF, RF, BF, and BFT	AM (6-runs, same as MI tasks)	_	2 EO, EC	109	∼ 9156 84 × 109 (AM: ∼ 9156)	84 21 × 4 (AM: 84)	∼ 2289 21 × 109 (AM: ∼ 2289)	1 (26 min) (_)
BCI-C IV-1[4] [133]	Uncued classifier application (classification of continuous EEG)	2 Two of 3 classes (LH, RH, FT)	_	_	IS	7 3 AR	3080 TD: 1400 ED: 1680	440 TD: 200 ED: 240	1540 TD: 100 × 7 ED: 120 × 7	1
BCI-C IV-2a[5] [131]	Continuous classifier application and eye movement artifacts	4 LH, RH, FT, T	_	EOG	3 EO, EC, EM	9	5184 (4800 valid)	576 TD: 288 ED: 288	1296 TD: 72 × 9 ED: 72 × 9	2 (_) (DD)
BCI-C IV-2b[6] [101]	Session-to-session transfer and eye movement artifacts	2 LH, RH	_	EOG	3 EO, EC, EM	9	6480 (720 × 9)	720 SP: 120 × 2 FP: 160 × 3	3240 SP: 120 × 9 FP: 240 × 9	5 SP:2, FP:3 (_) (5 days)
BCI-C III-3a[8] [136]	Multi-class problems	4 LH, RH, FT, T	_	_	_	3	840	sub1: 360 sub2: 240 sub3: 240	∼ 210 90 + 60 × 2	1
BCI-C III-3b[9] [136]	Non-stationarity problems	2 LH, RH	_	_	_	3	2800	sub1: 640 sub2: 1080 sub3: 1080	∼ 1400 320 + 540 × 2	3

Table 3 (continued)

References	Key features	# MI classes (Type)	Non-MI	Non-EEG	# Non-TR EEG (Rest/Noise) (Type)	# Subjects	#Trials			# Sessions (duration) (time between sessions)
							Total	Per subject	Per class	
BCI-C III-4a[10] [136]	Small training sets, (subject-to-subject transfer)	2 RH, FT	_	_	_	5	1400 TD: 560 ED: 840	280 140 × 2	700 140 × 5	4
BCI-C III-4b[11] [136]	Classification of continuous EEG without trial structure	2 LH, FT	_	_	_	1	TD: 210	TD: 210	_	7 TD:3 ED: 4
BCI-C III-4c[12] [136]	Non-stationarity problems	2 LH, FT	_	_	_	1	630 TD: 210 ED: 420	630	_	7 TD:3 ED: 4
BCI-C II-3[13] [102]	First public MI dataset	2 LH, RH	_	_	_	1 F	280 TD: 140 ED: 140	280	140	1

References	#Runs per session (duration) (rest between runs)	# Trials per run (# MI in a trial)	Trial duration [B ǀ MI ǀ A] (s)	SW/HW	# Electrodes (type)	Sampling rate (Freq. band) (Hz)	Voltage resolution	Signal quality validation (during recording)	MI data validation (after recording)
MIJoint [145]	1	MI: 40 RS: 50 (ST)	8 [2 + 1 ǀ 4 ǀ 1]	E7	64 EEG (Gel)	1000 (0.5–100)	_	SQV-1 (≤ 5 kΩ) SQV-4	MIV-4 MIV-5
Lee et al. [50]	2 (offline/online) (22 min) (7 min)	100 (ST)	13 ± 1.5 [3 ǀ 4 ǀ 6 ± 1.5]	E6, E11	62 EEG 4 EMG	1000	_	SQV-1 (≤ 10 kΩ)	_
MISCP [158]	3 (15 min) (2 min)	300 (OT)	3 ± 0.5 [–ǀ 1 ǀ 2 ± 0.5]	S2, E3	19 EEG (Gel)	200 (0.53–70) 1000 (0.53–100)	24 bit (0.1 uV)	SQV-1 (≤ 10 kΩ) SQV-2 (within ± 0.25 cm)	MIV-4 MIV-5
GigaDB [132]	5–6 MI:4–5, AM:1 (140 s) (4 min max)	20 (OT)	7 [2 ǀ 3 ǀ 2]	S1, E1, E2	64 EEG 2 EMG	512	_	_	MIV-1 (Amp > ± 100 μV) MIV-2, MIV-3, MIV-4
ULMov [142]	10	42 (ST)	7.5 ± 0.5 [2 ǀ 3 ǀ 2–3]	E14	61 EEG 3 EOG 19 GS 13 ES	512 (0.01–200)	_	_	_
IndImag [157]	8 (_) (4)	25 (10 for MI) (ST)	13 ± 0.5 [3 ǀ 7 ǀ 3 ± 0.5]	S3, E13	30 EEG	256 (0.5–100)	_	_	_
Steyrl et al. [144]	8 (TD: 5, ED: 3)	20 (ST)	5 [–ǀ 5 ǀ–]	E15	15 EEG	512	_	_	_
OpenViBE [159]	4–5 (total 14 runs)	40 (ST)	6 [–ǀ 6 ǀ–]	E12	11 EEG	512	_	_	_
EEGMMIDB [139]	14 BL: 2, MI: 6, AM: 6 (BL: 1, MI: 2, AM: 2 min)	14 (ST)	8 [2 ǀ 4 ǀ 2]	S1	64 EEG	160	_	_	_
BCI-C IV-1 [133]	6 TD: 2, ED: 4 (_) (5–15 min)	TD: 100 ED: 60 (ST)	TD: 8 [2 ǀ 4 ǀ 2] ED: 8 [–ǀ 1.5–8 ǀ 1.5–8]	E4, E5	59 EEG	1000 (0.05–200)	16 bit (0.1 uV)	_	_
BCI-C IV-2a [131]	6 (∼ 6 min) (short break)	48 (ST)	8 ± 0.5 [2 ǀ 4 ǀ 2 ± 0.5]	E6	22 EEG 3 EOG	250 (0.5–100)	_	SQV-3	_
BCI-C IV-2b [101]	10 SP: 6, FP: 4	SP: 20 FP: 40 (ST)	SP: 9 ± 0.5 [3 ǀ 4 ǀ 2 ± 0.5] FP: 9 ± 0.5 [3 ǀ 4.5 ǀ 1.5 ± 0.5]	_	3 EEG 3 EOG	250 (0.5–100)	_	SQV-3	_
BCI-C III-3a [136]	6–7	40 (ST)	7 [3 ǀ 4 ǀ–]	E8	60 EEG	250 (1–50)	_	_	_

Table 3 continued

References	#Runs per session (duration) (rest between runs)	# Trials per run (# MI in a trial)	Trial duration [B | MI | A] (s)	SW/HW	# Electrodes (type)	Sampling rate (Freq. band) (Hz)	Voltage resolution	Signal quality validation (during recording)	MI data validation (after recording)
BCI-C III-3b [136]	4–9	_	8 [3 | 5 |–]	E9	2 EEG	125 (0.5–30)	_	_	_
BCI-C III-4a [136]	_	_	5.5 ± 0.25 [2 ± 0.25 | 3.5 |–]	E10	118 EEG	1000 (0.5–200)	16 bit (0.1 uV)	_	_
BCI-C III-4b [136]	_	_	TD: 5.5 ± 0.25 [2 ± 0.25 | 3.5 |–] ED: 3.25–10.25 [2 ± 0.25 | 1.5–8 |–]	E10	118 EEG	1000 (0.5–200)	16 bit (0.1 uV)	_	_
BCI-C III-4c [136]	_	_	TD: 5.5 ± 0.25 [2 ± 0.25 | 3.5 |–] ED: 3 ± 0.25 [2 ± 0.25 | 1 |–]	E10, E11	118 EEG	1000 (0.5–200)	16 bit (0.1 uV)	_	_
BCI-C II-3 [102]	7 (_) (several minutes)	40 (ST)	9 [3 | 6 |–]	E6, E9	3 EEG	128 (0.5–30)	_	_	_

BCI-C BCI competition

[1] http://gigadb.org/dataset/100295

[2] https://www.physionet.org/content/eegmmidb/1.0.0/

[3]https://doi.org/10.6084/m9.figshare.c.3917698

[4]http://www.bbci.de/competition/iv/#dataset1

[5]http://www.bbci.de/competition/iv/#dataset2a

[6]http://www.bbci.de/competition/iv/#dataset2b

[7]https://doi.org/10.7910/DVN/RBN3XG

[8]http://bbci.de/competition/iii/#data_set_iiia

[9]http://bbci.de/competition/iii/#data_set_iiib

[10]http://bbci.de/competition/iii/#data_set_iva

[11]http://bbci.de/competition/iii/#data_set_ivb

[12]http://bbci.de/competition/iii/#data_set_ivc

[13]http://www.bbci.de/competition/ii/

[14]http://openvibe.inria.fr/datasets-downloads/

[15]http://bnci-horizon-2020.eu/database/data-sets

[16]https://lampx.tugraz.at/~bci/database/004-2015/description.pdf

[17]https://lampx.tugraz.at/~bci/database/001-2017/dataset_description.pdf

[18]https://lampx.tugraz.at/~bci/database/002-2014/description.pdf

MI motor imagery, *F* female, *M* male, *TD* training data, *ED* evaluation data, *A* after MI trial, *B* before MI trial, *SP* screening paradigm, *FP* feedback paradigm, *BL* baseline, *s* seconds, *OT* one-MI per trial, *ST* sustained (several/long) MIs per trial, *DD* different days

Sensors: *EEG* electroencephalogram, *EOG* electrooculography, *EMG* electromyography, *GS* glove sensors, *ES* exoskeleton sensors

MI tasks (18): *LH* left hand, *RH* right hand, *LF* left fist, *RF* right fist, *BF* both fists, *HO* hand open, *HC* hand close, *LL* left leg, *RL* right leg, *FT* foot/feet, *BFT* both foot, *T* tongue, *FN* finger, *RE* right elbow, *EF* elbow flexion, *EE* elbow extension, *FS* forearm supination, *FP* forearm pronation

Non-MI tasks: *AM* actual motor movement, *MWS* mental word association, *MS* mental subtraction, *SN* spatial navigation, *ERP* event-related potential, *SSVEP* steady-state visually evoked potentials

Table 3 continued

Non-task-related (Non-TR) EEG: *EB* eye blinking, *EO* eye-open, *EC* eye-closed, *EUD* eyeball up/down, *EM* eye movement, *ELR* eyeball left/right, *HM* head movement, *JC* jaw clenching, *RS* resting state, *IS* idle state

SW software, *S1* BCI2000, *S2* Neurofax, *S3* g.tec GAMMAsys system

HW hardware (equipment)

E1: Biosemi ActiveTwo. (with Ag/AgCl active electrodes)

E2: 3D coordinate digitizer (Polhemus Fastrak)

E3: EEG-1200 JE-921A (Nihon Kohden, Japan)

E4: BrainAmp MR plus amplifiers (Brain Products GmbH, Munich, Germany)

E5: Ag/AgCl electrode cap (EASYCAP GmbH)

E6: Ag/AgCl electrodes

E7: Neuroscan SynAmps2 amplifier (Neuroscan, Inc.)

E8: 64-channel EEG amplifier from Neuroscan

E9: G.tec amplifier

E10: 128 channel Ag/AgCl electrode cap from ECI

E11: BrainAmp amplifier

E12: Mindmedia NeXus32B amplifier

E13: g.LADYbird active electrodes and two g.USBamp biosignal amplifiers (Guger Technologies, Graz, Austria)

E14: active electrodes (g.tec medical engineering GmbH, Austria)

E15: biosignal amplifier and active Ag/AgCl electrodes (g.USBamp, g.LADYbird, Guger Technologies OG, Schiedlberg, Austria)

Signal quality validation (during recording)

SQV-1: Electrode impedance

SQV-2: Distances between reference electrodes

SQV-3: Visual inspection by an expert to detect trials containing artifacts

SQV-4: EMG Validation to detect actual movements

MI data validation (After recording)

MIV-1: Mark bad trials

MIV-2: EMG/EEG correlation

MIV-3: Spectral analysis (ERD/ERS)

MIV-4: Classification of MI tasks

MIV-5: Examining ERP curves

sampling rate of 1000 Hz and 118 EEG electrodes and support session-to-session transfer.

The PhysioNet dataset (EEGMMIDB) [139] supports a large number of subjects: 109 subjects performing 84 trials of four MI tasks (imagining opening/closing left fist, right fist, both fists, both feet) in one session. The GigaDB [132] and Lee et al. [50] datasets also provide a large number of subjects with 52 and 54 subjects for each dataset, respectively. The ULMov [142] and MIJoint [145] datasets present imagined joints movement from the same limb. R. Scherer et al. [157] created a dataset (IndImag) for several mental activities including two MI tasks performed by people with disabilities.

Newly available MI datasets focus on increasing the number of trials with several MI classes. The MISCP [158] dataset gives the largest number of classes under five different paradigms that include 10 MI classes: left and right hand, left and right leg, tongue, and the five finger movements. Moreover, it has a large number of trials with a total of 60,000 trials for 13 subjects.

Figure 14 shows a visual representation of some key information for the public MI datasets presented in Table 3. The reported classification accuracy and kappa values for public MI datasets, as well as the training and evaluation methodologies used to arrive at the reported results, are shown in Fig. 15.

6 MI-Based applications

Formerly, the analysis of brain activity was limited to medical fields or laboratory exploration of brain functions. This restriction was due to the high cost of acquiring brain data in addition to the unreliability, high variability, and limited resolution of the acquired brain signals. However, this situation has improved significantly over the last two decades due to technological advancements. In the meantime, BCI research has extended to non-medical uses. MI BCI research has had an impact in a variety of areas, covering both medical and non-medical fields. Figure 16 shows the fields using MI-based BCI.

MI-based BCI has been used in several medical applications to recover, restore, or replace lost or weakened human functions. Post-stroke rehabilitation is one of the main applications of MI BCI in the medical field [4, 5]. Stroke rehabilitation aims to use brain waves to help stroke patients recover their lost motor abilities. Ramos-Murguialday et al. from the University of Tübingen in 2013 were the first to scientifically prove that BCI has a rehabilitative impact on stroke patients [160]. In rehabilitation, the BCI uses activity-dependent plasticity mechanisms by linking brain activation during action intention with peripheral feedback on the paralyzed limb to recover the motor function [5]. MI BCI has also been used in other medical applications to help patients with limb weakness or paralysis, as a result of a spinal cord injury, to restore lost or impaired limb function (e.g., movement) by controlling prostheses, exoskeleton, or electrical stimulation with brain waves [6, 7]. Other studies have used MI BCI to replace the walking function for people who cannot walk with a robotic wheelchair [8–10]. MI BCI systems have been used in simple applications, including cursor control, to help patients control a computer by moving the cursor on the screen [12], and in a speller that helps patients interact with a graphical interface that displays symbols or characters and allows them to choose the desired character using brain

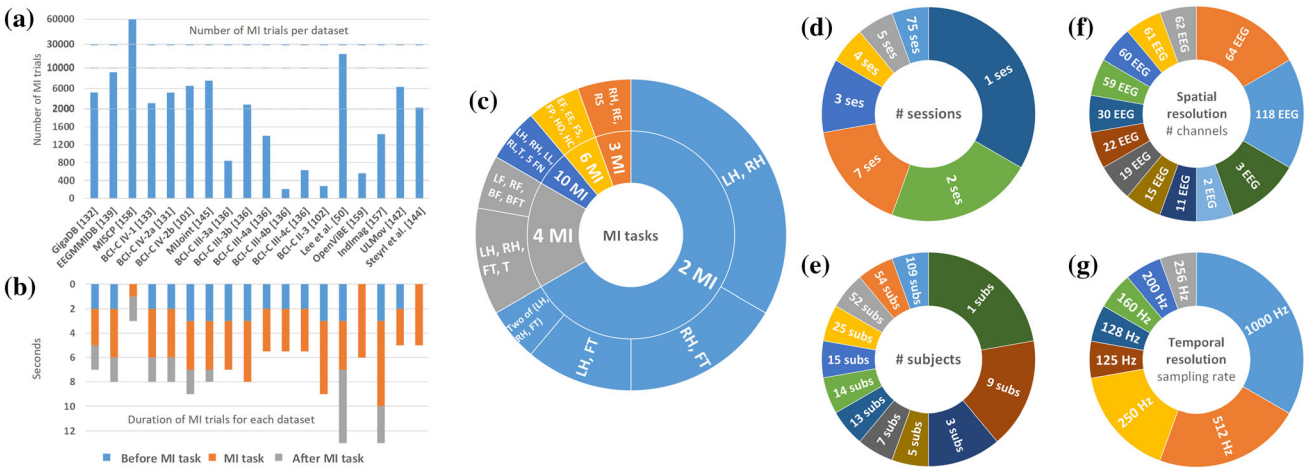

Fig. 14 Visualization of key information for the public MI datasets presented in Table 3, including the number of MI trials per dataset (**a**), duration of MI trials per data set (**b**), proportional representation of MI tasks (**c**), number of sessions (**d**), number of subjects (**e**), number of channels (spatial resolution) (**f**), and the sampling rate (temporal resolution) (**g**) in the public MI datasets

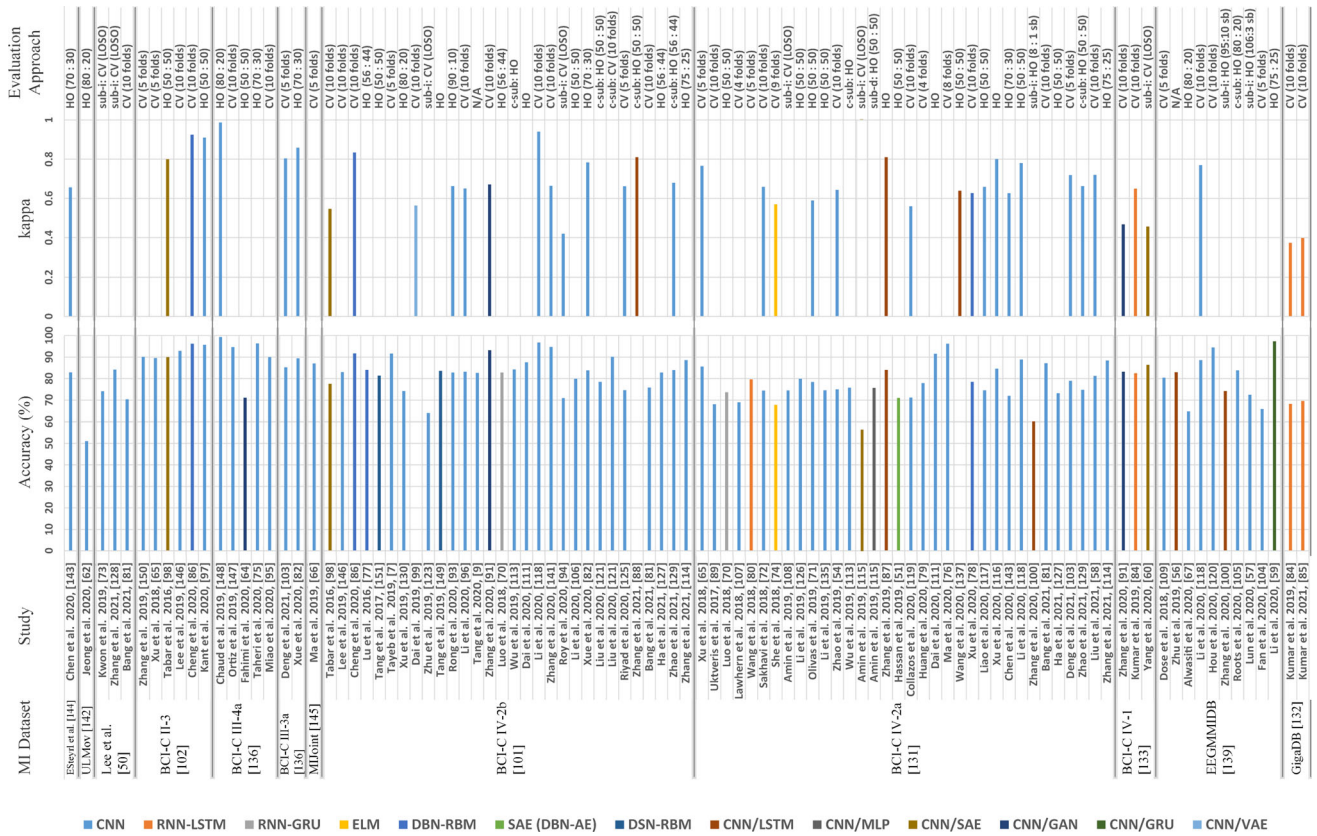

Fig. 15 Reported classification accuracies and kappa values for public MI datasets with the presentation of the applied evaluation approaches on which the reported results were based. HO: hold-out (train: test), CV: cross-validation, LOSO: leave-one-subject-out, c-sub: cross-subject, sub-d: subject-dependent, sub-i: subject-independent, sb: subjects, "(x: y sb)": x subjects for training and y subjects for testing

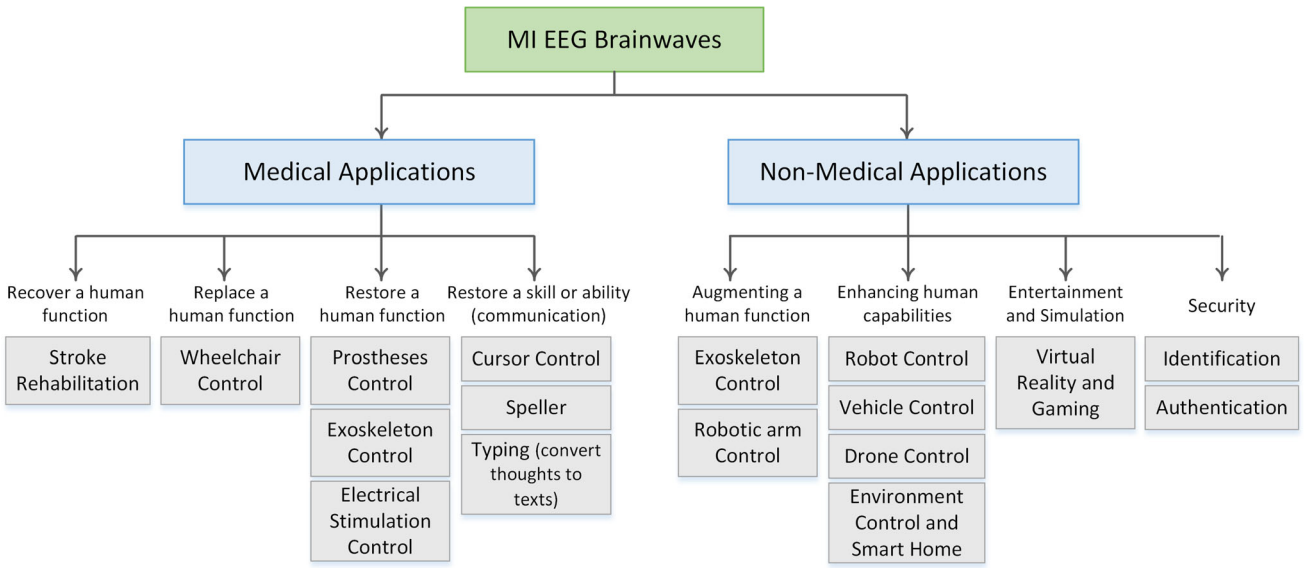

Fig. 16 Taxonomy of the motor imagery-based BCI applications

signals [11, 49]. In another study [161], the authors proposed a system that used MI-EEG signal to transform people's thoughts into text, which might help people with disabilities communicate with the outside world.

In non-medical applications, healthy people can use MI BCI for improving human capabilities, controlling unmanned vehicles, gaming, and entertainment. Improving human capabilities includes controlling an exoskeleton or a

robotic arm (a third arm) [162]. However, BCIs for augmenting human functions raise important ethical issues and debates [163] that have led to them receiving less attention than applications that improve quality of life for people with disabilities, which require little justification. MI BCIs have also been used to control unmanned ground or aerial vehicles [121, 164, 165], as well as the environment in smart homes [166]. The researchers in [164] conducted experiments with healthy users to control a robotic drone in three-dimensional space. In another study [165], researchers proposed a vehicle control model using MI-EEG brain signals. However, the practical use of such schemes in the physical world has yet to be investigated. Gaming and entertainment is another area that has been investigated in recent studies, where MI-EEG signals are used to control actions in the virtual world [16, 167, 168]. MI-EEG signals have also been used in security for identification and authentication [169, 170].

7 Discussion

In this section, three major questions about DL-based MI classification are addressed: (1) Is preprocessing required for DL-based techniques? (2) What input formulations are best for DL-based techniques? (3) What are the current trends in DL-based techniques?

7.1 Is preprocessing required for DL-based techniques?

7.1.1 Channel selection

Fewer EEG channels are preferred as they help reduce computational time, memory required, system complexity as well as preparation time during electrode placement and equipment cost. It can also help reduce the risk of overfitting that may occur by using irrelevant channels. However, by selecting fewer electrodes in inappropriate positions, important information may be lost; Hence, it is crucial to choose the optimal number of electrodes and their appropriate positions [57]. Several studies have investigated the effect of channel selection on MI classification [51, 54–57].

The study in [51] examined the effect of channel selection on the accuracy of MI classification using two DBN networks, DBN-AE and DBN-RBM. The authors compared the performance of the DBN models using all 22 channels in the dataset (BCI-C IV-2a [131]) with different numbers of channels (ranging from 16 to 21). The results showed that by reducing the number of EEG channels, the average accuracy of DBN models increased slightly in DBN-AE but decreased slightly in DBN-RBM. Because the changes in accuracy for both models were minimal, channel selection may be considered a redundant step with regard to the accuracy; however, it can help reduce the amount of processed MI data and thus speed up the deep model. The research in [54] studied the performance of a CNN model with a different number of EEG channels using the same dataset adopted in the aforementioned study. They divided the EEG channels into five groups according to the number (9, 13, or 22), the distribution (sparse or dense), and the position (front or back) of the electrodes. The study confirmed that the DL model could achieve competitive accuracy for MI classification with fewer electrodes (i.e., 9 or 13 compared with 22). Avilov et al. [55] demonstrated that using as few as 6 channels is sufficient to discriminate the motor imagery features. The findings in [57] also showed that CNN-based MI classification can achieve high accuracy with only two channels.

In another paper [56], Zhu et al. performed more investigation on the effect of different selections of EEG channels for MI classification using a dataset with 64 channels (EEGMMIDB [139]). The authors selected eight groups of channels in a cumulative manner, i.e., starting with a group of 3 electrodes up to 64, where subsequent groups included the same electrodes as the previous groups in addition to new electrodes. The channel selection was based on the position of the electrodes relative to the sensorimotor cortex. The results using CNN and CNN-LSTM models showed that the greater the number of channels, the higher the accuracy, indicating that all EEG channels include useful MI information. However, after a certain number of channels (i.e., 11 channels), the performance tended to be constant (slight improvement), indicating that a group of 11 channels, placed at the center of the sensorimotor cortex, was the optimal channel selection (high performance with low computation). As a conclusion, the number of EEG channels can be reduced without appreciable impact on accuracy while reducing the computational time and required memory.

7.1.2 Signal frequency filtering

Several studies have investigated the best selection of frequency bands for MI classification [54, 55, 92, 116, 125]. The study in [116] investigated a narrow frequency range (9–20 Hz) and compared it with that of the 9–30 Hz frequency range, showing that the narrower frequency range performed better in MI classification with a CNN model. However, other studies [54, 55, 92, 125] have shown that using relatively wider frequency bands improves the performance of MI classification. Avilov et al. [55] and Riyad et al. [125] analyzed the performance of CNN models for MI classification using (4–38 Hz) and (0–38 Hz) frequency bands. Avilov et al.

showed that the inclusion of the theta (4–8 Hz) and lower gamma (30–38 Hz) bands led to little improvement in performance. In [125], the authors demonstrated that the inclusion of the delta band (0–4 Hz) gave better performance. The researchers in [92] showed that the accuracy of MI classification using the CNN model with a 5-channel EEG signal was higher using a wide frequency band (1–100 Hz) compared to 8–30 Hz. The study in [54] analyzed the performance of a 3D multi-branch CNN model for MI classification using three different frequency bands (0.5–4 Hz, 0.5–38 Hz, and 0.5–100 Hz) and compared them with that of the 4–38 Hz frequency band. The findings showed that all three bands gave better performance than the commonly used frequency range (4–38 Hz), and the best performance was achieved using the raw full-band, i.e., without frequency filtering, (0.5–100 Hz) and the 0.5–38 Hz band, with a slight superiority for the full-band. This study also showed that there are distinct features related to MI tasks in very low frequencies (e.g., 0.5–4 Hz), which can be utilized and combined with those of the common range (4–38 Hz) to significantly improve the performance of MI classification. Based on the analyzed studies, for DL-based approaches, it is recommended to either use the raw MI-EEG signal without frequency filtering or filtering the MI-EEG signal with a low-pass filter, e.g., below 38 Hz.

7.1.3 Artifact removal

The authors in [51] compared the performance of DBN-RBM and SAE networks in MI classification using FFT features without artifact removal and with artifact removal using synchrosqueezed wavelet transforms (SWT). The results showed a slight increase in the average overall accuracy, 68.55% and 71.08%, with SWT-based artifact removal for the two DL models, DBN-RBM and SAE, respectively, compared with 68.43% and 71.0% without artifact removal. However, the specific results showed different changes in the accuracy (in both DL models) for different subjects, i.e., the accuracy was slightly increased for some subjects while slightly decreased for others, indicating the possible loss of some useful MI information from the EEG data for some subjects. Based on this study, because the accuracy improvement was minimal, artifact removal can be considered a redundant process for DL models.

Although the previous study showed that DL models could achieve competitive accuracy without artifact removal, the question of whether filtering MI-EEG data is still needed, as DL models are capable of extracting MI information from unfiltered data, remains to be answered.

7.2 What input formulations are best for DL-based techniques?

The choice of input formulation was largely dependent on the architecture of the deep learning model, as shown in Table 4 and Fig. 17.

For CNN-based DL networks, the four input formulations of MI-EEG signals were investigated by the studies reviewed, as shown in Fig. 17. Raw signal data was the most-used input formulation in CNN models (37%), followed by spectral images (29%) and extracted features (23%). However, CNNs had the best performance when topological images were used as inputs.

While topological maps could be generated from an MI-EEG signal in the time or spectral domain, time-domain points were the most popular option. The researchers in [118] used CNN models to compare the effect of constructing topological images from the MI-EEG signal in the time domain and the frequency domain and showed that the time-domain-based maps remarkably outperformed their counterparts in the frequency domain. This research also demonstrated that placing EEG data from each channel in incorrect order in the map reduced the classification accuracy by 12.3%, indicating that the information for electrode coordination according to their exact location on the scalp significantly enhanced the classification performance.

Although the topological map inputs showed the best performance in CNN models, the construction of these images is highly dependent on the number of EEG channels. The resolution of the topological image depends on the number of electrodes, i.e., the higher the number of electrodes, the higher the resolution of the image, and thus, the better the representation of spatial information. When the number of electrodes is small, the spatial information is compressed into a low-resolution image, which may not be able to make a valuable contribution to the results. Therefore, all proposed research that used topological mapping utilized all EEG channels, with a minimum of 22 channels. Zhao et al. [54] showed that the higher the number of electrodes, the better the MI classification using topological maps inputs. The study in [118] examined the proposed topological map in three datasets with 64, 22, and 3 channels and reported competitive results for all datasets, showing promising results that were insensitive to the number of channels.

Raw MI-EEG data were used as input in most of the CNN-based studies (37%), and the majority of those studies did not restrict the number of EEG channels. Raw MI data with CNNs have achieved competitive results compared to other input formulations, implying that CNN models are robust to handle the large size and high

Table 4 Summary of the different deep learning models with the input formulation used

		Input formulation								
		Extracted features				Spectral images		Raw values	Topological maps	
		Time features	Frequency features	Time–frequency features	Spatial features	Time–frequency images	Spatial-frequency images	2D matrices	Time domain	Spectral domain
Deep learning model										
CNN		[66, 81]	[75, 76]	[9, 75, 76, 79, 120]	[71–73, 75, 82, 83, 123, 126, 143]	[7, 65, 67, 68, 90, 92–94, 96–98, 130, 141, 146–148, 148]	[89, 95]	[55, 57, 62, 103–114, 121, 122, 124, 125, 127–129, 135, 140]	[54, 58, 116–118]	[67, 118, 119]
RNN	LSTM	[80]			[70, 84, 85]	[7, 90]				
	GRU				[70]					
ELM					[74]					
DSN	DSN-RBM							[149, 151]		
DBN	DBN-AE		[51]			[98]				
	DBN-RBM	[86]	[51, 77] [61]	[77, 78]	[78]					
Hybrid	CNN/RNN					[7]				
	CNN/GRU								[59]	
	CNN/LSTM				[87, 88]			[56, 63, 137, 138]	[100]	
	CNN/AE					[98, 99]		[60, 115]		
	CNN/MLP							[115]		
	CNN/GAN					[91]		[64]		
Generative	VAE					[91, 99]				
	GAN					[91]		[64]		

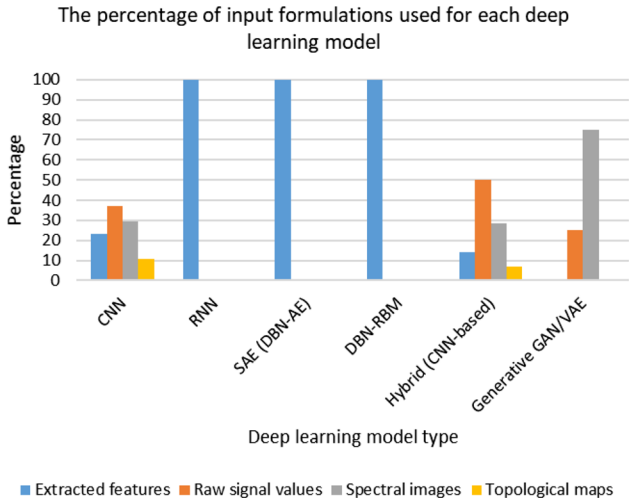

Fig. 17 Input formulation percentages by type of deep learning model. This figure considers the main proposed deep model for each paper, i.e., for studies that examined different deep models, the higher performance model was considered

dimensionality of raw EEG signals with minimal preprocessing.

Spectral images also yielded competitive results in the MI-EEG classification using CNN models. Time–frequency images were suggested in 97% of the studies: 55% of them used the STFT method, and 23% used the CWT method. The study in [148] compared the performance of a CNN model using time–frequency images generated by STFT and CWT techniques. The results achieved with the CWT method were superior to those achieved with the STFT technique, indicating that CWT images provide more distinct features than STFT in the time–frequency domain.

Manual feature extraction was used in 23% of studies that adopted CNN-based DL techniques. The average accuracy achieved by CNNs with handcrafted features across all public MI datasets was comparable to those achieved by CNNs with other input formulations. This indicates that the time consumption of extracting features manually does not add a significant gain in the learning ability of deep CNN models. Among all feature extraction techniques, CSP was the most used approach (56%) followed by time–frequency features (31%) (e.g., WPD, DWT, EMD, and HHT) and frequency features (13%) (e.g., FFT, DCT, and PSD).

RNN studies, including LSTM and GRU, used only extracted features as inputs. The average accuracy of the RNN models across all public datasets using extracted features was 77.34%. However, because there are fewer studies that used RNN, more research is needed to investigate the different input formulations of RNN models.

DBN followed the same pattern as in the RNN studies in terms of the input formulation, where extracted features were used as inputs for all DBN models (DBN-RBM and DBN-AE), as shown in Fig. 17. The average accuracy achieved by the reviewed DBN studies using extracted features was 84.29% (across all public datasets). The study in [77] compared the performance of the DBN-RBM model using spectral features, extracted by FFT or WPD, and raw signal values. Study results showed that the classification accuracy of MI-EEG data, when spectral features were fed as inputs to DBN-RBM, was significantly better than that of raw signal values. In addition, the results showed that the frequency features of FFT outperformed the time–frequency features of WPD, implying that the time information of the frequency components for MI signals is not as discriminative as typical frequency information.

Overall, for the studies that used raw EEG signal values as inputs, the average accuracy based on the most common MI dataset (BCI-C IV-2a [131]) was 77.65%, compared with 77.02, 76.80, and 76.5% for the extracted features, spectral images, and topological maps inputs, respectively. The second most-used MI dataset (BCI-C IV-2b [101]) showed similar results, where the studies that used raw EEG signal values as inputs achieved an average accuracy of 83.22% compared with 81.56 and 82.07% for extracted features and spectral images inputs, respectively. This observation runs counter to the intuition that the more time put into the preprocessing steps, the higher the performance achieved. This leads to the unexpected inference that future research may maximize performance by feeding raw signal values directly into deep learning models without having to spend more time processing the input data.

7.3 Current trends in DL-based techniques

Based on the reviewed studies, 64% of all studies adopted CNN-based DL strategies and 15% proposed combining CNN with other DL models, such as recurrent (e.g., LSTM), generative (e.g., GAN), or representative (e.g., AE) models. The widespread use of CNN-based models can be justified by the following points. First, a CNN architecture is able to extract deep discriminative features and spatial patterns from MI-EEG data. Therefore, CNNs have been used for classification as well as, in some studies, for feature extraction. Second, CNNs have had considerable success in many fields, such as image and video processing; thus, CNNs are well-known and accessible (public codes). As a result, MI BCI researchers have a better chance of understanding and adapting CNNs in their studies. Third, EEG brain signals can be processed by a CNN as 2-D images in various forms, e.g., raw time-series (Time × Channel), spectral images, or topological maps. Fourth, CNNs have many variants that are suitable for a variety of MI-EEG tasks and forms.

Previous research has shown that CNNs outperform other deep learning strategies. In [90], the authors

compared the performance of LSTM and CNN networks and found that CNN significantly outperformed the LSTM network. However, the architecture of the LSTM model was basic, having a single hidden layer, and the input to the model was spectral images, which is more suitable for CNN than for LSTM. Another study [7] compared the performance of a hybrid CNN/RNN model with standard CNN and LSTM models. The results showed that the CNN model outperformed the other models with an accuracy of 84.24% compared with 77.72% and 66.2% for the hybrid CNN/RNN and LSTM models, respectively.

While the identified studies did not specifically compare different numbers of convolutional and fully connected layers in the CNN models, 86% of them used a number of convolutional layers ranging from one to five, with two and three convolutional layers being most commonly used. Some other studies have suggested very deep CNN models with more than ten convolutional layers. Li et al. [118] proposed a very deep CNN model with 31 convolutional layers, which was compared to different structures, while the number of convolutional layers was varied. The study concluded that the deeper CNN models achieved higher classification accuracy. For classifier layers (fully connected layers), 87% of the studies used one or two fully connected layers. However, the study in [118] showed that using up to three fully connected layers increased performance compared with using only one layer.

Regarding the activation function, the most commonly used activation functions in the convolutional layers were the rectified linear unit (ReLU) (65%) and the exponential linear unit (ELU) (26%). The study in [115] compared the performance of CNNs with ReLU and ELU and showed that ELU achieved better results for both time and accuracy. Other activation functions were used less frequently, such as leaky rectified linear unit (LReLU) (5%), scaled exponential linear unit (SELU) (1%), and hyperbolic tangent (tanh) (3%). Wang et al. [90] showed that SELU outperformed both ELU and ReLU. Given the large number of studies using ReLU, the recommendation is to employ ReLU as an activation function in the first construction of convolutional layers; then, the potential for performance improvement can be investigated using other activation functions, such as ELU or SELU.

Only 7% of all reviewed studies used RNN, and 8% of studies suggested combining RNN with CNN. This is less than was expected, given the RNN's proven effectiveness for learning time-series features. One explanation for these findings is that RNNs consume a great deal of time and memory, especially for long sequences. For instance, the MI-EEG signal with a typical duration of 4 s and a sampling frequency of 512 Hz contains 2048 time points. Experience shows [30] that for an EEG signal with 3000 samples, the RNN takes a training time 20 times longer than that taken by the CNN. The MLP is also unpopular due to its simplistic deep learning design and lower efficacy (e.g., nonlinear capability) compared with other techniques.

ALL RNN-based studies in this review used LSTM networks, except for one study that proposed the use of GRU. LSTM has been shown to be better than GRU in large datasets. The explanation for this is that LSTM has two control gates (forget and input gates) more than GRU and thus provides better nonlinear capability. GRU, on the other hand, has fewer gates, which means it has fewer parameters and is, therefore, less complex than LSTM and requires fewer data to generalize. The research in [70] compared the performance of recurrent GRU and LSTM networks in MI classification tasks using two different datasets with a different number of subjects and showed that GRU outperformed LSTM on both datasets. Generally, GRU and LSTM have similar performance. As a result, it is recommended to test both of them for any specific MI task and then determine which one works best.

The majority of RNN-based studies used two RNN layers, and one study explicitly investigated the effect of the number of RNN layers [80]. Wang et al. [80] compared the performance of the LSTM network with different numbers of LSTM layers and noticed a major improvement when they used two layers against a single layer, whereas the accuracy did not change significantly with the addition of a third LSTM layer.

DBN was employed in more than 7% of the reviewed papers. DBN, especially DBN-RBM, is the most common feature extraction model among representative models for two reasons: (1) it efficiently learns generative parameters that expose the relationship between variables in adjacent layers; (2) it allows the direct computation of the values of the variables inherent in hidden layers [30]. Most of the reviewed DBN studies used stacked RBMs (70%). In the study [51], the performance of AE-based and RBM-based DBN models was compared for MI classification using FFT features. The results showed the superiority of DBN-AE with 71% accuracy compared with 68.4% accuracy for DBN-RBM.

Generative models were rarely used in the reviewed studies. Data augmentation, e.g., VAE- or GAN-based, demonstrated improvements in classifier training ability and its performance [64, 91, 99]. Thus, this is a promising direction in future research.

Hybrid DL models were proposed in 15% of the reviewed publications. All hybrid models employed CNN as a powerful feature extractor. More than 53% of the hybrid models suggested combining CNN and RNN (especially LSTM). CNN-RNN hybrid models demonstrated a strong ability to extract both spatial and temporal features. Another type of combination of two

discriminative models is the fusion of CNN and MLP. However, CNN-MLP was adopted by only one study. Typically, in this type of fusion, CNN is used for feature extraction and MLP for classification. The fusion of representative and discriminative models was proposed in 27% of the hybrid models. This fusion makes sense because the former is used for feature extraction and the latter is used for classification. The most commonly used representative model is AE, whereas CNN is the most common discriminative model.

Hybrid models usually perform better than standalone DL models. A hybrid model that combines CNN and SAE architectures was compared with standard CNN and SAE models in [98]. The CNN's performance greatly surpassed that of SAE, while the CNN/SAE hybrid model had the best accuracy. Another study [56] compared the performance of a CNN model with a hybrid CNN/LSTM model and showed that the hybrid model achieved the best performance.

8 Challenges and future directions

Although deep learning has improved the decoding performance of MI-EEG signals, the real-world broad application of MI-based BCI systems remains limited by technical and usability challenges. These challenges are discussed in this section, along with potential research directions.

8.1 Robust and general framework

The classification performance and generalization ability of MI-based BCI systems remain limited, as shown in this review, and further research is needed. One possible research direction for improving classification accuracy is the hybrid fusion of different DL networks and multi-layer DL architectures. Another promising research direction is to develop a general deep learning model that can handle various MI-EEG signals regardless of the number of channels, time points (sampling rate), subject dependency, session dependency, etc. To achieve that, one possible solution is to utilize the attention mechanism to focus the deep model on the most important portions of the high-dimensional structure of the MI-EEG signal. The attention approach can be implemented based on attention scores or by different machine learning algorithms such as reinforcement learning. Reinforcement learning demonstrated good ability to identify the most important components through policy search [30]. In future research, CNNs could be used as a core method for feature learning at different levels, and they could be combined with appropriate attention strategies to form a robust and general classification system.

Generative models can also be used to augment the MI-EEG data and help enhance the generalizability of the MI classification system at the subject and session levels. Generative GAN models showed improvements in the MI classifier's training ability and consequently its performance [64, 91]. Hence, this is a promising research direction.

Other types of studies can be performed to interpret how deep learning networks distinguish the MI task-related features and discriminate them from artifacts and other parts of the EEG signal. These types of studies have not been conducted explicitly yet, although interpretation may be more crucial than classification performance.

8.2 Subject-independent classification

Most of the current studies focused on subject-dependent MI classification, in which training and test data are obtained from the same person. However, for real-world BCI applications, we need to develop a calibration-free BCI with subject-independent classification ability, e.g., one that does not need to use training samples from target subjects. When the subjects are well trained, MI BCI usually delivers good results. For an MI-based BCI to be controlled by a target user, the target user must complete a large number of training trials, making calibration time unacceptably long for real-world deployment. As a result, the research should focus on developing a zero-calibration BCI or reducing the calibration time with appropriate training techniques.

Building a customized model with transfer learning is one potential approach to achieving this objective. Using deep network adaptation is another possible solution [40]. Input data can be broken down into two components: a subject-dependent component that depends on the subject, and a subject-independent component that is common to all subjects. A multi-task hybrid model can operate simultaneously on two tasks, one focusing on individual identification and the other on class recognition.

8.3 Real-time employment

Real-time classification of MI-EEG signals, which includes online classification with very low latency, is one of the directions that should be investigated in future research. Most of the reviewed articles focused on offline MI task classification, as the dataset is collected and analyzed in offline mode. However, in real-world deployment, the MI BCI systems need to deal with live streams of EEG data and deliver real-time classification outcomes, which remains extremely difficult.

Live signals acquired from EEG sensors using an online procedure are unstable and noisier compared with offline mode for a variety of reasons, such as a subject's lack of concentration, or instability inherent in the device (for example, fluctuating sampling rate). According to experimental results, the performance of online EEG signal classification is significantly lower than its offline counterpart [30]. One of the future lines of research towards online and real-time classification is the development of a set of robust and low-latency algorithms to control the influence variables and extract the distinctive features inherent in the noisy online EEG signal. Post-processing approaches, such as voting and aggregation, can be used to increase classification accuracy by combining the results of several continuous samples. These approaches, however, would eventually lead to increased latency. Therefore, post-processing necessitates a compromise between low latency and high accuracy.

8.4 Real-environment employment

A major issue with BCIs is that nearly all BCI studies are performed in a controlled laboratory environment, regardless of the actual environment of the intended users. In the real world, various sensory stimuli in the surrounding environment, such as movements, sounds, and smells, as well as the variation in heart rate, cortisol, fatigue, and concentration level outside the laboratory may significantly affect the quality of brain waves [29]. Therefore, for real-world employment, researchers should understand the specific context in which the proposed system would be used when designing any BCI device. For instance, the design requirements for controlling devices at home using brain waves can differ from those for detecting a pilot's attention level during flight. As a result, it is critical to undertake a thorough review of the system's basic requirements, environmental factors, and target users during the system design phase.

9 Conclusion

In this survey, we summarized the deep learning networks for MI-EEG classification. Differing from traditional methods, DL can automatically learn high-level and latent complex features through deep architecture from raw MI-EEG signals, while eliminating time-consuming preprocessing and feature extraction. We analyzed the preprocessing strategy, input formulation, deep learning strategy, network structure, and performance evaluation of recent DL methods. CNN was the most frequently implemented method for MI classification in the reviewed papers. Raw signal data were used extensively with deep learning methods, with or without minimal preprocessing. In addition, we reviewed public MI-EEG datasets that could be used with deep learning techniques.

Acknowledgements The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number (DRI-KSU-1354).

Declarations

Conflict of interest The authors declare that they do not have any type of conflict of interest.

References

1. Alshehri F, Muhammad G (2021) A comprehensive survey of the Internet of Things (IoT) and AI-based smart healthcare. IEEE Access 9:3660–3678
2. Masud M, Gaba GS, Alqahtani S, Muhammad G, Gupta BB, Kumar P, Ghoneim A (2020) A lightweight and robust secure key establishment protocol for internet of medical things in COVID-19 patients care. IEEE Inter Things J
3. Muhammad G, Alshehri F, Karray F, El Saddik A., Alsulaiman M, Falk TH (2021). A comprehensive survey on multimodal medical signals fusion for smart healthcare systems. Inf Fusion 76:355–375
4. Cantillo-Negrete J, Carino-Escobar RI, Carrillo-Mora P, Elias-Vinas D, Gutierrez-Martinez J (2018) Motor imagery-based brain-computer interface coupled to a robotic hand orthosis aimed for neurorehabilitation of stroke patients. J Healthc Eng 2018:1–10
5. López-Larraz E, Sarasola-Sanz A, Irastorza-Landa N, Birbaumer N, Ramos-Murguialday A (2018) Brain-machine interfaces for rehabilitation in stroke: a review. NeuroRehabilitation 43(1):77–97
6. Al-Quraishi MS, Elamvazuthi I, Daud SA, Parasuraman S, Borboni A (2018) EEG-based control for upper and lower limb exoskeletons and prostheses: a systematic review. Sensors 18(10):3342
7. Tayeb Z et al (2019) Validating deep neural networks for online decoding of motor imagery movements from EEG signals. Sensors 19(1):210
8. Fernández-Rodríguez Á, Velasco-Álvarez F, Ron-Angevin R (2016) Review of real brain-controlled wheelchairs. J Neural Eng 13(6):61001
9. Tang X, Li W, Li X, Ma W, Dang X (2020) Motor imagery EEG recognition based on conditional optimization empirical mode decomposition and multi-scale convolutional neural network. Expert Syst Appl 149:113285
10. Li J, Liang J, Zhao Q, Li J, Hong K, Zhang L (2013) Design of assistive wheelchair system directly steered by human thoughts. Int J Neural Syst 23(03):1350013
11. Cao L, Xia B, Maysam O, Li J, Xie H, Birbaumer N (2017) A synchronous motor imagery based neural physiological paradigm for brain computer interface speller. Front Hum Neurosci 11:274
12. Das Chakladar D, Chakraborty S (2018) Multi-target way of cursor movement in brain computer interface using unsupervised learning. Biol Inspired Cogn Archit 25:88–100
13. Delorme A, Sejnowski T, Makeig S (2007) Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage 34(4):1443–1449

14. Jafarifarmand A, Badamchizadeh MA (2019) EEG artifacts handling in a real practical brain–computer interface controlled vehicle. IEEE Trans Neural Syst Rehabil Eng 27(6):1200–1208
15. Pawar D, Dhage S (2020) Feature extraction methods for electroencephalography based brain-computer interface: a review. IAENG Int J Comput Sci 47(3)
16. Djamal EC, Abdullah MY, Renaldi F (2017) Brain computer interface game controlling using fast fourier transform and learning vector quantization. J Telecommun Electron Comput Eng 9(2–5):71–74
17. Kousarrizi MRN, Ghanbari AA, Teshnehlab M, Shorehdeli MA, Gharaviri A (2009) Feature extraction and classification of EEG signals using Wavelet transform, SVM and artificial neural networks for brain computer interfaces. In: 2009 international joint conference on bioinformatics, systems biology and intelligent computing, pp 352–355
18. Wang L, Lan Z, Wang Q, Yang R, Li H (2019) ELM_Kernel and Wavelet packet decomposition based EEG classification algorithm. Autom Control Comput Sci 53(5):452–460
19. Ramoser H, Muller-Gerking J, Pfurtscheller G (2000) Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans Rehabil Eng 8(4):441–446
20. Zhang L, Wen D, Li C, Zhu R (2020) Ensemble classifier based on optimized extreme learning machine for motor imagery classification. J Neural Eng 17(2):26004
21. Wang K, Zhai DH, Xia Y (2019) Motor imagination EEG recognition algorithm based on DWT, CSP and extreme learning machine. In: 2019 Chinese control conference (CCC), pp 4590–4595
22. Jin Z, Zhou G, Gao D, Zhang Y (2018) EEG classification using sparse Bayesian extreme learning machine for brain–computer interface. Neural Comput Appl 32:1–9
23. Ang KK, Chin ZY, Wang C, Guan C, Zhang H (2012) Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. Front Neurosci 6:39
24. Chen CY, Wu CW, Lin CT, Chen SA (2014) A novel classification method for motor imagery based on brain-computer interface. In: 2014 International joint conference on neural networks (IJCNN), pp 4099–4102
25. Arvaneh M, Guan C, Ang KK, Quek C (2011) Optimizing the channel selection and classification accuracy in EEG-based BCI. IEEE Trans Biomed Eng 58(6):1865–1873
26. Samek W, Vidaurre C, Müller K-R, Kawanabe M (2012) Stationary common spatial patterns for brain–computer interfacing. J Neural Eng 9(2):26013
27. Samek W, Kawanabe M, Müller K-R (2013) Divergence-based framework for common spatial patterns algorithms. IEEE Rev Biomed Eng 7:50–72
28. Wu W, Chen Z, Gao X, Li Y, Brown EN, Gao S (2014) Probabilistic common spatial patterns for multichannel EEG analysis. IEEE Trans Pattern Anal Mach Intell 37(3):639–653
29. Rashid M et al (2020) Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review. Front Neurorobot 14:25
30. Zhang X, Yao L, Wang X, Monaghan JJM, Mcalpine D, Zhang Y (2020) A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers. J Neural Eng 18:031002
31. Altaheri H, Alsulaiman M, Muhammad G (2019) Date Fruit classification for robotic harvesting in a natural environment using deep learning. IEEE Access 7(1):117115–117133
32. Qamhan M, Altaheri H, Meftah AH, Muhammad G, Alotaibi YA (2021) Digital audio forensics: microphone and environment classification using deep learning. IEEE Access 9:62719–62733
33. Muhammad G, Hossain MS, Kumar N (2020) EEG-based pathology detection for home health monitoring. IEEE J Sel Areas Commun 39(2):603–610
34. Muhammad G, Alhamid MF, Long X (2019) Computing and processing on the edge: Smart pathology detection for connected healthcare. IEEE Netw 33(6):44–49
35. Muhammad G, Rahman SKMM, Alelaiwi A, Alamri A (2017) Smart health solution integrating IoT and cloud: a case study of voice pathology monitoring. IEEE Commun Mag 55(1):69–73
36. Lotte F et al (2018) A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. J Neural Eng 15(3):31005
37. Craik A, He Y, Contreras-Vidal JL (2019) Deep learning for electroencephalogram (EEG) classification tasks: a review. J Neural Eng 16(3):31001
38. Padfield N, Zabalza J, Zhao H, Masero V, Ren J (2019) EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges. Sensors 19(6):1423
39. Aggarwal S, Chugh N (2019) Signal processing techniques for motor imagery brain computer interface: a review. Array 1:100003
40. Wan Z, Yang R, Huang M, Zeng N, Liu X (2020) A review on transfer learning in EEG signal analysis. Neurocomputing 421:1–14
41. Lashgari E, Liang D, Maoz U (2020) Data augmentation for deep-learning-based electroencephalography. J Neurosci Methods 2020:108885
42. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6(7):e1000097
43. Millán JDR et al (2010) Combining brain–computer interfaces and assistive technologies: state-of-the-art and challenges. Front Neurosci 4:161
44. Greenfield LJ, Geyer JD, Carney PR (2012) Reading EEGs: a practical approach. Lippincott Williams and Wilkins, Philadelphia
45. Ball T, Kern M, Mutschler I, Aertsen A, Schulze-Bonhage A (2009) Signal quality of simultaneously recorded invasive and non-invasive EEG. Neuroimage 46(3):708–716
46. Kandel ER, Schwartz JH, Jessell TM, Siegelbaum S, Hudspeth AJ, Mack S (2000) Principles of neural science. McGraw-Hill, New York
47. CHB-MIT Scalp EEG Database. Available: https://archive.physionet.org/physiobank/charts/chbmit.png. (Accessed 12 Apr 2020)
48. Lacey S, Lawson R (2013) Multisensory imagery. Springer Science and Business Media, Berlin
49. Rezeika A, Benda M, Stawicki P, Gembler F, Saboor A, Volosyak I (2018) Brain–computer interface spellers: a review. Brain Sci 8(4):57
50. Lee MH et al (2019) EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. Gigascience 8(5):giz002
51. Hassanpour A, Moradikia M, Adeli H, Khayami SR, Shamsinejadbabaki P (2019) A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals. Expert Syst 36(6):e12494
52. Pfurtscheller G, Brunner C, Schlögl A, Da Silva FHL (2006) Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. Neuroimage 31(1):153–159
53. Wang Y, Nakanishi M, Zhang D (2019) EEG-based brain-computer interfaces, in neural interface: frontiers and applications. Springer, Berlin, pp 41–65
54. Zhao X, Zhang H, Zhu G, You F, Kuang S, Sun L (2019) A multi-branch 3D convolutional neural network for EEG-based

motor imagery classification. IEEE Trans Neural Syst Rehabil Eng 27(10):2164–2177
55. Avilov O, Rimbert S, Popov A, Bougrain L (2021) Optimizing motor intention detection with deep learning: towards management of intraoperative awareness. IEEE Trans Biomed Eng. https://doi.org/10.1109/TBME.2021.3064794
56. Zhu K, Wang S, Zheng D, Dai M (2019) Study on the effect of different electrode channel combinations of motor imagery EEG signals on classification accuracy. J Eng 2019(23):8641–8645
57. Lun X, Yu Z, Chen T, Wang F, Hou Y (2020) A simplified CNN classification method for MI-EEG via the electrode pairs signals. Front Hum Neurosci 14. https://doi.org/10.3389/fnhum.2020.00338
58. Liu T, Yang D (2021) A densely connected multi-branch 3D convolutional neural network for motor imagery EEG decoding. Brain Sci 11(2):197
59. Li Y, Yang H, Li J, Chen D, Du M (2020) EEG-based intention recognition with deep recurrent-convolution neural network: performance and channel selection by Grad-CAM. Neurocomputing 415:225–233
60. Yang J, Ma Z, Wang J, Fu Y (2020) A novel deep learning scheme for motor imagery EEG decoding based on spatial representation fusion. IEEE Access 8:202100–202110
61. Chu Y, Zhao X, Zou Y, Xu W, Han J, Zhao Y (2018) A decoding scheme for incomplete motor imagery EEG with deep belief network. Front Neurosci 12:680
62. Jeong J-H, Lee B-H, Lee D-H, Yun Y-D, Lee S-W (2020) EEG classification of forearm movement imagery using a hierarchical flow convolutional neural network. IEEE Access 8:66941–66950
63. Yang J, Yao S, Wang J (2018) Deep fusion feature learning network for MI-EEG classification. IEEE Access 6:79050–79059
64. Fahimi F, Dosen S, Ang KK, Mrachacz-Kersting N, Guan C (2020) Generative adversarial networks-based data augmentation for brain-computer interface. IEEE Trans Neural Netw Learn Syst 2020:1–13
65. Xu B et al (2018) Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification. IEEE Access 7:6084–6093
66. Ma X, Qiu S, Wei W, Wang S, He H (2019) Deep channel-correlation network for motor imagery decoding from the same limb. IEEE Trans Neural Syst Rehabil Eng 28(1):297–306
67. Alwasiti H, Yusoff MZ, Raza K (2020) Motor imagery classification for brain computer interface using deep metric learning. IEEE Access 8:109949–109963
68. Alazrai R, Abuhijleh M, Alwanni H, Daoud MI (2019) A deep learning framework for decoding motor imagery tasks of the same hand using EEG signals. IEEE Access 7:109612–109627
69. Gómez-Herrero G, et al. (2006) Automatic removal of ocular artifacts in the EEG without an EOG reference channel. In: Proceedings of the 7th nordic signal processing symposium-NORSIG 2006, pp 130–133
70. Luo T, Chao F (2018) Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network. BMC Bioinform 19(1):344
71. Olivas-Padilla BE, Chacon-Murguia MI (2019) Classification of multiple motor imagery using deep convolutional neural networks and spatial filters. Appl Soft Comput 75:461–472
72. Sakhavi S, Guan C, Yan S (2018) Learning temporal information for brain-computer interface using convolutional neural networks. IEEE Trans Neural Networks Learn Syst 29(11):5619–5629
73. Kwon OY, Lee MH, Guan C, Lee SW (2019) Subject-independent brain–computer interfaces based on deep convolutional neural networks. IEEE Trans Neural Networks Learn Syst 31(10):3839–3852
74. She Q, Hu B, Luo Z, Nguyen T, Zhang Y (2018) A hierarchical semi-supervised extreme learning machine method for EEG recognition. Med Biol Eng Comput 57(1):147–157
75. Taheri S, Ezoji M, Sakhaei SM (2020) Convolutional neural network based features for motor imagery EEG signals classification in brain–computer interface system. SN Appl Sci 2(4):1–12
76. Ma X, Wang D, Liu D, Yang J (2020) DWT and CNN based multi-class motor imagery electroencephalographic signal recognition. J Neural Eng 17(1):16073
77. Lu N, Li T, Ren X, Miao H (2016) A deep learning scheme for motor imagery classification based on restricted Boltzmann machines. IEEE Trans neural Syst Rehabil Eng 25(6):566–576
78. Xu J, Zheng H, Wang J, Li D, Fang X (2020) Recognition of EEG signal motor imagery intention based on deep multi-view feature learning. Sensors 20(12):3496
79. Huang W, Xue Y, Hu L, Liuli H (2020) S-EEGNet: electroencephalogram signal classification based on a separable convolution neural network with bilinear interpolation. IEEE Access 8:131636–131646
80. Wang P, Jiang A, Liu X, Shang J, Zhang L (2018) LSTM-based EEG classification in motor imagery tasks. IEEE Trans Neural Syst Rehabil Eng 26(11):2086–2095
81. Bang JS, Lee MH, Fazli S, Guan C, Lee SW (2021) Spatio-spectral feature representation for motor imagery classification using convolutional neural networks. IEEE Trans Neural Networks Learn Syst 2021:1–12
82. Xue J et al (2020) A multifrequency brain network-based deep learning framework for motor imagery decoding. Neural Plast 2020:1–11
83. Zhao X, Zhao J, Liu C, Cai W (2020) Deep neural network with joint distribution matching for cross-subject motor imagery brain-computer interfaces. Biomed Res Int 2020:1–15
84. Kumar S, Sharma A, Tsunoda T (2019) Brain wave classification using long short-term memory network based OPTICAL predictor. Sci Rep 9(1):1–13
85. Kumar S, Sharma R, Sharma A (2021) OPTICAL+: a frequency-based deep learning scheme for recognizing brain wave signals. PeerJ Comput Sci 7:e375
86. Cheng L, Li D, Yu G, Zhang Z, Li X, Yu S (2020) A motor imagery EEG feature extraction method based on energy principal component analysis and deep belief networks. IEEE Access 8:21453–21472
87. Zhang R, Zong Q, Dou L, Zhao X (2019) A novel hybrid deep learning scheme for four-class motor imagery classification. J Neural Eng 16(6):66004
88. Zhang R, Zong Q, Dou L, Zhao X, Tang Y, Li Z (2021) Hybrid deep neural network using transfer learning for EEG motor imagery decoding. Biomed Signal Process Control 63:102144
89. Uktveris T, Jusas V (2017) Application of convolutional neural networks to four-class motor imagery classification problem. Inf Technol Control 46(2):260–273
90. Wang Z, Cao L, Zhang Z, Gong X, Sun Y, Wang H (2018) Short time Fourier transformation and deep neural networks for motor imagery brain computer interface recognition. Concurr Comput Pract Exp 30(23):e4413
91. Zhang K et al (2020) Data augmentation for motor imagery signal classification based on a hybrid neural network. Sensors 20(16):4485
92. Shajil N, Mohan S, Srinivasan P, Arivudaiyanambi J, Murrugesan AA (2020) Multiclass classification of spatially filtered motor imagery EEG signals using convolutional neural network for BCI based applications. J Med Biol Eng 40(5):663–672

93. Rong Y, Wu X, Zhang Y (2020) Classification of motor imagery electroencephalography signals using continuous small convolutional neural network. Int J Imaging Syst Technol 30(3):653–659
94. Roy S, Chowdhury A, McCreadie K, Prasad G (2020) Deep learning based inter-subject continuous decoding of motor imagery for practical brain-computer interfaces. Front Neurosci 14. https://doi.org/10.3389/fnins.2020.00918
95. Miao M, Hu W, Yin H, Zhang K (2020) Spatial-frequency feature learning and classification of motor imagery EEG based on deep convolution neural network. Comput Math Methods Med 2020:1–13
96. Li F, He F, Wang F, Zhang D, Xia Y, Li X (2020) A novel simplified convolutional neural network classification algorithm of motor imagery EEG signals based on deep learning. Appl Sci 10(5):1605
97. Kant P, Laskar SH, Hazarika J, Mahamune R (2020) CWT based transfer learning for motor imagery classification for brain computer interfaces. J Neurosci Methods 345:108886
98. Tabar YR, Halici U (2016) A novel deep learning approach for classification of EEG motor imagery signals. J Neural Eng 14(1):16003
99. Dai M, Zheng D, Na R, Wang S, Zhang S (2019) EEG classification of motor imagery using a novel deep learning framework. Sensors 19(3):551
100. Zhang D, Chen K, Jian D, Yao L (2020) Motor imagery classification via temporal attention cues of graph embedded EEG signals. IEEE J Biomed Heal Inform 24(9):2570–2579
101. Leeb R, Brunner C, Müller-Putz G, Schlögl A, Pfurtscheller G (2008) BCI Competition 2008–Graz data set B. Inst Knowl Discov Graz Univ Technol 16:1–6
102. Blankertz B et al (2004) The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. IEEE Trans Biomed Eng 51(6):1044–1051
103. Deng X, Zhang B, Yu N, Liu K, Sun K (2021) Advanced TSGL-EEGNet for motor imagery EEG-based brain-computer interfaces. IEEE Access 9:25118–25130
104. Fan CC, Yang H, Hou ZG, Ni ZL, Chen S, Fang Z (2021) Bilinear neural network with 3-D attention for brain decoding of motor imagery movements from the human EEG. Cogn Neurodyn 15(1):181–189
105. Roots K, Muhammad Y, Muhammad N (2020) Fusion convolutional neural network for cross-subject EEG motor imagery classification. Computers 9(3):72
106. Li D, Xu J, Wang J, Fang X, Ying J (2020) 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 28:2615–2626
107. Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ (2018) EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. J Neural Eng 15(5):56013
108. Amin SU, Alsulaiman M, Muhammad G, Bencherif MA, Hossain MS (2019) Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification. IEEE Access 7:18940–18950
109. Dose H, Møller JS, Iversen HK, Puthusserypady S (2018) An end-to-end deep learning approach to MI-EEG signal classification for BCIs. Expert Syst Appl 114:532–542
110. Tang Z, Li C, Sun S (2017) Single-trial EEG classification of motor imagery using deep convolutional neural networks. Optik (Stuttg) 130:11–18
111. Dai G, Zhou J, Huang J, Wang N (2020) HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification. J Neural Eng 17(1):16025
112. Lee B-H, Jeong J-H, Lee S-W (2020) SessionNet: feature similarity-based weighted ensemble learning for motor imagery classification. IEEE Access 8:134524–134535
113. Wu H et al (2019) A parallel multiscale filter bank convolutional neural networks for motor imagery EEG classification. Front Neurosci 13:1275
114. Zhang C, Kim Y-K, Eskandarian A (2021) EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification. J Neural Eng 18(4):46014
115. Amin SU, Alsulaiman M, Muhammad G, Mekhtiche MA, Hossain MS (2019) Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion. Futur Gener Comput Syst 101:542–554
116. Xu M et al (2020) Learning EEG topographical representation for classification via convolutional neural network. Pattern Recognit 105:107390
117. Liao JJ, Luo JJ, Yang T, So RQY, Chua MCH (2020) Effects of local and global spatial patterns in EEG motor-imagery classification using convolutional neural network. Brain Computer Interfaces 7(3–4):47–56
118. Li M-A, Han J-F, Duan L-J (2019) A novel MI-EEG imaging with the location information of electrodes. IEEE Access 8:3197–3211
119. Collazos-Huertas DF, Álvarez-Meza AM, Acosta-Medina CD, Castaño-Duque GA, Castellanos-Dominguez G (2020) CNN-based framework using spatial dropping for enhanced interpretation of neural activity in motor imagery classification. Brain Inform 7(1):1–13
120. Hou Y, Zhou L, Jia S, Lun X (2020) A novel approach of decoding EEG four-class motor imagery tasks via scout ESI and CNN. J Neural Eng 17(1):16048
121. Liu X, Shen Y, Liu J, Yang J, Xiong P, Lin F (2020) Parallel spatial–temporal self-attention CNN-based motor imagery classification for BCI. Front Neurosci 14. https://doi.org/10.3389/fnins.2020.587520
122. Amin SU, Altaheri H, Muhammad G, Alsulaiman M, Abdul W (2021) Attention based inception model for robust EEG motor imagery classification. In: 2021 IEEE international instrumentation and measurement technology conference (I2MTC), pp 1–6. https://doi.org/10.1109/I2MTC50364.2021.9460090
123. Zhu X, Li P, Li C, Yao D, Zhang R, Xu P (2019) Separated channel convolutional neural network to realize the training free motor imagery BCI systems. Biomed Signal Process Control 49:396–403
124. Musallam YK et al (2021) Electroencephalography-based motor imagery classification using temporal convolutional network fusion. Biomed Signal Process Control 69:102826
125. Riyad M, Khalil M, Adib A (2021) MI-EEGNET: A novel convolutional neural network for motor imagery classification. J Neurosci Methods 353:109037
126. Li D, Wang J, Xu J, Fang X (2019) Densely feature fusion based on convolutional neural networks for motor imagery EEG classification. IEEE Access 7:132720–132730
127. Ha K-W, Jeong J-W (2021) Temporal pyramid pooling for decoding motor-imagery EEG signals. IEEE Access 9:3112–3125
128. Zhang K, Robinson N, Lee S-W, Guan C (2021) Adaptive transfer learning for EEG motor imagery classification with deep convolutional neural network. Neural Netw 136:1–10
129. Zhao H, Zheng Q, Ma K, Li H, Zheng Y (2020) Deep representation-based domain adaptation for nonstationary EEG classification. IEEE Trans Neural Networks Learn Syst 32:535–545
130. Xu G et al (2019) A deep transfer convolutional neural network framework for EEG signal classification. IEEE Access 7:112767–112776

131. Brunner C, Leeb R, Müller-Putz G, Schlögl A, Pfurtscheller G (2008) BCI Competition 2008–Graz data set A. Inst Knowl Discov Graz Univ Technol 16:1–6
132. Cho H, Ahn M, Ahn S, Kwon M, Jun SC (2017) EEG datasets for motor imagery brain–computer interface. Gigascience 6(7):gix034
133. Blankertz B, Dornhege G, Krauledat M, Müller K-R, Curio G (2007) The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects. Neuroimage 37(2):539–550
134. Kingma DP, Welling M (2013) Auto-encoding variational bayes. arXiv preprint, http://arxiv.org/abs/1312.6114
135. Li Y, Zhang X-R, Zhang B, Lei M-Y, Cui W-G, Guo Y-Z (2019) A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding. IEEE Trans Neural Syst Rehabil Eng 27(6):1170–1180
136. Blankertz B et al (2006) The BCI competition III: validating alternative approaches to actual BCI problems. IEEE Trans neural Syst Rehabil Eng 14(2):153–159
137. Wang L, Huang W, Yang Z, Zhang C (2020) Temporal-spatial-frequency depth extraction of brain-computer interface based on mental tasks. Biomed Signal Process Control 58:101845
138. Freer D, Yang G-Z (2020) Data augmentation for self-paced motor imagery classification with C-LSTM. J Neural Eng 17(1):16041
139. Goldberger AL et al (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 101(23):e215–e220
140. Xiaoling L (2020) Motor imagery-based EEG signals classification by combining temporal and spatial deep characteristics. Int J Intell Comput Cybern 13:437–453
141. Zhang K et al (2020) Instance transfer subject-dependent strategy for motor imagery signal classification using deep convolutional neural networks. Comput Math Methods Med 2020:1–10
142. Ofner P, Schwarz A, Pereira J, Müller-Putz GR (2017) Upper limb movements can be decoded from the time-domain of low-frequency EEG. PLoS ONE 12(8):e0182578
143. Chen J, Yu Z, Gu Z, Li Y (2020) Deep temporal-spatial feature learning for motor imagery-based brain-computer interfaces. IEEE Trans Neural Syst Rehabil Eng 28(11):2356–2366
144. Steyrl D, Scherer R, Förstner O, Müller-Putz GR (2014) Motor imagery brain-computer interfaces: random forests vs regularized LDA-non-linear beats linear. In: Proceedings of the 6th international brain-computer interface conference, pp 241–244
145. Ma X, Qiu S, He H (2020) Multi-channel EEG recording during motor imagery of different joints from the same limb. Sci Data 7(1):1–9
146. Lee HK, Choi Y-S (2019) Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface. Entropy 21(12):1199
147. Ortiz-Echeverri CJ, Salazar-Colores S, Rodríguez-Reséndiz J, Gómez-Loenzo RA (2019) A new approach for motor imagery classification based on sorted blind source separation, continuous wavelet transform, and convolutional neural network. Sensors 19(20):4541
148. Chaudhary S, Taran S, Bajaj V, Sengur A (2019) Convolutional neural network based approach towards motor imagery tasks EEG signals classification. IEEE Sens J 19(12):4494–4500
149. Tang X-L, Ma W-C, Kong D-S, Li W (2019) Semisupervised deep stacking network with adaptive learning rate strategy for motor imagery EEG recognition. Neural Comput 31(5):919–942
150. Zhang Z et al (2019) A novel deep learning approach with data augmentation to classify motor imagery signals. IEEE Access 7:15945–15954
151. Tang X, Zhang N, Zhou J, Liu Q (2017) Hidden-layer visible deep stacking network optimized by PSO for motor imagery EEG recognition. Neurocomputing 234:1–10
152. Deng L, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: IEEE conference on computer vision and pattern recognition, pp 248–255
153. Altaheri H, Alsulaiman M, Muhammad G, Amin SU, Bencherif M, Mekhtiche M (2019) Date fruit dataset for intelligent harvesting. Data Br 26:104514
154. Alsulaiman M, Muhammad G, Bencherif MA, Mahmood A, Ali Z (2013) KSU rich Arabic speech database. Information 16(6B):4231–4253
155. Graz University of Technology (2021) Data sets-BNCI Horizon 2020. Available: http://bnci-horizon-2020.eu/database/data-sets. Accessed 05 Feb 2021
156. Lotte F (2021) Fabien Lotte's professional homepage-links. Available: https://sites.google.com/site/fabienlotte/bci-community/links?authuser=0#h.p_ID_172. Accessed 05 Feb 2021
157. Scherer R et al (2015) Individually adapted imagery improves brain-computer interface performance in end-users with disability. PLoS ONE 10(5):e0123727
158. Kaya M, Binli MK, Ozbay E, Yanar H, Mishchenko Y (2018) A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. Sci Data 5:180211
159. Brodu N, Lotte F, Lécuyer A (2012) Exploring two novel features for EEG-based brain–computer interfaces: multifractal cumulants and predictive complexity. Neurocomputing 79:87–94
160. Ramos-Murguialday A et al (2013) Brain–machine interface in chronic stroke rehabilitation: a controlled study. Ann Neurol 74(1):100–108
161. Zhang X, Yao L, Sheng QZ, Kanhere SS, Gu T, Zhang D (2018) Converting your thoughts to texts: enabling brain typing via deep feature learning of EEG signals. In: 2018 IEEE international conference on pervasive computing and communications (PerCom), pp 1–10
162. Van Erp J, Lotte F, Tangermann M (2012) Brain-computer interfaces: beyond medical applications. Computer (Long Beach Calif) 45(4):26–34
163. Yuste R et al (2017) Four ethical priorities for neurotechnologies and AI. Nat News 551(7679):159
164. LaFleur K, Cassady K, Doud A, Shades K, Rogin E, He B (2013) Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface. J Neural Eng 10(4):46003
165. Yu Y et al (2016) Toward brain-actuated car applications: Self-paced control with a motor imagery-based brain-computer interface. Comput Biol Med 77:148–155
166. Zhang X, Yao L, Huang C, Sheng QZ, Wang X (2017) Intent recognition in smart living through deep recurrent neural networks. In: International conference on neural information processing, pp 748–758
167. Li T, Zhang J, Xue T, Wang B (2017) Development of a novel motor imagery control technique and application in a gaming environment. Comput Intell Neurosci 2017:1–16
168. Kreilinger A, Hiebel H, Müller-Putz GR (2015) Single versus multiple events error potential detection in a BCI-controlled car game with continuous and discrete feedback. IEEE Trans Biomed Eng 63(3):519–529
169. Zhang X, Yao L, Kanhere SS, Liu Y, Gu T, Chen K (2018) Mindid: Person identification from brain waves through attention-based recurrent neural network. Proc ACM Interactive Mobile Wearable Ubiquitous Technol 2(3):1–23

170. Zhang X, Yao L, Huang C, Gu T, Yang Z, Liu Y (2017) DeepKey: an EEG and gait based dual-authentication system. arXiv Preprint, http://arxiv.org/abs/1706.01606

Authors and Affiliations

Hamdi Altaheri[1,4] · Ghulam Muhammad[1,4] · Mansour Alsulaiman[1,4] · Syed Umar Amin[1,4] · Ghadir Ali Altuwaijri[1,2,4] · Wadood Abdul[1,4] · Mohamed A. Bencherif[1,4] · Mohammed Faisal[3,4]

✉ Ghulam Muhammad
ghulam@ksu.edu.sa

Hamdi Altaheri
haltaheri@ksu.edu.sa

Mansour Alsulaiman
mmalsulaiman@ksu.edu.sa

Syed Umar Amin
samin@ksu.edu.sa

Ghadir Ali Altuwaijri
ga.altuwaijri@mu.edu.sa

Wadood Abdul
aabdulwaheed@ksu.edu.sa

Mohamed A. Bencherif
mabencherif@ksu.edu.sa

Mohammed Faisal
mfaisal@ksu.edu.sa

[1] Department of Computer Engineering, College of Computer and Information Sciences (CCIS), King Saud University, Riyadh 11543, Saudi Arabia

[2] Computer Sciences and Information Technology College, Majmaah University, Majmaah, Saudi Arabia

[3] College of Applied Computer Sciences, King Saud University, Riyadh, Saudi Arabia

[4] Center of Smart Robotics Research, CCIS, King Saud University, Riyadh, Saudi Arabia