



Deep learning for motor imagery EEG-based classification: A review

Ali Al-Saegh ^{*}, Shefa A. Dawwd, Jassim M. Abdul-Jabbar

Computer Engineering Department, College of Engineering, University of Mosul, Mosul, Iraq



ARTICLE INFO

Keywords:

BCI
Deep learning
Deep neural network
EEG
Motor imagery

ABSTRACT

Objectives: The availability of large and varied Electroencephalogram (EEG) datasets, rapidly advances and innovations in deep learning techniques, and highly powerful and diversified computing systems have all permitted to easily analyzing those datasets and discovering vital information within. However, the classification process of EEG signals and discovering vital information should be robust, automatic, and with high accuracy. Motor Imagery (MI) EEG has attracted us due to its significant applications in daily life.

Methods: This paper attempts to achieve those goals throughout a systematic review of the state-of-the-art studies within this field of research. The process began by intensely surfing the well-known specialized digital libraries and, as a result, 40 related papers were gathered. The papers were scrutinized upon multiple noteworthy technical issues, among them deep neural network architecture, input formulation, number of MI EEG tasks, and frequency range of interest.

Outcomes: Deep neural networks build robust and automated systems for the classification of MI EEG recordings by exploiting the whole input data throughout learning salient features. Specifically, convolutional neural networks (CNN) and hybrid-CNN (h-CNN) are the dominant architectures with high performance in comparison to public datasets with other types of architectures. The MI related datasets, input formulation, frequency ranges, and preprocessing and regularization methods were also reviewed.

Inferences: This review gives the required preliminaries in developing MI EEG-based BCI systems. The review process of the published articles in the last five years aims to help in choosing the appropriate deep neural network architecture and other hyperparameters for developing those systems.

1. Introduction

Normally the process of analyzing any physiological signal is basically about a practitioner to inspect that signal, then to give a comment or a decision about it. This process becomes slow when more and more signals are to be analyzed. Moreover, those decisions may be accompanied by the suspicion which in turn highly dependent on the

practitioner's mood and whether he/she suffers from tiredness. The physiological Electroencephalogram (EEG) recordings have been intensively investigated [1–5] due to their wide range of applications such as neuromarketing [6], neuroentertainment [7], epileptic seizure detection [8], Person identification and authentication [9–11], and social interaction [12], among others. The research community of brain-computer interface (BCI) systems has been attracted by a variety

Abbreviation: 1d-AX, OneDimension-Aggregate Approximation; AAR, AutomaticArtifact Removal; ADAM, Adaptive Moment Estimation; AE, Autoencoder; BCI, Brain-Computer Interface; BGRU, BidirectionalGated Recurrent Unit; BSS, Blind Source Separation; CNN, Convolutional Neural Network; CPU, Central Processing Unit; CSP, Common Spatial Pattern; CWD, Choi-Willimas Distribution; CWT, Continuous Wavelet Transform; DBN, Deep Belief Network; DL, Deep Learning; DNN, Deep Neural Network; EOG, Electrooculogram; ERD, Event-Related Desynchronization; ERP, Event-Related Potential; ERS, Event-Related Synchronization; EEG, Electroencephalography; ELM, Extreme Learning Machine; ELU, Exponential Linear Unit; EMD, Empirical Mode Decomposition; FBCSP, Filter Bank Common Spatial Pattern; FC, Fully Connected; FFT, Fast Fourier Transform; fNIRS, Functional Near-Infrared Spectroscopy; GD, Gradient Descent; GPU, Graphical Processor Unit; GRU, Gated Recurrent Unit; HGD, High-Gamma Dataset; ICA, IndependentComponent Analysis; LSTM, Long Short-Term Memory; moVEP, Motion Onset Visual Evoked Potential; PCA, Principal Component Analysis; PSD, Power Spectral Density; PSO, Particle Swarm Optimization; RBM, Restricted Boltzmann Machine; ReLU, Rectified Linear Unit; RNN, Recurrent Neural Network; SAE, Stacked Autoencoders; SELU, Scaled Exponential Linear Unit; SGD, Stochastic Gradient Descent; STFT, Short-Time Fourier Transform; SSVEP, Steady-StateVisual Evoked Potentials; ULM, Upper Limb Movement; VAE, Variational Autoencoder; VEP, Visually Evoked Potential; WPD, Wavelet Package Decomposition.

* Corresponding author.

E-mail address: ali.alsaegh@uomosul.edu.iq (A. Al-Saegh).

of EEG signals such as the steady-state visual evoked potentials (SSVEP), P300 evoked potentials, motion onset visual evoked potential (moVEP), motor imagery (MI), etc [13,14]. These types of EEG signals facilitate developing EEG-based BCI systems due to their noninvasive nature of recording [15], which makes the process of collecting data much easier.

Nowadays, MI EEG-based BCI is a promising technology due to its enormous domain in both medical and non-medical implementations. The MI task is accomplished by imagining performing a specific task without actually performing it [13]. The widely used MI tasks in researches are the imaginations of the right hand, left hand, right foot, left foot, both feet, and tongue; many other tasks are also under research like those movements related to the elbow, fists, and fingers. The MI-based BCI application involves clarification of the EEG signals and the determination of responses to those signals in real-time. Usually, analyzing EEG signals encountered the curse of dimensionality problem [16] arising from non-stationarity of signals, multi-channel recording paradigm, channel correlation, and the existence of noise and artifacts. Hence, robust and rapid systems are required for uncovering latent variables from those physiological signals. Specifically, deployment of MI EEG signals in real-world systems faces some challenges [17] such as a) EEG sensors need to be technically developed in order to establish more user-friendly equipment, b) signal processing techniques need to be improved such as sampling rate and classification methods, c) choice of technology needs to be considered carefully to ensure fast, reliable, robust, cheap, wearable, and portable end-user device.

Researchers used traditional neural networks (NNs) for a long time for automating different tasks. However, with the exponential growth of all digital data forms (time-series signals, images, and videos); it is soon realized the inadequacy of those neural networks to cover all states of diversity within a large amount of data [18]. Recently, a revolution began in adopting deep learning (DL) techniques for analyzing large-scale datasets [17]. Whereby DL techniques are capable of learning a large number of the required parameters for uncovering valuable and in-depth information more robustly than traditional NNs. Moreover, DL does not rely on hand-engineered methods of extracting features that are time-consuming, subject-dependent [19], and suffer from losing information [20]. Deep neural networks (DNNs) involve an extremely large number, tens of millions, of learnable parameters as well as intense floating-point matrix multiplication [21–23]. This makes a heavy computation load on the computing system. Usually, researches promote incorporating graphical processing units (GPUs) to tackle up that heavy computation problem.

DNNs are trained by examples, this paradigm of training requires the training dataset to be large enough, has diverse examples, and not biased to specific examples. The current main challenge that restricts researchers throughout using DL by MI EEG signals is the unavailability of large enough datasets. Whereby, the available MI EEG datasets are recorder for several subjects and contain a limited number of MI trials. However, there exist several methods to tackle such a problem; further discussion regarding such methods are included in this paper.

Plenty of studies have shown the capability of DNNs in diagnosing the physiological signals rapidly and reliably [24]. During the last decade, numerous researchers have employed DNNs in analyzing MI signals [25] due to their wide range of applications such as: controlling wheelchairs or robots [26], helping paralyzed patients to walk [27], upper limb exoskeleton [28], and many others [17]. However, some inherent drawbacks within MI signals make the analysis process not that straightforward. Whereby, MI signals have very small amplitude and rather low signal-to-noise ratio and they are usually contaminated with artifacts [29,30]. Indeed, such properties rise different challenges throughout working on such types of signals.

Li et al. [31] used the convolutional neural network (CNN) to classify MI signals. They first invested primary features by extracting the dependency of EEG channels and the temporal features as well. Then CNN is adopted for extracting high-level features. Alazrai et al. [13] applied Choi-Williams distribution (CWD) transformation on EEG signals in

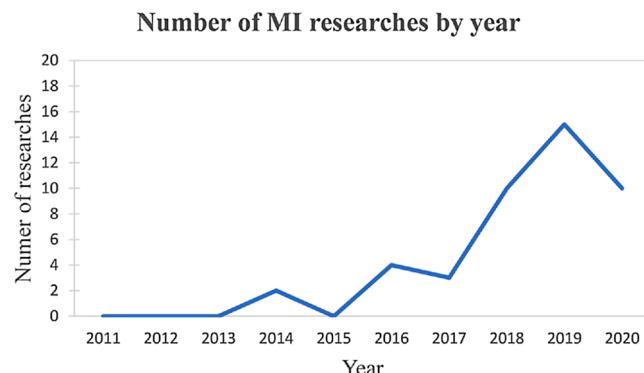


Fig. 1. Number of researches by year. The keywords (Motor Imagery OR MI) AND (deep learning OR deep neural network) were used to search within the titles of the researches using Google scholar search engine.

order to formulate 2D images as input to the CNN. Those images reflect the energy distribution over both time and frequency domains. Lu et al. [32] attained the frequency representation of EEG recordings using Fast Fourier transform (FFT) and wavelet package decomposition (WPD) separately for comparison purposes. Then deep belief network (DBN) that is composed of restricted Boltzmann machines (RBM) is trained using the frequency domain data. Xu et al. [18] used transfer learning for training their CNN that was proposed for the classification of MI signals. They got the benefit of the VGG-16 (visual geometry group) pre-trained network by means of transferring its parameters to their proposed CNN model. Zhao et al. [33] proposed a multi-branch CNN for the classification of MI signals. In their network, the EEG data is applied to three separated CNNs in parallel fashion then quantifying the final classification result based on summing the results retained by the three CNNs. Tabar and Halici [34] suggested a deep network that is composed of CNN and stacked autoencoders (SAE). They used the short-time Fourier transform (STFT) to construct 2D images for training their network. The features within the MI signals were extracted by the CNN then they classified using the SAE. Lee and Choi [15] used the continuous wavelet transform (CWT) to construct 2D images for training a CNN model.

The mentioned studies are just like a simple snapshot of the research field; it gives an impression about the ongoing exploration of the most suitable decoding algorithms for MI EEG data. Fig. 1 shows the number of MI related researches during the last ten years. Google Scholar search engine was used for establishing the curve by searching only within the titles of articles for the following keywords: (Motor Imagery OR MI) AND (deep learning OR deep neural network). The figure shows the increase in the number of researches in the last few years. Note that the stated number for the year 2020 is the search result for the first three months only.

Several review papers within the field of DL have been published recently, the following lines give a brief overview of their main contributions. Then we show our main contribution of this paper. Paper [19] overviews several medical imaging techniques and the use of different DL methods for the classification, segmentation, and object detection of those images. Paper [24] presents four types of physiological signals namely Electromyogram (EMG), Electroencephalogram (EEG), Electrocardiogram (ECG), and Electrooculogram (EOG); then it summarizes the DL algorithms used for the classification of those physiological signals related to a specific application. Paper [35] reviews the recent different architecture types of DNNs and presents state-of-the-art architectures that have been trained on natural images. Paper [36] reviews some of the related topics to the DL field of study namely types of DL mechanisms, DL platforms, and applications of DL. Paper [37] reviews different tasks carried out by EEG and discusses the possible DNN architecture for each of them.

The use of DL for the classification of MI EEG signals particularly

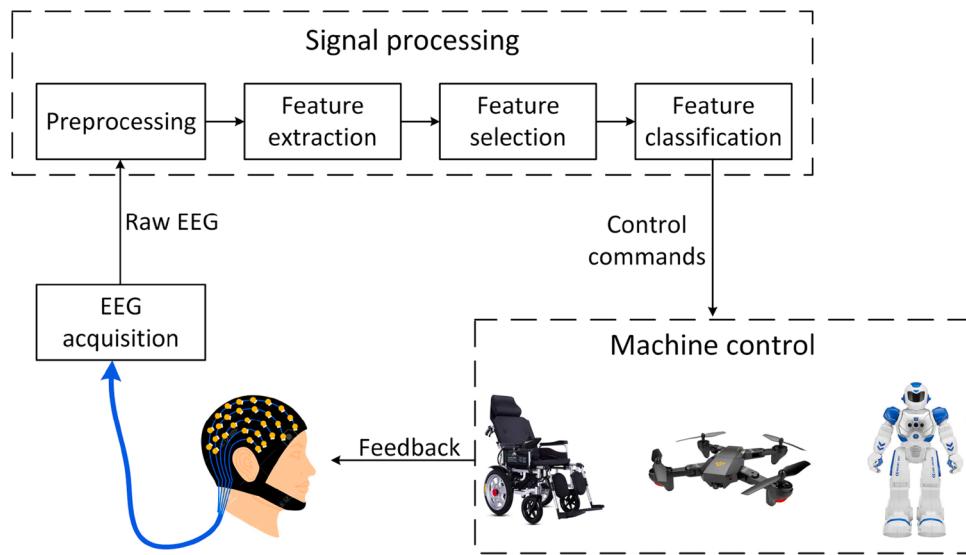


Fig. 2. A block diagram for the general BCI system.

arises these essential questions: (1) Which deep neural network architecture is best suited for the classification process? (2) Which structure of input data has a more positive impact on deep learning? (3) What frequency range must be considered during the analysis? To the best of our knowledge, this is the first literature review on the use of DNNs for the classification of MI EEG data. Whereby, this paper discusses in a specific manner the MI EEG signals and states their available different datasets, the used DNN architectures for the classification of those signals, the most important frequency bands related to MI tasks, the used regularization methods, input formulation to the DNNs, and the used DL platforms. Hence, the paper could be a good starting point before diving into the analysis of EEG and especially the MI signals.

The rest of this paper is organized as follow: Section 2 provides background information regarding two main subjects; first, the BCI system including EEG recording process, preprocessing methods, and feature identification; and second the DL including DNN architectures, input formulation, types of learning, and regularization methods. Section 3 presents the strategy for collecting the reviewed articles and the types of captured data from those articles. Section 4 presents the extracted data as the results of the review process and section 5 discusses those results. Section 6 concludes the paper.

2. Background

This section puts on view details about fundamental concepts related to the research field in scope. First, we give details about BCI systems, EEG data and its recording process, and preprocessing techniques that possibly need to be applied to eliminate undesirable things from the raw data, and features identification. Then, we give details about DNN architectures (CNN, RNN, AE, and DBN), input formulations (calculated features, time-series, and images), DL and its frameworks, and regularization methods.

2.1. MI EEG-based BCI system

A general methodology of EEG-based BCI systems is described in the following first subsection. Then more details about the BCI parts are given in the following subsections. Details concerning MI EEG-based BCI systems are given wherever it is required.

2.1.1. BCI systems

Neurophysiological signals necessitate well-trained specialists and a considerable amount of time for reliable diagnostics which is considered

a constraint. Machine learning techniques can automate such a process and hasten it as well as delivering much more accurate results [38]. Brain-computer interface (BCI), also called brain-machine interface (BMI), is a computer-based technology that takes knowledge and methods from neuroscience, signal processing, and machine learning. The main idea of BCI is interpreting the user's intent in order to take control of an external device like an artificial arm. Nowadays, BCI is intensively investigating by researchers because of its valuable achievements during the last decade [39]. Among the fields that BCI found its way into are neuroscience, industry, smart home, security, art, and entertainment [40–42].

A BCI system can be depicted in six main processing phases: (i) collecting data i.e. recording private datasets or use publicly available ones, (ii) preprocessing the raw data to be clean from background noise, (iii) extracting application-specific features from the clean version of data, (iv) selecting the most salient features from the extracted ones, (v) classifying the obtained features to conclude decisions and send commands to an end machine, and (vi) providing feedback to the user based on the concluded decisions after executing the commands by the machine [43]. Fig. 2. depicts the overall methodology of a BCI system. Sometimes BCI systems do not involve feature selection methods to avoid more computations, however, those systems that involve a selection procedure seek salient features to achieve higher accuracy [44]. Many researchers tried to improve the performance of those phases and especially the preprocessing, feature extraction, and classification phases [17,45]. In fact, the feature extraction phase represents the heart of each BCI system as it draws the line for the classification phase. A large number of feature extraction techniques exist where each of them is based on specific information in data such as time, frequency, spatial, or a combination of them [17]. Recent studies [19,46,47] proved that those hand-engineered feature extraction techniques deliver not that robust BCI system, which is extremely relying on the elected features.

BCI systems can be implemented using electrooculogram (EOG) [48], functional near-infrared spectroscopy (fNIRS) [49], EEG [50], electrocorticography (ECoG) [51], or any other neurophysiological signals. Nevertheless, EEG is non-invasive, cheap, portable, and with high temporal resolution technology [52]. Hence, it is more nominated than other types of data to be used in BCI implementations. Two main categories of EEG-based BCI systems exist: evoked and spontaneous. As the names infer, the evoked systems require external stimulation while the spontaneous ones do not. Different stimulations can be used in evoked systems like visually, auditory, or sensory. Visually evoked potential (VEP) and event-related potential (ERP) are two main classes of

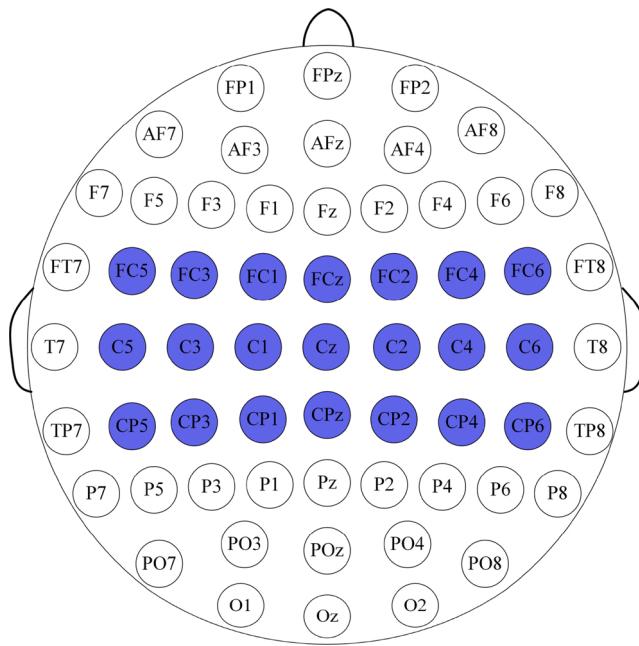


Fig. 3. EEG electrodes placement according to the 10-20 system. The related motor imagery electrodes are identified in blue color.

evoked systems and both of them are being employed in many applications. Motor imagery (MI) signals are of spontaneous type because they are generated based on a user's imagination of performing a task [17]. This spontaneity property is considered interesting whereby the MI technique permits both healthy and disabled users to streamline brain signals with no need for an external stimulus.

Dealing with high-dimensional MI signals is problematic due to the following reasons that are related to the signals: non-linearity, non-stationarity, weakness of signal-to-noise ratio, and high complexity [45]. Moreover, the statistical distribution of MI data is highly variable across subject-to-subject and session-to-session because there exist alternative physiological and psychological features for each person at each time. Hence, using hand-engineered feature extraction techniques with MI data necessitates a calibration process in order to train the decoder [53,54]. The calibration process takes about thirty minutes to start recording sufficient data and requires the assistance of an expert as well; in the end, the subsequent BCI system will be subject-dependent [53], i.e. not general. Accordingly, the mission of MI related feature extraction and classification is challenging, which in turn leads to suspicion in reading brain dynamics.

The emerging DL techniques have proven their ability to tackle the mentioned problems and challenges [37,53]. DL techniques do not rely on hand-engineered features, can merge the feature extraction, selection, and classification phases, sometimes require no preprocessing, build subject-independent systems with no calibration, and can deal with highly complex data [52,53]. Details concerning architectures and learning methods of DNNs are given in section 2.2.

An evolving trend in this context is the hybrid BCI (hBCI) in which a BCI system is merged with another BCI system or any other type of brain interfaces. The two interfaces are merged either in a parallel or in a sequential fashion. In the parallel merge, the classification results of both systems are regarded in the final decision. In the sequential merge, intermediate results are made by the first system, then the second system will deliver the final decision based on the intermediate results. In the case of merging two BCIs, the neurophysiological signals can be of the same type or two different types [17,55,56]. Indeed, such a combination of systems with possibly different types of signals will finally construct a robust and more general BCI system.

Table 1

Frequency bands of EEG and MI related tasks.

Band name	Frequency band (Hz)	MI related frequency bands (Hz)
Delta	1–3	–
Theta	4–7	–
Alpha	8–13	8–12
Beta	13–30	16–24
Gamma	above 30	30–35

2.1.2. EEG recordings

Electroencephalogram (EEG) captures physiological activities of the body by recording electrical signals which are the produced post-synaptic potentials by cortical neurons [57]. The electrical signals are recorded via conducting electrodes that are placed on the scalp according to the well-known 10–20 international placement system as stated in Fig. 3, motor imagery related electrodes are identified in blue color [53]. Each electrode records a one-dimensional vector of raw EEG data. The signals are obtained on the three-dimensional scalp surface through volume conduction across multiple brain tissue [57]. Hence, EEG signals are prone to artifacts originating from different parts of the body like the eye, head, neck, or any other muscle. Moreover, the power cable of the recording device and electrode displacement might both cause some artifacts. This form of recording EEG produces weak, non-stationary, and low signal-to-noise ratio signals. Consequently, it complicates the classification and interpretation of signals belonging to a particular case of consideration.

Despite the inherent drawbacks in raw EEG, it has some advantages over other neurological imaging techniques. It characterized by low cost, portability, and causes no side effects because of its non-invasive style [20,52]. Thus, EEG has a variety of applications whether as a screening method or for hypothesis-based diagnostics. Depending on the shape of waves, e.g. rapid spikes waves or slow waves, several forms of brain disorders can be assessed, such as epilepsy [58], tumors [58], Alzheimer [59], sleep disorder [60], etc.

Basically, the amplitude and frequency values in EEG signals are used for discriminating various physiological activities. The amplitude is normally fluctuating in microvolts. The frequency range in EEG signals can be split into several bands as shown in Table 1 [61,62]. The most popular motor imagery frequency bands are also stated in the table. It should be noted that when the Alpha band is recorded from the sensorimotor cortex then it is called mu band. The Gamma band can be recorded consistently using internal electrodes in order to better capture it as it is being weak at the scalp [17]. Slightly different frequency bands or ranges may be found in the literature.

EEG signals are complex in their nature and there exists acute dependency of signal quality on the mental state of the user. Studies proved that the classification accuracy of an intelligent system for a certain task was better at the beginning of the trial than the accuracy at the end of that trial [17]. Therefore, recording or selecting an EEG dataset is crucial for training, validating, and testing machine learning models.

Regardless of the way of obtaining EEG datasets i.e. recording them or getting them ready to use, it should be taken into consideration the key distinctions in the datasets. The MI EEG datasets have the following key distinctions: number and type of MI tasks, number of EEG channels, number of participated volunteers, number of sessions accomplished with each volunteer, number of trials within a session, length of a trial, and the period between two successive sessions.

2.1.3. Preprocessing of raw data

Raw EEG signals usually contain plenty of undesirable background noise that requires elimination before beginning the real analysis. Also, sometimes it is desirable to enhance the raw EEG to better fit the requirements. For those reasons, one or more of the listed below preprocessing methods [47,63–66] might be applied.

Features identification	Extraction	Time-domain	Autoregressive (AR) Adaptive autoregressive Root-mean-square (RMS) Integrated EEG (IEEG)
		Frequency-domain	Fast Fourier transform (FFT) Welch's method Local characteristic-scale decomposition (LCD)
		Time-frequency domain	Short-time Fourier transform (STFT) Wavelet transform (WT) Discrete wavelet transform (DWT)
		Spatial domain	Common spatial pattern (CSP) Common spatio-spectral pattern (CSSP) Common sparse spatio-spectral patterns (CSSSP) Sub-band common spatial pattern (SBCSP)
	Selection	Statistical transformation	Principal component analysis (PCA) Independent component analysis (ICA)
Classification	Filter bank	Filter bank	Filter bank CSP (FBCSP) Discriminant filter bank CSP (DFBCSP) Sparse filter bank CSP (SFBCSP)
		Evolutionary algorithms (EAs)	Particle swarm optimization (PSO) Differential evolution (DE) Artificial bee colony (ABC) Ant colony optimization (ACO) Genetic algorithms (GAs)
		Linear	Support-vector machines (SVMs) Linear discriminant analysis (LDA) Naive Bayes Logistic regression (LR)
		Nonlinear	Bayes quadratic Hidden Markov model (HMM)
	Nearest neighbor		k-nearest neighbor analysis (k-NN)
Artificial neural network (ANN)			Multilayer perceptron (MLP) Radial basis function (RBF) Deep neural network (DNN)

Fig. 4. A summary list of signal processing techniques for features extraction, selection, and classification.

- Notch filtering to remove power line noise at 50 Hz or 60 Hz.
- High-pass filtering with a low cutoff frequency to remove baseline drift.
- Band-pass filtering to select the desirable band(s).
- Clipping the amplitude of EEG signals either by forcing it to be within a specific range or based on the mean and the standard deviation.
- Canceling several samples from the beginning and/or from the end of the signal to remove possible acute artifacts.
- Normalizing the data to zero mean and unit variance using z-score. This can speed up the convergence and avoid trapping by local minima.
- Downsampling the signal for speeding up the computation and reducing memory storage.
- Selection of specific electrodes based on the goal of the application.
- Interpolating of corrupted signals.
- Artifact rejection such as the electrooculogram (EOG) and the electromyogram (EMG) artifacts. This can be achieved in different ways such as a thresholding-based method or data-driven method like the Independent Component Analysis (ICA).
- Referencing using either an electrode (Cz and Fz are often employed as reference electrodes) or the average of signals from all electrodes. Common average referencing (CAR) and Laplacian are widely used

spatial filters. Referencing helps to eliminate some of the background noise.

2.1.4. Dealing with features

As it is stated in Fig. 2, the processed EEG data go through feature extraction, feature selection, and feature classification phases in order to achieve a decision after that. According to [62], the types of usually used EEG features in BCI systems can be grouped into manually selected, statistical, and data-driven adaptive features. Those three phases intensively employed numerous traditional signal processing approaches. Fig. 4 summarizes the most widely used signal processing approaches for dealing with features in EEG applications. Based on the fact that EEG signals are dynamic and highly complex [15], selecting a set of tools for dealing with features would be crucial.

Feature extraction can be achieved based on time, frequency, and/or spatial information contained in the signals. Extracting features depending only on temporal information discards spectral information, likewise, if features are extracted based on only spectral information then the temporal information will be discarded. Hence, either of these techniques is considered weak in extracting salient features. Time-frequency approaches are more vigorous since they relate temporal information with spectral information into each single extracted feature [15]. It must be kept in mind the differences among the time-frequency

approaches. For example, STFT is considered a static method in terms of time and frequency resolution because of the fixed-length window used along with the signal analysis; while wavelet transform delivers dynamic features due to the used multiresolution window [50,67]. Relying on time-frequency approaches is beneficial in the analysis because of the nonstationary nature of EEG signals. Spatial domain approaches can be combined with time and/or frequency domain approaches aiming at increasing the classification accuracy. With the spatial domain approaches the most effective EEG channels can be identified [68] and used with higher weights than those less effective channels.

Normally, high dimensional feature sets are extracted from EEG data. For this, statistical transformation techniques such as PCA and ICA are used for dimensionality reduction and feature selection, however, these techniques are computationally demanding and they may decrease the classification accuracy [32]. Evolutionary algorithms (EAs) approaches involve optimization methods for selecting features from large sets of features, thus EAs could manage the high dimensionality problem [69]. The use of filter bank approaches such as CSP had a high impact on dealing with features in EEG data [17]. Neural network (NN) as a wealth approach uses a framework that combines all of the three phases extraction, selection, and classification into a single pipeline. Despite the lengthy training stage in NN, new unseen data can be analyzed as soon as the network parameters are fixed [62]. This results in more efficient computations which in turn permits using NN in online EEG analysis.

A large number of studies [44,70–73] used more than one approach in each of the three phases extraction, selection, and classification. Noteworthy, many BCI systems do not involve selection phase, however, those systems that involve a selection procedure seek for salient features to achieve higher accuracy.

2.1.5. Feedback

After classifying the signal and making a decision, the result is sent to a translation algorithm in order to translate the classification result to the desired command. The command is executed on a connected end machine based on the recorded brain signal. The taken action by the machine is captured by the user, at this point the loop is closed. The captured feedback can be in different forms such as visually, auditory, and/or haptically.

2.1.6. MI EEG data

MI recording technique captures signals based on a user's imagination of performing a specific task for example limb movement but without actually moving that limb. The imagination of moving a unilateral limb causes a variation of activations in a specific cortex area [16] which are translated into electrical signals propagated by volume conduction through multiple brain tissue. MI applies to different parts of the body such as the movement of the left hand, right hand, tongue [20], left foot, right foot movement [74], wrist movement (flexion, extension, pronation, and supination) [75], elbow flexion/extension, forearm pronation/supination, hand open/close [76], and finger movements [13]. Since those actions show some diversity in terms of spot position and size on the cortical area emitting the signals and the signal strength itself as well, therefore, those actions exhibit noteworthy and discriminative characteristics related to each other and to the contextual EEG [17]. Such facts give aspiration to design a reliable decoder for several different MI actions.

Fundamentally, distinguishing variant imagined movements can be achieved depending on the two well-known phenomena event-related synchronization (ERS) and event-related desynchronization (ERD) [77]. Simply speaking, ERS and ERD occur during the user's imagination of moving a unilateral limb and their meaning, respectively, the recorded signals are enhanced at the contralateral motor-sensory cortex while the signals are attenuated at the ipsilateral. Both the synchronization and desynchronization can be detected in the Alpha (8–12 Hz), often called mu, and Beta (16–24 Hz) bands [73]. Previous studies illustrated the impossibility of decoding movements of left foot, right foot, and a

particular finger because the spots of motor-sensory cortex related to those movements are small to provide distinctive ERS and ERD [17]. However, recent studies proved the possibility of decoding movements of the left foot and right foot [61,74], and movements of a particular finger [13] as well. Hence, it seems that both of the phenomena (ERS and ERD) can be used for extracting distinctive MI features and learning artificial neural networks.

As mentioned earlier in this paper, MI EEG signals are highly variable [45] due to physiological and psychological features for each person at each time. It means that MI recordings may differ from subject-to-subject and session-to-session for a particular subject. This variability in recorded signals can be interpreted as slightly shifting ERS/ERD to higher/lower frequency. Hence, hand-engineered features using traditional signal processing techniques would manufacture subject-dependent features i.e. the resultant BCI system is not general.

2.2. Deep learning

Uncovering all the valuable information within large datasets requires in-depth analysis. Traditional classification systems underachieve in large and dynamic datasets because they cannot cover all states of diversity within the data [18]. DL is a subfield of machine learning seeking to mimic the operation of the human brain by establishing complex interconnected neural construction and hence retrieving a generic model with the ability to handle various data types. DL aims to surpass the shortcomings of traditional neural networks throughout covering whole information within a training dataset. DL is an end-to-end approach [78]; this means that raw data can be directly fed into the DNN for learning the parameters and hyperparameters. Otherwise stated, DL allows to feed DNNs with the raw data and with little or even without any preprocessing; also, DL accomplishes feature extraction, selection, and classification as a single pipeline. Moreover, the DL approach is considered universal, robust, general, and scalable [35]. Nevertheless, training DNN involves a massive number of parameters and hyperparameters which in turn extend the training time compared to other approaches and also exhausts much more hardware resources [71,79,80]. Still, it is possible to tackle those problems using different versed computing devices such as GPUs.

A considerable amount of studies proved the superiority of DL over stat-of-the-art approaches. In [18], researchers transformed the MI EEG signals into 2D time-frequency spectrum images using STFT to train a CNN, the classification results showed that CNN outperformed SVM and ANN classifiers. In [73], they used a CNN for extracting spatial features from MI EEG data then LSTM was used for extracting temporal features of signals, this framework showed better results than the SVM classifier. In [81], they used a CNN for extracting temporal and spatial features from MI EEG data then autoencoders were used for fusing features, this study showed better performance of this framework than a framework consisted of FBCSP as feature extractor with three different classifiers namely naïve Bayes, LDA, and SVM.

The following subsections provide details related to DNNs: the different types of architectures, types of input formulations, learning techniques and regularization methods, and DL platforms.

2.2.1. Architectures

This section reviews the most popular architectures of DNNs.

2.2.1.1. Convolutional neural network (CNN). CNN is a renowned DNN architecture that relies on a specialized sort of linear operation known as convolution. This type of networks can be suitable models for processing different types of signals: images [82], audio [83], videos [84], and EEG as a specific type of signals [13,18,45,76]. Specifically, The hierarchical structure of information within these signals can be learned using CNN throughout learning local features, then higher-level features can be composed of lower-level features [85].

Normally, a CNN consists of: an input layer, pairs of convolution-pooling layers, a fully connected layer, and an output layer. The convolution operation is accomplished by convolving the signals with several 2D filters, known as kernels, in order to extract parallel and complementary features [36]. These kernels play a crucial rule in extracting feature maps from raw data and the amount of their application is controlled by a specific stride. The feature extraction process starts with low-level features, e.g. edges in images, then goes deeper to end with high-level features, e.g. faces in images. The weights of kernels are adjusted throughout the back-propagation training process and using an optimization algorithm to reduce the classification error [86]. Increasing the number of layers means increasing the abstraction ability of the network. Meanwhile, as the network goes deeper the possibility of getting the network overfitted rises [87]. However, there exist methods for tackling the overfitting problem, section 2.2.3 presents some of those methods. The pooling layer works as a downsampling strategy and can be implemented using different types of pooling like sum, max, or average. As can be inferred, the jobs of both the convolution layer and pooling layer are reducing the data complexity and reducing the size of feature maps, respectively. The fully connected layer flattens the extracted features then passes the final classification results to the output layer.

Generic CNNs involve an abundant number of hyperparameters including the number of convolution layers, the size and number of kernels, the size of pooling windows, the size of the stride, and the pooling mode. These hyperparameters have a great influence on network activity, but there is no specific strategy to choose their values [88]. Hence, doing a large number of iterations is the only way to determine the values of the hyperparameters.

2.2.1.2. Recurrent neural network (RNN). RNN as the name infers, it has a circulation behavior throughout its calculations whereby the output depends on both the current input and the previous output. This type of network involves inbuilt memory cells for preserving the previous output states [24,89]. This form of functionality enabled these types of networks to deal with time-series signals analysis such as, among others, natural language [90] and speech recognition [91]. The integrated memory cell has three gates namely input, output, and forget gates for determining the output of the cell [24]. The most popular types of RNN is the long short-term memory (LSTM) network [92], gated recurrent unit (GRU), peephole connection LSTM [62]. The generic structure of these networks has the ability to remember and process past complex values for a long time.

As a comparison of the performance of both CNN and RNN in processing EEG signals; it can be deduced that CNN processes the complete

trial as a single object and extracts spatially-related features. While RNN subdivides a trial into several slices and extracts temporal-related features.

2.2.1.3. Stacked autoencoder. The autoencoder (AE) is a simple unsupervised neural network that is trained to encode and decode data efficiently by utilizing only one hidden layer. In the encoding phase, the input data is subjected to a lower-dimensional space than the data space whereby the most salient latent features are learned. Whereas in the decoding phase, a reconstruction of the real data is accomplished based on the latent features. The stacked autoencoder (SAE), also called deep autoencoder, can be constructed by stacking multiple hidden layers. The variational autoencoder (VAE) and split-brain autoencoder [35] are two variations of the autoencoder type of neural network.

2.2.1.4. Deep belief network. A deep belief network (DBN) is constructed by stacking several layers of restricted Boltzmann machine (RBM). The RBM is a generative model that is trained in unsupervised fashion; it consists of an input or a visible layer, a hidden layer, and bidirectional connectors between the two layers. Each node in the input layer is connected to all nodes in the hidden layer. The input data is represented by latent features in the hidden layer and also vice versa i.e. the latent features can be used for reconstructing the input data in a backward process, hence new data points can be generated [24,93]. During the training process, the first layer in the DBN is considered visible, and the second layer is considered the hidden layer. After this, the second layer in the network is considered the visible layer, and the third layer is the hidden layer, and so on. The process goes further until all layers in the network are trained.

2.2.2. Input formulation

EEG recordings suffer from low signal-to-noise ratio, a correlation among channels, and low spatial resolution, therefore, uncovering the salient features from those signals is not modest. A significant issue that could has some effects on several issues related to processing EEG data like learning mechanism and the amount of preprocessing is the structure of input data or input formulation. Moreover, choosing an input formulation depends on the data under analysis and the neural network architecture. Three types of input formulations related to the analysis of EEG signals exist in researches namely calculated (or hand-engineered) features, images, and time-series signals.

Calculated features input formulation is considered a traditional way of feeding training data as vectors into the input of neural networks. Different techniques, mentioned in section 2.1.4, can be used for crafting features from raw data. The feature extracting techniques that are based on information in time and frequency domains are relevant for dealing with the EEG data. Despite the interpretability and robustness of this approach, it has two main drawbacks: first, it works with small datasets and thus it can be time-consuming in case of large datasets; second, crafting features leads to a loss in latent information and consequently on the classification accuracy.

Images input formulation related to EEG data can be achieved using different approaches such as spectrograms, scalograms, and spatial filtering. The resultant feature maps can be as two-dimensional or three-dimensional matrices. STFT and wavelet can be used for retrieving spectrograms and scalograms, respectively. However, these spectral images discard the spatial information and also the correlation among the time samples whereas these two types of information are considered remarkable in EEG analysis [31]. Hence, a robust classification model can be achieved by considering all spectral, spatial, and temporal correlations of the EEG signals.

Time-series input formulation utilizes the amplitude in time-domain, it has the power of end-to-end training of a neural network i.e. neither preprocessing nor features calculation are required. In this formulation, raw data can be directly fed into the input of a DNN. Recently, alongside

Machine learning	Supervised	Classification
		Regression
	Unsupervised	Clustering
		Association
		Compression
	Semi-supervised	Classification
		Clustering
	Reinforcement	Classification
		Control

Fig. 5. A summary list of the machine learning mechanisms with their general learning tasks.

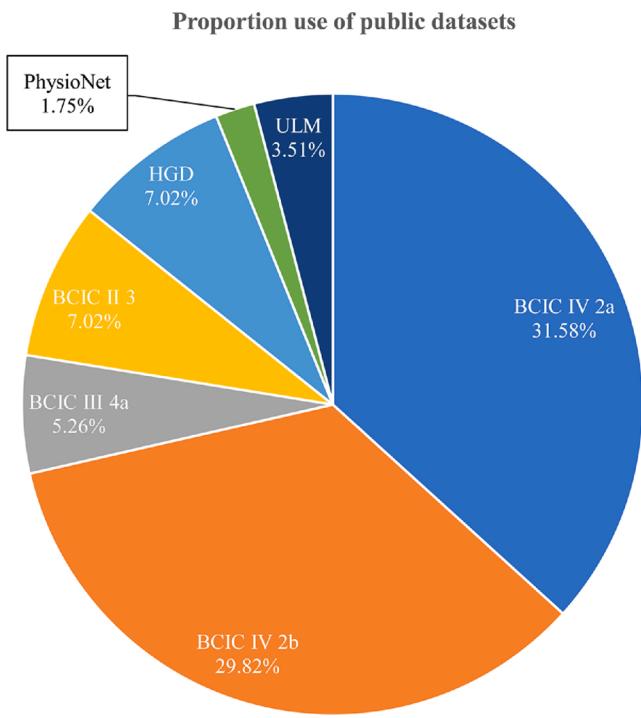


Fig. 6. Proportion use of MI EEG public datasets across all the reviewed studies.

the evolution of DL techniques and also the availability of vigorous computing hardware, researchers began moving to this type of input formulation. However, the overfitting problem arises when small datasets are used for the training process [94]. Nevertheless, thanks to cropping and other training strategies in treating small datasets. Several techniques that are useful in the training process are discussed in the next section.

2.2.3. Learning and regularization methods

DNNs involve a huge number of parameters and hyperparameters [79]; for distinguishing purpose, parameters includes weights and biases whereas hyperparameters include learning rate, momentum, choice of the activation function, number of iterations, number of hidden layers, number of hidden neurons, number of convolution layers, size and number of kernels, size of pooling windows, size of the stride, pooling mode, and others may exist. Learning is the process of tuning those parameters and hyperparameters aiming to accomplish the main task such as classification, clustering, association, compression, and control. The process of training DNNs can be accomplished using supervised learning, unsupervised learning, semi-supervised learning, or reinforcement learning. Fig. 5 summarizes the types of learning mechanisms in machine learning with their renowned learning tasks.

The supervised learning mechanism means there must be a teacher during the training of a neural network in order to accomplish either of two main predictive tasks specifically classification and regression [36]. The investigated dataset contains examples of data with their labels, this dataset is partitioned into three parts: training part is for tuning the network parameters aiming to lessen the loss function; validating part is to validate the trained model, and testing part is for examining the model by presenting unseen data from the environment then determine the classification accuracy using a statistical metric. In a classification task, a sample data is presented to the network to be assigned to one of the multiple predetermined classes, whereas in the regression task a probability is assigned to the sample data of being belonging to a specific class.

The unsupervised learning mechanism does not involve a teacher i.e.

no labels in datasets whereby the model learns the inherent latent features in data. Clustering, association, and compression [36] are the main renowned learning tasks that can be achieved with unsupervised learning.

The semi-supervised learning mechanism falls between the two preceding types [95] whereby a mixture of labeled data and unlabeled data is used for network training. This type of learning has emerged due to the difficulty and expensiveness of gathering fully labeled datasets in some applications whereas gathering unlabeled data is relatively easier and cheaper. Hence, semi-supervised learning aims to maximize the learning performance and at the same time minimize the required cost of gathering data. Spam filtering, webpage classification, and natural language processing are examples of applications that can be achieved using semi-supervised learning. Hence, classification and clustering are two main tasks that can be accomplished using this type of learning [95].

The reinforcement learning mechanism lets the model or the agent learn from the surrounding environment interactively [35]. The agent continuously collects observations from the environment then it takes an action to either maximize the reward feedback or minimize the risk. The process is repeated and the agent always depends on the previous action until achieving the required behavior. Two main tasks can be accomplished using this learning mechanism namely classification [96] and control [97].

The remainder of this section presents several addons or regularization methods that could improve the performance of DNNs, among the recently established addons are data augmentation, dropout, batch normalization, transfer learning, and feature fusion.

Data augmentation can be used for enlarging small datasets and hence tackling the scarcity problem in data. Cropping is a widely used technique for implementing augmentation in order to increase the number of EEG trials in small datasets and thus preventing overfitting problem [85]. The process is to subdivide each trial, before presenting them to the network, into a number of slices using a sliding window with a specific stride and overlapping. Moreover, data augmentation is used for enforcing possible cases that could be faced in different EEG datasets. For example, EEG recording electrodes may be shifted during the experiment, this case can be achieved by spatially rotating the raw EEG data [98] then adding the new artificial data to the dataset. Also, perturbations can be added on the spectrums of EEG data after constructing 2D images [31]. Hence, the data augmentation procedures enrich the labeled data and permit getting more robust decoders which are mainly required in DL especially in case of a small dataset.

Overfitting can be encountered when small datasets are used for training large scale DNNs whereby neurons in fully connected layers are developed together in a dependent manner and thus preventing individual neurons to show their power in learning features. Dropout is a technique that can be used for preventing overfitting problem [36] by randomly removing a number of connectors between neurons.

Batch normalization is a beneficial process in case of changing the distribution of feature maps during the training process which imposes selecting small learning rate [78]. Batch normalization makes the training process smoother and faster by normalizing the output of intermediate layers to zero mean and unit variance [85]; the process is accomplished for a batch of training samples.

Transfer learning is a method that permits training a DNN depending on knowledge taken from another pre-trained network [14]. The process is extremely useful in case only a small dataset is available for the case study, also it reduces the training time and computational cost [18]. A pre-trained network on a large dataset for solving a specific task can be exploited in solving a similar task with a small dataset throughout fine-tuning the network's parameters instead of training it from scratch. Moreover, transferring knowledge from a subject to another and from a session to another [14] can solve the problem of inconsistent EEG signals across different subjects and sessions.

Feature fusion is a technique that combines extracted features from

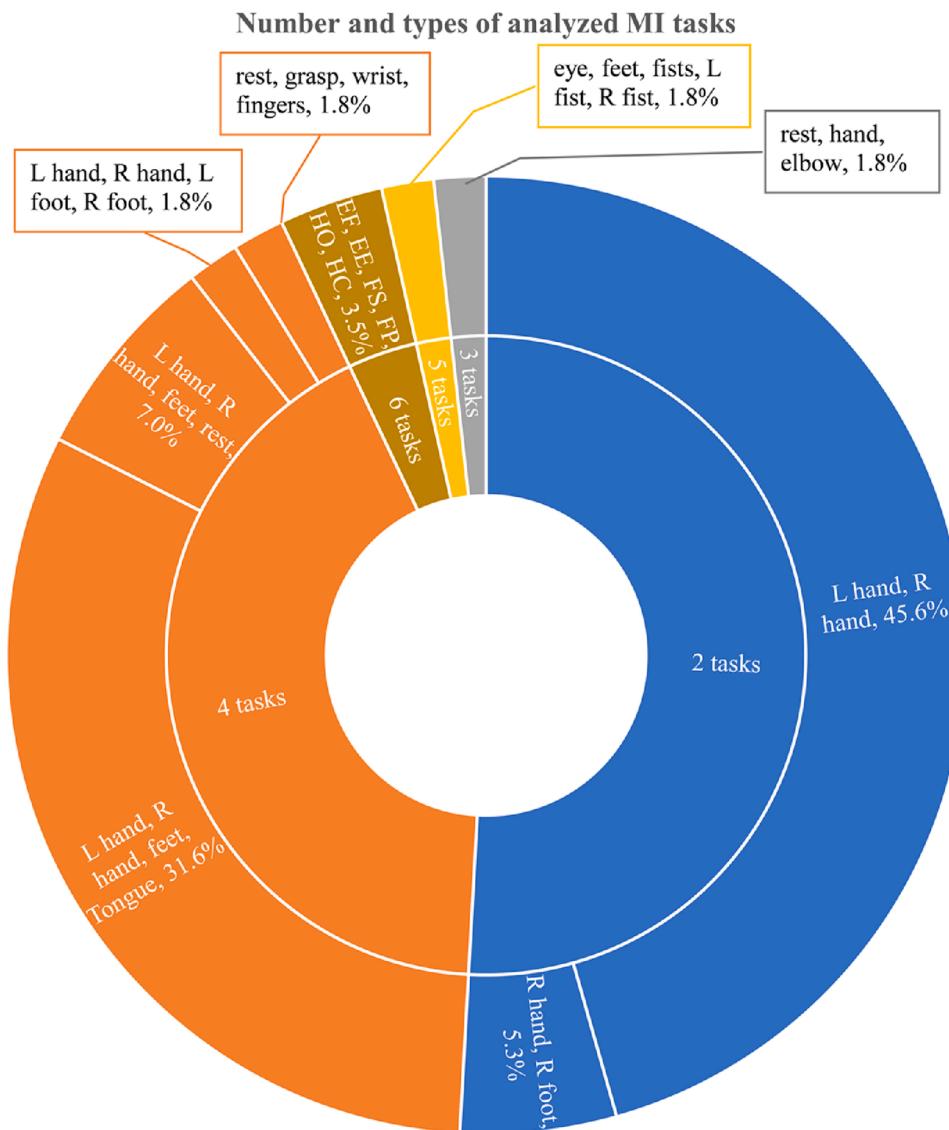


Fig. 7. Proportion of the reviewed experiments according to the number and types of the classified MI tasks. L: left, R: right, EF: elbow flexion, EE: elbow extension, FS: forearm supination, FP: forearm pronation, HO: hand open, HC: hand close.

multiple layers with the aim to extract domain-specific knowledge [99] and hence increasing the classification accuracy. Feature fusion can be accomplished in different manners. For example, a DNN can be trained first on a target dataset, then features from selected layers are combined and passed to a fusion layer like a multilayer perceptron, an autoencoder, or a fully connected layer [99]. Also, feature fusion can be achieved by training multiple DNNs then combining the extracted features from those networks into one fusion layer [100]. Another possible way of fusing features can be done during the training procedure, extracted features at each layer are passed to all subsequent layers [101], hence, in this case, the fusion process will be carried out at all layers of the DNN.

2.2.4. Platforms

One of the main helpful tools behind the popularity and ease of implementation of DL is the existence of open-source DL frameworks. These software packages afford easy-to-use functions that are implementable on GPUs. The list of the most popular ones contains, in alphabetical order: Caffe, DeepLearning4J, Keras, Microsoft Cognitive Toolkit (CNTK), TensorFlow, and Theano. These frameworks can be used under Python, Matlab, Java, and/or C++.

3. Methods

This section presents the searching strategy for articles of interest and the types of the collected data.

3.1. Searching procedure

The process of searching for journal articles was carried out on 31 March 2020 within IEEE, Springer, ScienceDirect, and PubMed databases. We used the following way of arranging keywords during the search: ('motor imagery' OR 'MI') AND ('deep learning' OR 'deep neural network') AND ('electroencephalogram' OR 'electroencephalography' OR 'EEG'). After retrieving the search results, first, the duplicated articles among the databases are excluded. Then based on the criteria of including articles that are listed below, articles that are not within the scope of this review are also excluded. The search process ended up with 40 articles after a full-text screening of the remaining collected articles.

Articles inclusion criteria:

- EEG: only articles with EEG signals were included. Other articles with any other physiological signals such as fMRI and ECoG or a combination of those signals with EEG were excluded.

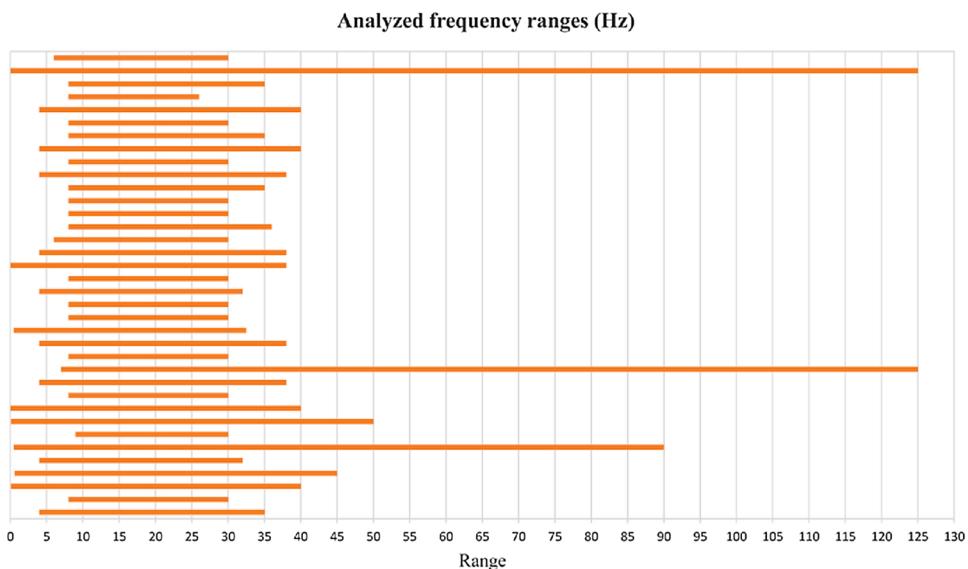
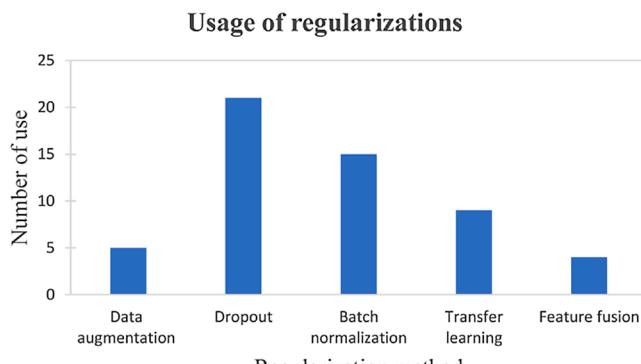


Fig. 8. analyzed frequency range for MI related tasks in each of the reviewed studies.



3.2. Collecting data and demonstration

- Datasets
- Preprocessing and regularization methods
- Input formulation
- DNN architectures
- Number and types of MI tasks
- Highest performance metric
- Computing hardware
- Frameworks of deep learning

This review focuses on the important information that anyone should have before diving into the field of classification of MI EEG-based signals. Such information was gathered from the experiments and results in the revised studies. The collected data can be categorized into:

The preprocessed and regularization methods that have been used by the reviewed studies are presented in Table A1. Preprocessing methods such as band-pass filtering, notch filtering, were applied to the public datasets before publishing them; therefore, such methods did not mention in the table. The challenging issue to the DNNs is keeping artifacts in the data whereby these networks are capable of dealing with raw data. The number of studies that have used an artifact removal method is only 5 out of the 40; and they have used ICA, thresholding, or Automatic artifact removal (AAR) toolbox.

Normally, experiments on EEG signals consider a specific range of frequencies to limit the bandwidth of the signals to be analyzed [45]. The reviewed studies used either band-pass filter or low and high-pass filters in order to retain the frequency range of interest and dismiss the rest. Fig. 8 depicts the analyzed frequency ranges by the reviewed studies.

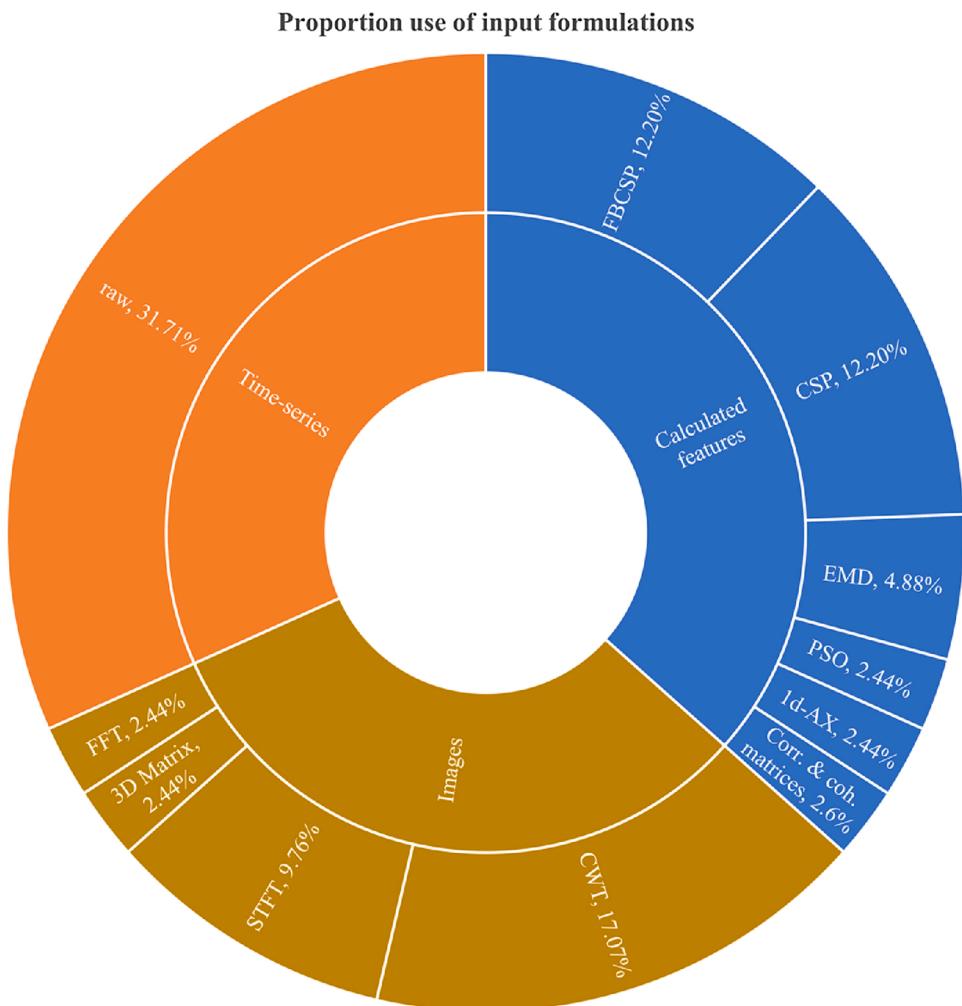


Fig. 10. Proportion use of input formulation types across all the reviewed studies with the used techniques for achieving them.

Regularizations are addon methods that are applied during the training of DNNs [35,52], discussed in section 2.2.3. The registered methods from the reviewed articles are five methods namely: data augmentation (cropping is included), dropout, batch normalization, transfer learning, and feature fusion. Fig. 9 shows the amount of using the different regularization methods across the reviewed articles.

4.3. Input formulation

All the reviewed studies used one of three input formulation types which are calculated features, time-series, and images [37]. The number of studies that have used each of the three types is as follows: calculated features (14), time-series (13), and images (13). About 70% of the studies used CSP in the process of features calculation. Wavelet transform and STFT were used by 54% and 31% of the studies that adopted images in the classification process, respectively. Fig. 10 depicts the proportion of using those three formulations with the used techniques for achieving them.

4.4. DNN architectures

The most predominant observed architecture types from the reviewed papers are CNN, RNN (LSTM and GRU), and hybrid-CNN (h-CNN). The dominant architecture was CNN at about 73% of the total number of the reviewed studies while 13% of h-CNN and 14% of other types. The h-CNN architectures contain in addition to convolutional layers other architecture types like RNN or autoencoders [73,89].

Generally, the design characteristics of the different architectures fell within the architecture type, the number of hidden layers, the type of activation functions, and the type of end classifiers. In fact, in this study, we concentrated on the architecture types and the activation functions that have been employed for classifying MI EEG-based signals.

The reviewed studies employed different activation functions in their proposed DNNs. The employed activation functions in the studies that used CNN and h-CNN were as follows: 66% ReLU, 23% ELU, 8% linear, and 3% SELU. The proposed architectures in the remaining studies employed the sigmoid and tanh activation functions. Fig. 11 depicts the proportion use of the different DNN architectures by all the reviewed studies with the employed activation functions.

During the training process, the parameters of the DNNs are updated using different optimization algorithms. The majority of the reviewed studies used one of the renowned optimization algorithms namely gradient descent (GD), stochastic gradient descent (SGD), and Adam. Single studies used the following optimization algorithms: Bayesian, Adadelta, and conjugate gradient (CG). Fig. 12 depicts the proportion use of the different optimization algorithms across all the reviewed studies.

All the reviewed studies which have mentioned the type of computing hardware have used either a GPU or a powerful computer for training their proposed DNN. Also, the proposed DNNs were developed using different DL frameworks. Fig. 13 depicts the proportion use of those frameworks.

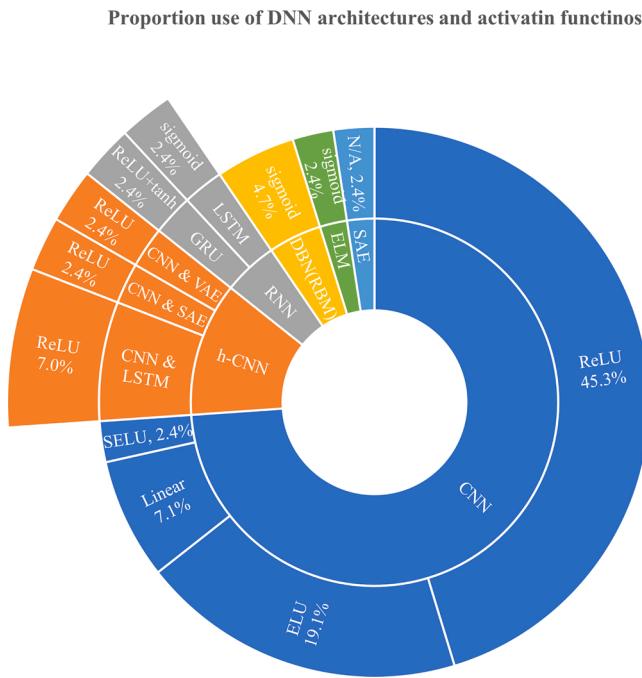


Fig. 11. Proportion use of the DNN architectures across all the reviewed studies with the employed activation functions.

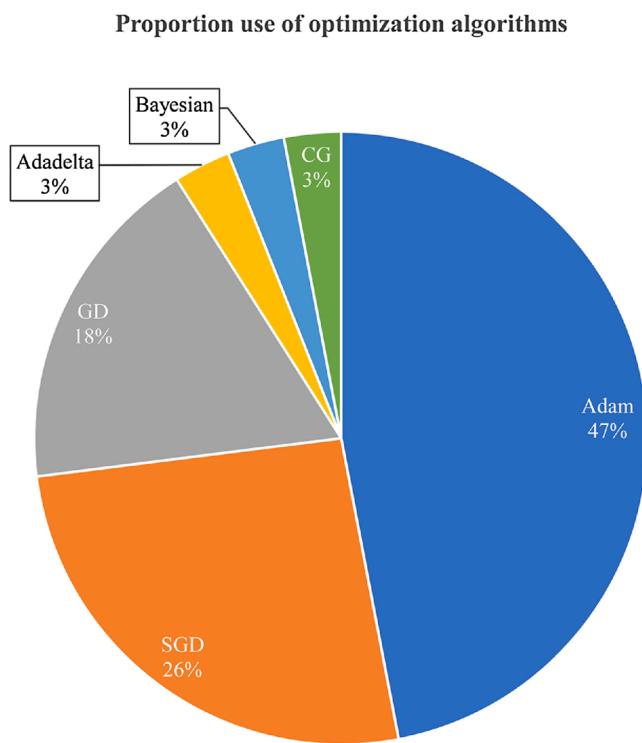


Fig. 12. Proportion use of the different optimization algorithms across all the reviewed studies.

5. Discussion

This section presents discussions about the attained statistics of the collected data from the reviewed articles. Also, it gives recommendations for choosing some of the design parameters.

Different datasets of MI EEG recordings are publicly available and can be used for testing the proposed architecture and comparing the

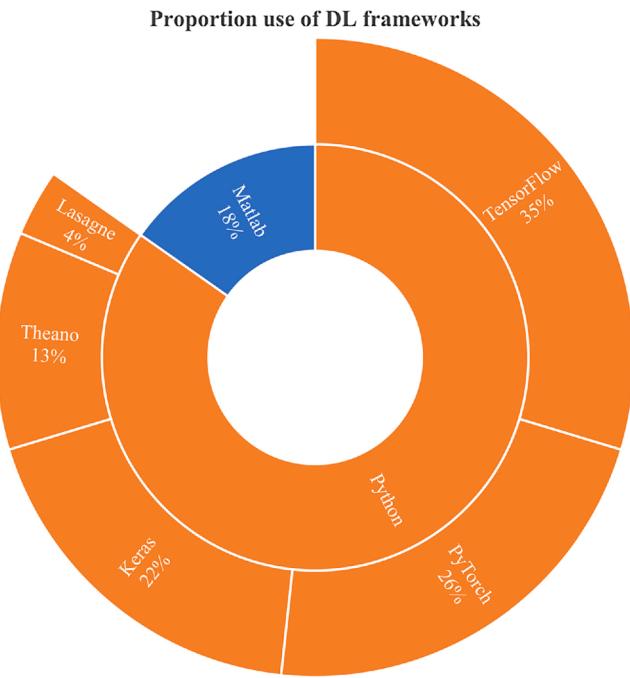


Fig. 13. Proportion use of the DL frameworks across all the reviewed studies.

results with other studies. The most renowned MI EEG datasets are dataset 2a and dataset 2b from BCI competition IV; these two datasets were used by most of the reviewed studies. Other MI EEG datasets are also available but were less frequently used such as dataset 4a from BCI competition III, dataset 3 from BCI competition II, high-gamma dataset, upper limb movement, and PhysioNet.

Different physiological tasks reflect different electrical signals on the scalp. The frequency range of the signals that are related to those tasks is one of these differences. Studies have suggested somehow different frequency ranges for dealing with the MI tasks during the analysis process. However, from the reviewed studies and as shown in Fig. 8, the observed starting point of the MI related frequency range was within 6–8 Hz; while the observed ending point of the range was within 38–40 Hz. Several studies that dealt with the high-gamma dataset have analyzed high frequencies because this dataset contains high frequencies.

CSP and its variations can be recommended in case of adopting calculated features as inputs to the neural network. While wavelet transform-based methods are recommended in case of using images as inputs to the neural network. Wavelet transforms were used more frequently than STFT in capturing images from the EEG data. Whereby the wavelet methods analyze the signals in multiple frequencies and time resolutions which are highly recommended for the analysis of EEG signals due to the non-stationarity nature of those signals; whereas STFT analyzes the signals using fixed-length window. To achieve a simple comparison among the three types of the discussed input formulations, the observed studies that used CNN in the analysis of dataset 2a from BCI competition IV have used input as follows: four studies used images, three studies used calculated features, and four studies used time-series. The average classification accuracy of the studies that used time-series was higher than the average accuracies of studies that used both images and calculated features. Based on that and also the fact that DL exploits the whole training data whereas loss of information could be encountered using images and calculated features, time-series input formulation can be recommended.

The most frequently used DNN architecture for the classification of MI tasks across the reviewed studies was CNN. Several studies proved the superiority of CNN over other types of architectures for MI tasks. Also, other architectures such as RNN and autoencoders were used in

combination with CNN. Moreover, single studies proved the ability of DBN and RNN for MI tasks with sufficient classification accuracies. Hence, based on the review process, CNN is highly recommended for the classification of MI tasks. ReLU and ELU are recommended to be used with CNN, this recommendation is based on the large number of studies that used these two activation functions with CNN architecture.

Aiming to improve the performance of the DNN classifiers by means of accuracy or computation complexity, several regularization methods can be suggested to be used during the training process. Among those are the dropout and batch normalization methods for preventing the possible overfitting problem during training the network. Also, transfer learning was used by several studies for transferring knowledge from a trained network to untrained one; this has a high impact on shortening the required training time. Data augmentation was also used by those studies that encountered small datasets; it could increase the training trials by a large factor.

6. Conclusion

During the last decades, researchers began to exploit and employ MI EEG recordings into various real-world application domains. However, it is soon realized the challenges behind dealing with those signals, whereas, in addition to the characteristics of general EEG signals which are weak, non-stationary, and low signal-to-noise ratio signals; MI EEG signals are shown to be highly variable due to the physiological and psychological features for each subject at each time. It means that MI recordings may differ from subject-to-subject and session-to-session for a particular subject. Many studies have been implemented to handle those signals. The studies used various methods for feature extraction and classification before DNN came into existence.

DNNs have been extensively applied to the classification of different MI tasks. Classification of the MI signals using DNNs suffers from the small size of the available datasets, whereby, these classifiers require a large number of examples to be learned. However, different

regularization methods, such as cropping, augmentation, and transfer learning have been proposed and implemented to overcome the shortage in datasets.

Based on the concluded results from the 40 reviewed studies, the used DNNs in the classification of MI EEG signals differ in several design parameters such as architecture type, input formulation, and activation function. Deep CNN architecture with the ReLU activation function is found to be the most effective architecture for the classification of MI tasks. DL has the capability of exploiting the whole input data for training networks, hence, raw EEG data could be reasonably used in the training process. Moreover, it is observed that the classification of MI tasks that are related to a single limb, e.g. tasks of different fingers, are more challenging than the classification of tasks that related to different limbs. This is because of the high similarity of the recorded signals in the former case and also the near cortical regions that are emitting those signals. In another context, it must be realized that a large number of hidden layers does not always lead to better performance; in contrast, it may lead to overfitting problem as well as more computational complexity.

CRediT authorship contribution statement

Ali Al-Saegh: Methodology, Resources, Writing - original draft, Visualization. **Shefa A. Dawwd:** Conceptualization, Supervision, Writing - review & editing. **Jassim M. Abdul-Jabbar:** Conceptualization, Supervision.

Declaration of Competing Interest

The authors report no declarations of interest.

Appendix A

Table A1
Summary of the collected data from the 40 reviewed articles.

Study	DL architecture	Activation function	Optimization strategy	Dataset	Accuracy (%)	Input formulation	Preprocessing and regularization methods	Frequency range (Hz)	Deep learning hardware	DL platform
[45]	CNN	ReLU	GD	• BCIC IV 2b • BCIC IV 2b • private	83.2 82.61 85.83	images (CWT) calculated features (EMD)	• Batch normalization • Electrode selection • Dropout • AAR • Referencing • Downsampling • Data augmentation • Dropout • Batch normalization	4–35 8–30	– –	– –
[63]	CNN	ReLU	SGD	• private	87.03	calculated features (correlation matrix and coherence matrix)	0.1–40	NVIDIA GeForce GTX 1080 Ti GPU	Keras	
[75]	CNN	ReLU	Adam	• BCIC III 4a	96.34	calculated features (EMD, CSP)	–	–	–	–
[105]	CNN	ReLU	–	• ULM • BCIC IV 2a • BCIC IV 2b	90.3 91.57 87.6	images (CWT) time-series	–	0.6–45 4–32	Two NVIDIA GeForce RTX 2080 Ti GPUs	Matlab
[76]	CNN	ReLU	Adam	• BCIC III 4a	94.66	calculated features (BSS, CWT)	• Dropout	0.5–90	PC with Intel Core i7–700HQ 2.80 GHz with 16 GB of RAM	Tensorflow and Keras
[103]	CNN	ELU	SGD	• BCIC IV 2a • Private	0.61* 76.96	calculated features (FBCSP)	• Data normalization	9–30	NVIDIA GeForce GTX 970 GPU	–
[61]	CNN	ReLU	SGD							

(continued on next page)

Table A1 (continued)

Study	DL architecture	Activation function	Optimization strategy	Dataset	Accuracy (%)	Input formulation	Preprocessing and regularization methods	Frequency range (Hz)	Deep learning hardware	DL platform
[15]	CNN	ReLU	GD	• BCIC IV 2b • BCIC II 3	83 91.4 75.7	images (CWT)	• Dropout • Batch normalization • Feature fusion • Transfer learning • Data normalization • Dropout • Batch normalization • Transfer learning • Data normalization • Data augmentation	0.1–50	–	–
[100]	CNN	ELU	Adam	• BCIC IV 2a • HGD	95.4	time-series	• Batch normalization • Transfer learning • Data normalization • Thresholding • Clipping • Data normalization • Cropping • Dropout • Batch normalization	0–40	NVIDIA GeForce GTX 1080 GPU	PyTorch
[14]	CNN	ReLU	Adam	• BCIC IV 2b • Private	64 73	calculated features (CSP)	• Data normalization • Transfer learning • Data normalization • Data augmentation	8–30	NVIDIA GeForce GTX 1070 GPU	Keras
[94]	CNN	linear	Adam	• BCIC IV 2a • BCIC IV 2b • HGD	75.9 84.7 94.9	time-series	• Data normalization • Dropout • Batch normalization • Transfer learning • Data normalization • Data augmentation	4–38	–	PyTorch.
[31]	CNN	ELU	Adam	• BCIC IV 2a • HGD	74.6 93.7	time-series	• Data augmentation • Dropout • Batch normalization • Thresholding • Clipping • Data normalization • Cropping • Dropout • Batch normalization	7–125	NVIDIA GeForce GTX 1080Ti GPU	PyTorch
[47]	CNN	ReLU	Adam	• BCIC IV 2b	84.24	time-series	• Data normalization • Cropping • Dropout • Batch normalization	8–30	NVIDIA GeForce GTX Titan X GPU	Theano & Keras
[52]	CNN & LSTM	Not stated	Adam	• PhysioNet	98.3	time-series	• Data normalization • Dropout • Cropping • Dropout • Batch normalization	–	NVIDIA GeForce GTX Titan X Pascal GPU	TensorFlow
[99]	CNN	ELU	SGD	• BCIC IV 2a	74.5	time-series	• Data normalization • Feature fusion • Transfer learning • Electrodes selection • Batch normalization	–	NVIDIA GeForce GTX 1080 GPU	PyTorch
[33]	CNN	ELU	Adam	• BCIC IV 2a	0.644*	images (3D matrices)	• AAR • Downsampling • Downsampling • Electrodes selection • Batch normalization	4–38	–	–
[13]	CNN	ReLU	SGD	• private	73.7	images (CWD)	• Downsampling • Feature fusion • Transfer learning • Interpolation • Dropout • Feature fusion	0.5–32.5	Nvidia Quadro K620 GPU PC with a dual-core i7–7700 4.2-GHz processor and 16-GB RAM	TensorFlow
[53]	CNN	ReLU	SGD	• private	75	calculated features (CSP)	• Feature fusion • Transfer learning • Dropout • Feature fusion	8–30	i7–7700 4.2-GHz processor and 16-GB RAM	TensorFlow
[101]	CNN	ReLU	Adam	• BCIC IV 2a	79.90	calculated features (CSP)	• Transfer learning • Feature fusion • Feature fusion	8–30	NVIDIA GeForce GTX 1080 GPU	TensorFlow
[18]	CNN	ReLU	–	• BCIC IV 2b	74.2	images (STFT)	• Transfer learning • Feature fusion • Feature fusion	4–32	NVIDIA GeForce GTX 1080 Ti GPU	PyTorch
[107]	CNN	ReLU	SGD	• BCIC II 3 • private	90.1 85.2	images (complex Morlet wavelets)	• Normalization • Data augmentation	8–30	PC with Intel 4.00 GHz Core i7 with 16 GB of RAM	Keras
[104]	CNN	Linear and ELU	Adam	• BCIC IV 2a • BCIC IV 2b	0.56* 83 31	Time-series	• Cropping • Transfer learning	0.01–38	–	–

(continued on next page)

Table A1 (continued)

Study	DL architecture	Activation function	Optimization strategy	Dataset	Accuracy (%)	Input formulation	Preprocessing and regularization methods	Frequency range (Hz)	Deep learning hardware	DL platform
[108]	CNN	ReLU	GD	• ULM • BCIC III 4a	99.35	images (CWT)	• Downsampling • Transfer learning	–	–	–
[73]	CNN & LSTM	ReLU	Adam	• BCIC IV 2a	0.80*	calculated features (FBCSP)	• Dropout	4–38	NVIDIA GeForce GTX 1070 GPU	TensorFlow
[71]	CNN & VAE	ReLU	GD	• BCIC IV 2b • Private	0.564* 0.603*	images (STFT)	• Interpolation	6–30	–	–
[21]	CNN	ReLU	Adam	• BCIC IV 2a • BCIC IV 2b	80.44 82.39	calculated features (FBCSP, PSO)	• ICA • Data augmentation • Interpolation • Referencing	8–36	PC with Intel 4.7 GHz core i7 with 16 GB of RAM	Python
[87]	CNN	ReLU	GD	• BCIC II 3	90	images (CWT)	–	8–30	–	–
[95]	ELM	Sigmoid	–	• BCIC IV 2b	67.76	calculated features (CSP)	–	8–30	PC with a 3.4 GHz processor and 8 GB RAM	Matlab
[109]	SAE	–	–	• BCIC IV 2a	71	time-series	• Data augmentation	8–35	NVIDIA GeForce GTX 960 M GPU	Theano
[77]	CNN	SELU	SGD	• BCIC IV 2b	78.44	images (STFT)	• Dropout • Batch normalization	4–38	NVIDIA Pascal Titan X GPU and NVIDIA GeForce GTX 1080 Ti GPU	Theano
[110]	DBN (RBM)	Sigmoid	Adadelta	• BCIC IV 2b	83.55	time-series	• Data normalization • Dropout	8–30	PC with Intel Core i7–7700, 3.6 GHz, and RAM is 16 GB	Matlab
[43]	CNN	Linear and ELU	Adam	• BCIC IV 2a	67	time-series	• Electrodes selection • Dropout • Batch normalization	4–40	NVIDIA Quadro M6000 GPU	Tensorflow
[62]	LSTM	Sigmoid	Adam	• BCIC IV 2a	71	calculated features (1d-AX)	• Normalization • Channel weighting • Dropout • Batch normalization	8–35	–	Tensorflow
[111]	GRU	ReLU & Tanh	SGD	• BCIC IV 2a • BCIC IV 2b	73.56 82	calculated features (FBCSP)	• Cropping • Dropout • Batch normalization	8–30	–	–
[16]	CNN	ReLU	Adam	• BCIC IV 2a	0.659*	calculated features (FBCSP)	• Dropout • Batch normalization	4–40	–	–
[89]	CNN & LSTM	ReLU	–	• private	86.71	time-series	• Referencing • ICA • Downsampling • Electrode selection • Dropout	8–26	–	TensorFlow and Theano
[32]	DBN (RBM)	Sigmoid	Conjugate gradient	• BCIC IV 2b	84	images (FFT)	• Data normalization • Transfer learning	8–35	PC with dual-core processor i7–3700 of 3.4 GHz and 12 GB	Matlab
[85]	CNN	ELU	SGD	• BCIC IV 2a • HGD	70.9 91.95	time-series	• Downsampling • Electrode selection • Cropping • Dropout • Batch normalization	0–125	Geforce GTX Titan Black GPU	Lasagne
[34]	CNN & SAE	ReLU	GD	• BCIC II 3 • BCIC IV 2b	77.6 90	images (STFT)	–	6–30	PC with Intel 4.00 GHz Core i7 and 16 GB of RAM	Matlab

The * sign indicated that the classification accuracy is given by kappa value.

Table A2

Details about the used MI EEG datasets by the 40 reviewed articles.

Dataset	Nu. of electrodes	Nu. of subjects	Nu. of sessions	Nu. of trials per session	Imagination duration (seconds)	Actions	Sampling rate (Hz)	reference
BCI competition IV dataset 2a (BCIC IV 2a)	22	9	2	288	4	<ul style="list-style-type: none"> • Left hand • Right hand • Feet • Tongue 	250	[112]
BCI competition IV dataset 2b (BCIC IV 2b)	3	9	5	First two sessions 120 Last three sessions 160	4	<ul style="list-style-type: none"> • Left hand • Right hand 	250	[112]
BCI competition III dataset 4a (BCIC III 4a)	118	5	1	280	2	<ul style="list-style-type: none"> • Right hand • Right foot 	1000	[113]
BCI competition II dataset 3 (BCIC II 3)	3	1	1	280	6	<ul style="list-style-type: none"> • Left hand • Right hand 	128	[87], [107]
High-Gamma Dataset (HGD)	128	14	13	1000 in total	4	<ul style="list-style-type: none"> • Left hand • Right hand • Both feet • Rest 		[31]
PhysioNet EEG Dataset	64	109	1	270	10	<ul style="list-style-type: none"> • Eye closed • Both feet • Both fists • Left fist • Right fist 	160	[52]
Upper limb movement (ULM)	61	15	2	180	3	<ul style="list-style-type: none"> • Elbow flexion • Elbow extension • Forearm supination • Forearm pronation • Hand open • Hand close 	512	[104]
Dataset used by [63]	16	5	1	180	4	<ul style="list-style-type: none"> • Left hand • Right hand 	128	[63]
Dataset used by [13]	16	22	–	Each of 18 subjects done 440 Each of 4 subjects done 560	<10	<ul style="list-style-type: none"> • Rest • Grasp-related tasks • Wrist-related tasks • Fingers-related tasks 	2048	[13]
Dataset used by [53]	62	54	2	100	4	<ul style="list-style-type: none"> • Left hand • Right hand 	1000	[53]
Dataset used by [61]	10	8	5	120	3	<ul style="list-style-type: none"> • Left hand • Right hand • Left foot • Right foot 	–	[61]
Dataset used by [71]	5	5	3	400 per subject	6	<ul style="list-style-type: none"> • Left hand • Right hand 	250	[71]
Dataset used by [75]	64	25	7	5 sessions of 40 2 sessions of 50	4	<ul style="list-style-type: none"> • Rest • Hand • Elbow 	1000	[75]
Dataset used by [14]	15	25	4	50	5	<ul style="list-style-type: none"> • Left hand • Right hand 	1000	[14]
Dataset used by [116]	14	5	2	60	2	<ul style="list-style-type: none"> • Left hand • Right hand 	256	[116]

References

- [1] G. Li, et al., Towards emerging EEG applications: a novel printable flexible Ag/AgCl dry electrode array for robust recording EEG signals at forehead sites, *J. Neural Eng.* 17 (2) (2020), <https://doi.org/10.1088/1741-2552/ab71ea>, p. 026001.
 - [2] Z.A.A. Alyasseri, A.T. Khader, M.A. Al-Betar, J.P. Papa, O.A. Alomari, S. N. Makhadmeh, Classification of EEG mental tasks using multi-objective flower pollination algorithm for person identification, *Int. J. Integr. Eng.* 10 (7) (2018) 102–116.
 - [3] Z.A.A. Alyasseri, A.T. Khader, M.A. Al-Betar, A. Abasi, S. Makhadmeh, N.S. Ali, The effects of EEG feature extraction using multi-wavelet decomposition for mental tasks classification, *ACM Int. Conf. Proceeding Ser.* (2019) 139–146, <https://doi.org/10.1145/3321289.3321327>.
 - [4] Z.A.A. Alyasseri, A.T. Khader, M.A. Al-Betar, A.K. Abasi, S.N. Makhadmeh, EEG signals denoising using optimal wavelet transform hybridized with efficient metaheuristic methods, *IEEE Access* 8 (2020) 10584–10605, <https://doi.org/10.1109/ACCESS.2019.2962658>.
 - [5] Z.A.A. Alyasseri, A.T. Khader, M.A. Al-Betar, J.P. Papa, O.A. Alomari, EEG feature extraction for person identification using wavelet decomposition and multi-
 - [6] M. Aldayel, M. Ykhlef, A. Al-Nafjan, Deep learning for EEG-based preference classification in neuromarketing, *Appl. Sci.* 10 (4) (2020) 1–23, <https://doi.org/10.3390/app10041525>.
 - [7] J. Heo, G. Yoon, EEG studies on physical discomforts induced by virtual reality gaming, *J. Electr. Eng. Technol.* 15 (2020) 1323–1329, <https://doi.org/10.1007/s42835-020-00373-1>.
 - [8] K. Abualsaoud, M. Mahmudin, M. Saleh, A. Mohamed, Ensemble classifier for epileptic seizure detection for imperfect EEG data, *Sci. World J.* 2015 (2015), <https://doi.org/10.1155/2015/945689>.
 - [9] Z.A.A. Alyasseri, A.T. Khader, M.A. Al-Betar, O.A. Alomari, Person identification using EEG channel selection with hybrid flower pollination algorithm, *Pattern Recognit.* 105 (2020) 107393, <https://doi.org/10.1016/j.patcog.2020.107393>.
 - [10] Z.A.A. Alyasseri, A.T. Khader, M.A. Al-Betar, J.P. Papa, O.A. Alomari, EEG-based person authentication using multi-objective flower pollination algorithm, 2018 IEEE Congr. Evol. Comput. CEC 2018 - Proc. (2018), <https://doi.org/10.1109/CEC.2018.8477895>.
 - [11] Z.A. Alkareem Alyasseri, A.T. Khader, M.A. Al-Betar, J.P. Papa, O.A. Alomari, S. N. Makhadmeh, An efficient optimization technique of EEG decomposition for user

- authentication system, 2nd Int. Conf. BioSignal Anal. Process. Syst. ICBAPS 2018 (2018) 1–6, <https://doi.org/10.1109/ICBAPS.2018.8527404>.
- [12] A. Czeszumski, et al., Hyperscanning: a valid method to study neural inter-brain underpinnings of social interaction, *Front. Hum. Neurosci.* 14 (2020) 39, <https://doi.org/10.3389/fnhum.2020.00039>.
- [13] R. Alazrai, M. Abuhijleh, H. Alwanni, M.I. Daoud, A deep learning framework for decoding motor imagery tasks of the same hand using EEG signals, *IEEE Access* 7 (2019) 109612–109627, <https://doi.org/10.1109/access.2019.2934018>.
- [14] X. Zhu, P. Li, C. Li, D. Yao, R. Zhang, P. Xu, Separated channel convolutional neural network to realize the training free motor imagery BCI systems, *Biomed. Signal Process. Control* 49 (2019) 396–403, <https://doi.org/10.1016/j.bspc.2018.12.027>.
- [15] H.K. Lee, Y. Choi, Application of continuous wavelet transform and convolutional neural network in decoding motor imagery brain-computer interface, *Entropy* 21 (12) (2019) 1–12, <https://doi.org/10.3390/e21121199>.
- [16] S. Sakhavi, C. Guan, S. Yan, Learning temporal information for brain-computer interface using convolutional neural networks, *IEEE Trans. neural networks Learn. Syst.* 29 (11) (2018) 5619–5629.
- [17] N. Padfield, J. Zabalza, H. Zhao, V. Masero, J. Ren, EEG-based brain-computer interfaces using motor-imagery: techniques and challenges, *Sensors (Switzerland)* 19 (6) (2019) 1–34, <https://doi.org/10.3390/s19061423>.
- [18] G. Xu, et al., A deep transfer convolutional neural network framework for EEG signal classification, *IEEE Access* 7 (2019) 112767–112776, <https://doi.org/10.1109/access.2019.2930958>.
- [19] G. Litjens, et al., A survey on deep learning in medical image analysis, *Med. Image Anal.* 42 (December 2017) (2017) 60–88, <https://doi.org/10.1016/j.media.2017.07.005>.
- [20] W. Qiao, X. Bi, Deep spatial-temporal neural network for classification of EEG-based motor imagery, *ACM International Conference Proceeding Series* (2019) 265–272, <https://doi.org/10.1145/3349341.3349414>.
- [21] I. Majidov, T. Whangbo, Efficient classification of motor imagery electroencephalography signals using deep learning methods, *Sensors (Switzerland)* 19 (7) (2019) 1–13, <https://doi.org/10.3390/s19071736>.
- [22] S.H. Khan, M. Hayat, F. Porikli, Regularization of Deep Neural Networks with Spectral Dropout, *Neural Netw.* 110 (February 2019) (2019) 82–90, <https://doi.org/10.1016/j.neunet.2018.09.009>.
- [23] E. Nurvitadhi, et al., Can FPGAs beat GPUs in accelerating next-generation deep neural networks? *FPGA 2017 - Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays* (2017) 5–14, <https://doi.org/10.1145/3020078.3021740>.
- [24] O. Faust, Y. Hagiwara, T.J. Hong, O.S. Lih, U.R. Acharya, Deep learning for healthcare applications based on physiological signals: a review, *Comput. Methods Programs Biomed.* 161 (April 2018) (2018) 1–13, <https://doi.org/10.1016/j.cmpb.2018.04.005>.
- [25] B. Rim, N. Sung, S. Min, M. Hong, Deep learning in physiological signal data : a survey, *Sensors (Switzerland)* 20 (4) (2020) 969, <https://doi.org/10.3390/s20040969>.
- [26] R. Ron-angevin, F. Velasco-álvarez, Á. Fernández-rodríguez, A. Díaz-estrella, M. J. Blanca-mena, F.J. Vizcaíno-martín, Brain-computer Interface application: auditory serial interface to control a two- class motor-imagery-based wheelchair, *J. Neuroeng. Rehabil.* 14 (2017) 1–16, <https://doi.org/10.1186/s12984-017-0261-y>.
- [27] H. Yu, H. Lu, S. Wang, K. Xia, Y. Jiang, P. Qian, A general common spatial patterns for EEG analysis with applications to vigilance detection, *IEEE Access* 7 (2019) 111102–111114, <https://doi.org/10.1109/access.2019.2934519>.
- [28] D.O. Souto, T.K.F. Cruz, K. Coutinho, A. Julio-Costa, P.L.B. Fontes, V.G. Haase, Effect of motor imagery combined with physical practice on upper limb rehabilitation in children with hemiplegic cerebral palsy, *NeuroRehabilitation* 46 (2020) 53–63, <https://doi.org/10.3233/NRE-192931>.
- [29] F. Wang, Motor imagery classification using geodesic filtering common spatial pattern and filter- bank feature weighted support vector machine Motor imagery classification using geodesic filtering common spatial pattern and filter-bank feature weighted support vec, *Rev. Sci. Instrum.* 91 (2020) 034106, <https://doi.org/10.1063/1.5142343>.
- [30] P.S. Ramya, K. Yashasvi, A. Anjum, A. Bhattacharyya, R.B. Pachori, Development of an effective computing framework for classification of motor imagery EEG signals for brain – computer interface. *Advances in Computational Intelligence Techniques*, Springer, Singapore, 2020, pp. 17–35.
- [31] Y. Li, X.R. Zhang, B. Zhang, M.Y. Lei, W.G. Cui, Y.Z. Guo, A channel-projection mixed-scale convolutional neural network for motor imagery EEG decoding, *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (6) (2019) 1170–1180, <https://doi.org/10.1109/TNSRE.2019.2915621>.
- [32] N. Lu, T. Li, X. Ren, H. Miao, A deep learning scheme for motor imagery classification based on restricted boltzmann machines, *IEEE Trans. Neural Syst. Rehabil. Eng.* 25 (6) (2017) 566–576, <https://doi.org/10.1109/TNSRE.2016.2601240>.
- [33] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, L. Sun, A multi-branch 3D convolutional neural network for EEG-Based motor imagery classification, *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (10) (2019) 2164–2177, <https://doi.org/10.1109/TNSRE.2019.2938295>.
- [34] Y.R. Tabar, U. Halici, A novel deep learning approach for classification of EEG motor imagery signals, *J. Neural Eng.* 14 (1) (2016) 16003, <https://doi.org/10.1088/1741-2560/14/1/016003>.
- [35] M.Z. Alom, et al., A state-of-the-art survey on deep learning theory and architectures, *Electron.* 8 (3) (2019) 1–67, <https://doi.org/10.3390/electronics8030292>.
- [36] W.G. Hatcher, W. Yu, A survey of deep learning: platforms, applications and emerging research trends, *IEEE Access* 6 (c) (2018) 24411–24432, <https://doi.org/10.1109/ACCESS.2018.2830661>.
- [37] A. Craik, Y. He, J.L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review, *J. Neural Eng.* 16 (3) (2019), <https://doi.org/10.1088/1741-2552/abQab5>.
- [38] R. Schirrmeister, L. Gemein, K. Eggensperger, F. Hutter, T. Ball, Deep learning with convolutional neural networks for decoding and visualization of EEG pathology, in: 2017 IEEE Signal Processing in Medicine and Biology Symposium, SPMB 2017 - Proceedings, 2018-Janua, 2017, pp. 1–7, <https://doi.org/10.1109/SPMB.2017.8257015>.
- [39] L. Employment, *Brain-computer interfaces : toward a daily life employment*, *Brain Sci.* 10 (157) (2020) 10–13.
- [40] B. Lei, et al., Walking imagery evaluation in brain computer interfaces via a multi-view multi-level deep polynomial network, *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (3) (2019) 497–506, <https://doi.org/10.1109/TNSRE.2019.2895064>.
- [41] X. Tang, T. Wang, Y. Du, Y. Dai, Motor imagery EEG recognition with KNN-based smooth auto-encoder, *Artif. Intell. Med.* 101 (2019) 101747, <https://doi.org/10.1016/j.artmed.2019.101747>.
- [42] H. Yi, Efficient machine learning algorithm for electroencephalogram modeling in brain – computer interfaces, *Neural Comput. Appl.* 3 (2020), <https://doi.org/10.1007/s00521-020-04861-3>.
- [43] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung, B.J. Lance, EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces, *J. Neural Eng.* 15 (5) (2018) 1–30, <https://doi.org/10.1088/1741-2552/aace8c>.
- [44] M.Z. Baig, N. Aslam, H.P.H. Shum, L. Zhang, Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG, *Nrl. Northumbria. Ac. UK* 90 (2017) 184–195, <https://doi.org/10.1108/17410391111097438>.
- [45] F. Li, F. He, F. Wang, D. Zhang, Y. Xia, X. Li, A novel simplified convolutional neural network classification algorithm of motor imagery EEG signals based on deep learning, *Appl. Sci.* 10 (2020) 1605, <https://doi.org/10.3390/app10051605>.
- [46] G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang, A. Etemad, Classification of hand movements from EEG using a deep attention-based LSTM network, *IEEE Sens. J.* 20 (6) (2020) 3113–3122, <https://doi.org/10.1109/JSEN.2019.2956998>.
- [47] Z. Tayeb, et al., Validating deep neural networks for online decoding of motor imagery movements from eeg signals, *Sensors (Switzerland)* 19 (1) (2019) 210, <https://doi.org/10.3390/s19010210>.
- [48] Y. Zhou, S. He, Q. Huang, Y. Li, A hybrid asynchronous brain-computer interface combining SSVEP and EOG signals, *IEEE Trans. Biomed. Eng.* (2020), <https://doi.org/10.1109/tbme.2020.2972747>, pp. 1–1.
- [49] K.S. Hong, U. Ghafour, M.J. Khan, Brain-machine interfaces using functional near-infrared spectroscopy: a review, *Artif. Life Robot.* 25 (2020) 204–218, <https://doi.org/10.1007/s10015-020-00592-9>.
- [50] C. Herff, D.J. Krusienski, P. Kubben, The potential of Stereotactic-EEG for brain-computer interfaces: current progress and future directions, *Front. Neurosci.* 14 (February) (2020) 1–8, <https://doi.org/10.3389/fnins.2020.00123>.
- [51] C. Herff, A. De Pesters, D. Heger, P. Brunner, G. Schalk, T. Schultz, An ECoG-Based BCI based on auditory attention to natural speech, *Brain-Computer Interface Researc* (April) 2017 7–19, <https://doi.org/10.1007/978-3-319-57132-4>.
- [52] D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, Y. Liu, Making sense of spatio-temporal preserving representations for EEG-Based human intention recognition, *IEEE Trans. Cybern.* (2019) 1–12, <https://doi.org/10.1109/TCYB.2019.2905157>.
- [53] O. Kwon, M. Lee, C. Guan, S. Lee, Subject-independent brain-computer interfaces based on deep convolutional neural networks, *IEEE Trans. Neural Networks Learn. Syst.* XX (X) (2019) 1–15, <https://doi.org/10.1109/TNNLS.2019.2946869>.
- [54] A. Singh, S. Lal, H.W. Gueggen, Reduce calibration time in motor imagery using spatially regularized symmetric positive-definite matrices based classification, *Sensors (Switzerland)* 19 (2) (2019) 379, <https://doi.org/10.3390/s19020379>.
- [55] A. Khalaf, E. Sejdic, M. Akcakaya, *Hybrid EEG – fTCd brain – computer interfaces. Neuroergonomics: Principles and Practice*, 2020, pp. 295–314.
- [56] C. Chen, et al., Quadcopter robot control based on hybrid brain-computer interface system, *Sensors Mater.* 32 (3) (2020) 991, <https://doi.org/10.18494/sam.2020.2517>.
- [57] M. Sazgar, M.G. Young, *Overview of EEG, electrode placement, and montages. Absolute Epilepsy and EEG Rotation Review*, 2019, pp. 117–125.
- [58] W. Mardini, M.M. Bani Yassein, R. Al-Rawashdeh, S. Aljawarneh, Y. Khamayseh, O. Meqdadi, Enhanced detection of epileptic seizure using EEG signals in combination with machine learning classifiers, *IEEE Access* 8 (2020) 24046–24055, <https://doi.org/10.1109/ACCESS.2020.2970012>.
- [59] R. Pandya, S. Nadiawalra, R. Shah, M. Shah, Buildout of methodology for meticulous diagnosis of K-Complex in EEG for aiding the detection of Alzheimer's by artificial intelligence, *Augment. Hum. Res.* 5 (1) (2020), <https://doi.org/10.1007/s41133-019-0021-6>.
- [60] M. Entropy, A novel microwave treatment for sleep disorders and classification of sleep stages using multi-scale entropy, *Entropy* 22 (3) (2020) 347, <https://doi.org/10.3390/e22030347>.
- [61] B.E. Olivas-Padilla, M.I. Chacon-Murguia, Classification of multiple motor imagery using deep convolutional neural networks and spatial filters, *Appl. Soft Comput.* J. 75 (2019) 461–472, <https://doi.org/10.1016/j.asoc.2018.11.031>.

- [62] P. Wang, A. Jiang, X. Liu, J. Shang, L. Zhang, LSTM-based EEG classification in motor imagery tasks, *IEEE Trans. Neural Syst. Rehabil. Eng.* 26 (11) (2018) 2086–2095, <https://doi.org/10.1109/TNSRE.2018.2876129>.
- [63] X. Tang, W. Li, X. Li, W. Ma, X. Dang, Motor imagery EEG recognition based on conditional optimization empirical mode decomposition and multi-scale convolutional neural network, *Expert Syst. Appl.* 149 (2020) 113285, <https://doi.org/10.1016/j.eswa.2020.113285>.
- [64] S.A. Ludwig, J. Kong, Investigation of different classifiers and channel configurations of a mobile P300-based brain-computer interface, *Med. Biol. Eng. Comput.* 55 (12) (2017) 2143–2154, <https://doi.org/10.1007/s11517-017-1658-2>.
- [65] T.K. Reddy, V. Arora, S. Kumar, L. Behera, Y.-K. Wang, C.-T. Lin, Electroencephalogram based reaction time prediction with differential phase synchrony representations using Co-operative multi-task deep neural networks, *IEEE Trans. Emerg. Top. Comput. Intell.* 3 (2019) 369–379, <https://doi.org/10.1109/tetci.2018.2881229>.
- [66] L. Jingwei, C. Yin, Z. Weidong, Deep learning EEG response representation for brain computer interface, 2015 34th Chinese Control Conference (CCC) (2015) 3518–3523, <https://doi.org/10.1109/ChiCC.2015.7260182>.
- [67] M. Elsayed, A. Badawy, M. Mahmudin, T. Elfouly, A. Mohamed, K. Abualsaud, FPGA implementation of DWT EEG data compression for wireless body sensor networks, in: 2016 IEEE Conf. Wirel. Sensors, ICWiSE 2016, vol. 2017-December, 2017, pp. 1–5, <https://doi.org/10.1109/ICWiSE.2016.8187756>.
- [68] B.H. Yilmaz, C.M. Yilmaz, C. Kose, Diversity in a signal-to-image transformation approach for EEG-based motor imagery task classification, *Med. Biol. Eng. Comput.* 58 (2) (2020) 443–459, <https://doi.org/10.1007/s11517-019-02075-x>.
- [69] P. Tan, X. Wang, Y. Wang, Dimensionality reduction in evolutionary algorithms-based feature selection for motor imagery brain-computer interface, *Swarm Evol. Comput.* 52 (2020) 100597, <https://doi.org/10.1016/j.swevo.2019.100597>.
- [70] J. Kevric, A. Subasi, Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system, *Biomed. Signal Process. Control* 31 (2017) 398–406, <https://doi.org/10.1016/j.bspc.2016.09.007>.
- [71] M. Dai, D. Zheng, R. Na, S. Wang, S. Zhang, EEG classification of motor imagery using a novel deep learning framework, *Sensors (Switzerland)* 19 (3) (2019) 1–16, <https://doi.org/10.3390/s19030551>.
- [72] S. Kumar, A. Sharma, T. Tsunoda, An improved discriminative filter bank selection approach for motor imagery EEG signal classification using mutual information, *BMC Bioinformatics* 18 (16) (2017) 545, <https://doi.org/10.1186/s12859-017-1964-6>.
- [73] R. Zhang, Q. Zong, L. Dou, X. Zhao, A novel hybrid deep learning scheme for four-class motor imagery classification, *J. Neural Eng.* 16 (6) (2019) 066004, <https://doi.org/10.1088/1741-2552/ab3471>.
- [74] M. Tariq, P.M. Trivailo, M. Simic, Mu-Beta event-related (de)synchronization and EEG classification of left-right foot dorsiflexion kinaesthetic motor imagery for BCI, *PLoS One* 15 (3) (2020) 1–20, <https://doi.org/10.1371/journal.pone.0230184>.
- [75] X. Ma, S. Qiu, W. Wei, S. Wang, H. He, S. Member, Deep channel-correlation network for motor imagery decoding from same limb, *IEEE Trans. Neural Syst. Rehabil. Eng.* 31 (1) (2020) 1, <https://doi.org/10.1109/TNSRE.2019.2953121>.
- [76] N. Mamnone, C. Ieracitano, F.C. Morabito, A deep CNN approach to decode motor preparation of upper limbs from time-frequency maps of EEG signals at source level, *Neural Netw.* 124 (2020) 357–372, <https://doi.org/10.1016/j.neunet.2020.01.027>.
- [77] K.W. Ha, J.W. Jeong, Motor imagery EEG classification using capsule networks, *Sensors (Switzerland)* 19 (13) (2019) 2854, <https://doi.org/10.3390/s19132854>.
- [78] I. Ullah, M. Hussain, Eul H. Qazi, H. Aboalsamh, An automated system for epilepsy detection using EEG brain signals based on deep learning approach, *Expert Syst. Appl.* 107 (April 2018) (2018) 61–71, <https://doi.org/10.1016/j.eswa.2018.04.021>.
- [79] J.J. Escobar, J. Ortega, M. Damas, R.S. Kiziltepe, J.Q. Gan, Energy-time analysis of convolutional neural networks distributed on heterogeneous clusters for EEG classification, International Work-Conference on Artificial Neural Networks (2019) 895–907, https://doi.org/10.1007/978-3-030-20518-8_74.
- [80] J. Thomas, T. Masiczak, N. Sinha, T. Kluge, J. Dauwels, Deep learning-based classification for brain-computer interfaces, in: 2017 IEEE Int. Conf. Syst. Man, Cybern. SMC 2017, vol. 2017-January, 2017, pp. 234–239, <https://doi.org/10.1109/SMC.2017.8122608>.
- [81] S.U. Amin, M. Alsulaiman, G. Muhammad, M.S. Hossain, M. Guizani, Deep learning for EEG motor imagery-based cognitive healthcare. *Connected Health in Smart Cities*, 2020, pp. 233–254.
- [82] A. Milosavljevic, Identification of salt deposits on seismic images using deep learning method for semantic segmentation, *ISPRS Int. J. Geo-Information* 9 (1) (2020) 24, <https://doi.org/10.3390/ijgi9010024>.
- [83] Mustaqueem, S. Kwon, A CNN-assisted enhanced audio signal processing for speech emotion recognition, *Sensors (Switzerland)* 20 (1) (2020) 183, <https://doi.org/10.3390/s20010183>.
- [84] S.J. Narayanan, B. Perumal, S. Saman, A.P. Singh, Deep Learning for Person Re-Identification in Surveillance Videos, vol. 865, 2020.
- [85] R.T. Schirrmeister, et al., Deep learning with convolutional neural networks for EEG decoding and visualization, *Hum. Brain Mapp.* 38 (11) (2017) 5391–5420, <https://doi.org/10.1002/hbm.23730>.
- [86] A.F. Pérez-Zapata, A.F. Cardona-Escobar, J.A. Jaramillo-Garzón, G.M. Díaz, Deep convolutional neural networks and power spectral density features for motor imagery classification of EEG signals, in: International Conference on Augmented Cognition, vol. 1, 2018, pp. 158–169, <https://doi.org/10.1007/978-3-319-91470-1>.
- [87] B. Xu, et al., Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification, *IEEE Access* 7 (MI) (2019) 6084–6093, <https://doi.org/10.1109/ACCESS.2018.2889093>.
- [88] T. Shi, L. Ren, W. Cui, Feature recognition of motor imaging EEG signals based on deep learning, *Pers. Ubiquitous Comput.* 23 (3–4) (2019) 499–510, <https://doi.org/10.1007/s00779-019-01250-z>.
- [89] J. Yang, S. Yao, J. Wang, Deep fusion feature learning network for MI-EEG classification, *IEEE Access* 6 (c) (2018) 79050–79059, <https://doi.org/10.1109/ACCESS.2018.2877452>.
- [90] X. Wu, Y. Zhao, D. Radev, A. Malhotra, Identification of patients with carotid stenosis using natural language processing, *Eur. Radiol.* (2020), <https://doi.org/10.1007/s00330-020-06721-z>.
- [91] G. Li, R. Hu, X. Wang, R. Zhang, A near-end listening enhancement system by RNN-based noise cancellation and speech modification, *Multimed. Tools Appl.* 78 (11) (2019) 15483–15505, <https://doi.org/10.1007/s11042-018-6947-8>.
- [92] Y. Liu, et al., Deep C-LSTM neural network for epileptic seizure and tumor detection using high-dimension EEG signals, *IEEE Access* 8 (2020) 37495–37504, <https://doi.org/10.1109/ACCESS.2020.2976156>.
- [93] J. Li, Z.L. Yu, Z. Gu, W. Wu, Y. Li, L. Jin, A hybrid network for ERP detection and analysis based on restricted boltzmann machine, *IEEE Trans. Neural Syst. Rehabil. Eng.* 26 (3) (2018) 563–572, <https://doi.org/10.1109/TNSRE.2018.2803066>.
- [94] H. Wu, et al., A parallel multiscale filter bank convolutional neural networks for motor imagery EEG classification, *Front. Neurosci.* 13 (November) (2019) 1–9, <https://doi.org/10.3389/fnins.2019.01275>.
- [95] Q. She, B. Hu, Z. Luo, T. Nguyen, Y. Zhang, A hierarchical semi-supervised extreme learning machine method for EEG recognition, *Med. Biol. Eng. Comput.* 57 (1) (2019) 147–157, <https://doi.org/10.1007/s11517-018-1875-3>.
- [96] E. Lin, Q. Chen, X. Qi, Deep reinforcement learning for imbalanced classification, *Appl. Intell.* (2020), <https://doi.org/10.1007/s10489-020-01637-z>.
- [97] W. Bai, T. Li, S. Tong, NN reinforcement learning adaptive control for a class of nonstrict-feedback discrete-time systems, *IEEE Trans. Cybern.* 1 (2020) 1–12, <https://doi.org/10.1109/TCYB.2020.2963849>.
- [98] M.M. Krell, S.K. Kim, Rotational data augmentation for electroencephalographic data, Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS (2017) 471–474, <https://doi.org/10.1109/EMBC.2017.8036864>.
- [99] S.U. Amin, M. Alsulaiman, G. Muhammad, M.A. Bencherif, M.S. Hossain, Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification, *IEEE Access* 7 (2019) 18940–18950, <https://doi.org/10.1109/ACCESS.2019.2895688>.
- [100] S.U. Amin, M. Alsulaiman, G. Muhammad, M.A. Mekhtiche, M. Shamim Hossain, Deep Learning for EEG motor imagery classification based on multi-layer CNNs feature fusion, *Future Gener. Comput. Syst.* 101 (2019) 542–554, <https://doi.org/10.1016/j.future.2019.06.027>.
- [101] D. Li, J. Wang, J. Xu, X. Fang, Densely feature fusion based on convolutional neural networks for motor imagery EEG classification, *IEEE Access* 7 (2019) 132720–132730, <https://doi.org/10.1109/access.2019.2941867>.
- [102] R. Yuvaraj, J. Thomas, T. Kluge, J. Dauwels, A deep learning Scheme for automatic seizure detection from Long-term scalp EEG, 2018 52nd Asilomar Conference on Signals, Systems, and Computers (2018) 368–372.
- [103] G. Dai, J. Zhou, J. Huang, N. Wang, HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification, *J. Neural Eng.* 17 (1) (2020) 016025.
- [104] D. Zhao, F. Tang, B. Si, X. Feng, Learning joint space-time-frequency features for EEG decoding on small labeled data, *Neural Netw.* 114 (2019) 67–77, <https://doi.org/10.1016/j.neunet.2019.02.009>.
- [105] S. Taheri, M. Ezoji, S. Mahmoud, Convolutional neural network based features for motor imagery EEG signals classification in brain – computer interface system, *SN Appl. Sci.* 2 (February) (2020), <https://doi.org/10.1007/s42452-020-2378-z>.
- [106] C.J. Ortiz-Echeverri, S. Salazar-Colores, J. Rodríguez-Reséndiz, R.A. Gómez-Lozano, A new approach for motor imagery classification based on sorted blind source separation, continuous wavelet transform, and convolutional neural network, *Sensors (Switzerland)* 19 (20) (2019) 1–14, <https://doi.org/10.3390/s19204541>.
- [107] Z. Zhang, et al., A novel deep learning approach with data augmentation to classify motor imagery signals, *IEEE Access* 7 (c) (2019) 15945–15954, <https://doi.org/10.1109/ACCESS.2019.2895133>.
- [108] S. Chaudhary, S. Taran, V. Bajaj, A. Sengur, Convolutional neural network based approach towards motor imagery tasks EEG signals classification, *IEEE Sens. J.* 19 (12) (2019) 4494–4500, <https://doi.org/10.1109/JSEN.2019.2899645>.
- [109] A. Hassanpour, M. Moradikia, H. Adeli, S.R. Khayami, P. Shamsinejadbabaki, A novel end-to-end deep learning scheme for classifying multi-class motor imagery electroencephalography signals, *Expert Syst. (August)* (2019) 1–21, <https://doi.org/10.1111/exsy.12494>.
- [110] X.-L. Tang, W.-C. Ma, D.-S. Kong, W. Li, Semisupervised deep stacking network with adaptive learning rate strategy for motor imagery eeg recognition, *Neural Comput.* 31 (5) (2019) 919–942, <https://doi.org/10.1162/NECO>.
- [111] T. Jian Luo, C. le Zhou, F. Chao, Exploring spatial-frequency-sequential relationships for motor imagery classification with recurrent neural network,

- BMC Bioinformatics 19 (1) (2018) 1–18, <https://doi.org/10.1186/s12859-018-2365-1>.
- [112] M. Tangermann, et al., Review of the BCI competition IV, Front. Neurosci. 6 (JULY) (2012) 1–31, <https://doi.org/10.3389/fnins.2012.00055>.
- [113] B. Blankertz, et al., The BCI competition III: validating alternative approaches to actual BCI problems, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2) (2006) 153–159, <https://doi.org/10.1109/TNSRE.2006.875642>.