IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 40, No. 5, October, 2023, pp. 1857-1868
Journal homepage: http://iieta.org/journals/ts

Advancements in Image Feature-Based Classification of Motor Imagery EEG Data: A Comprehensive Review

Cagatay Murat Yilmaz[1*], Bahar Hatipoglu Yilmaz[2]

[1] Department of Software Engineering, Karadeniz Technical University, Trabzon 61080, Turkey
[2] Department of Computer Engineering, Karadeniz Technical University, Trabzon 61080, Turkey

Corresponding Author Email: cmyilmaz@ktu.edu.tr

https://doi.org/10.18280/ts.400507

Received: 8 March 2023
Revised: 14 July 2023
Accepted: 28 July 2023
Available online: 30 October 2023

Keywords:
motor imagery, brain-computer interface, signal-to-image conversion, short-time Fourier transform, deep learning, convolutional neural networks, reviews

ABSTRACT

Non-invasive acquisition and analysis of human brain signals play a crucial role in the development of brain-computer interfaces, enabling their widespread applicability in daily life. Motor imagery has emerged as a prominent technique for the advancement of such interfaces. While initial machine and deep learning studies have shown promising results in the context of motor imagery, several challenges remain to be addressed prior to their extensive adoption. Deep learning, renowned for its automated feature extraction and classification capabilities, has been successfully employed in various domains. Notably, recent research efforts have focused on processing and classifying motor imagery EEG signals using two-dimensional data formats, yielding noteworthy advancements. Although existing literature encompasses reviews primarily centered on machine learning or deep learning techniques, this paper uniquely emphasizes the review of methods for constructing two-dimensional image features, marking the first comprehensive exploration of this subject. In this study, we present an overview of datasets, survey a range of signal-to-image conversion methods, and discuss classification approaches. Furthermore, we comprehensively examine the current challenges and outline future directions for this research domain.

1. INTRODUCTION

Electroencephalography (EEG) is a neuroimaging method for measuring, recording, and examining the brain activity. It has been used in brain-computer interfaces (BCIs), consumer neuroscience, psychiatry, diagnosis, rehabilitation, and treatment. The most common use case is BCIs, which establish a communication and control environment without muscles and peripheral devices. Some examples include limb motor imagery, controlling unmanned aerial vehicles, virtual drones, robotic arms, and virtual reality environments. Motor imagery (MI) is a universal way to develop such BCIs, where actions are performed by mentally simulating the motor action in the brain without using the motor system. Researchers have proposed various machine/deep learning techniques, but the flourishing applications of MI-BCIs still need to be improved. Some difficulties encountered include noise and artefacts, non-stationarity, nonlinearity, inconsistency, being distorted, session-to-session and subject-to-subject variability, trade-offs between training time and performance, and the time-consuming calibration. The problems become even tougher when considering non-invasive real-time MI-EEG systems.

Nowadays, deep learning (DL) has become very popular due to its success in automatic feature extraction and classification capabilities, which differs from classical machine learning. It has been used to classify different data types and has demonstrated great success. Most deep learning research on MI-EEG has used one-dimensional time series data. However, further studies are needed to investigate the processing of MI-EEG signals in deep learning aspects. For example, some deep learning studies used 2D inputs and processed them in 2D instead of 1D time series MI-EEG signals or features. These approaches have increasing applications and successful results have been achieved. In this respect, examining studies on this subject is highly important. So far, most reviews have focussed only on machine learning or deep learning. In this paper, we mainly focus on a review of the techniques for converting MI-EEG signals to images. This is important because the processing of MI-EEG signals, especially in deep learning, is a significant issue. In the literature, only some researchers have investigated this topic. This present study will contribute to this gap in the literature. The basic research questions are: (1) What are the signal-to-image conversion methods for MI-EEG signals? (2) What are the classification techniques using these image features? and (3) What are the challenges and future directions in the classification of MI-EEG signals?

The MI-EEG datasets used for model development and testing are detailed in Sec. 2. The signal-to-image conversion and classification methods are analysed in Sec. 3. The last sections of this article present challenges, future directions, and conclusions.

2. DATASETS

Supervised datasets are essential for the learning models. Compared to data sets in other fields, the number of MI-EEG data sets is restricted. Because it is challenging to collect large-scale and high-quality data due to the strict conditions for subjects and experimental environments. Also, simulating the motor action in the brain is difficult. Most machine learning studies attempting to classify MI with image features have generally been conducted on similar datasets. These data sets were usually recorded from healthy subjects sitting in comfortable armchairs with armrests. They are shown in Table 1, where studies, number of subjects, recording channels and MI tasks categories are given.

BCI Competition datasets have been used most to validate signal processing and classification methods. They presented new challenges, such as small training sets, classification of continuous MI-EEG data without trial structure and classification of MI-EEG signals affected by eye movement artefacts, etc. DS1 (BCI Competition II dataset III [1]) is one of the first instances. The data set consists of recordings from a normal subject. Its objective is to provide a continuous output that could be used as BCI feedback. The challenge is to control a feedback bar in one dimension using MI of left or right-hand movements. It was recorded from C3, Cz, and C4 locations. It includes seven runs (conducted on the same day) with 40 trials each, resulting in 280 trials. The EEG was sampled with 128Hz and filtered between 0.5 – 30Hz [1]. DS2 (BCI Competition III dataset IIIa [2]) is a four-class (left hand, right hand, foot, or tongue) cued dataset. The data set consists of recordings from three subjects. It includes 60-channel EEG data sampled with 250 Hz and filtered between 1 – 50 Hz with a Notch filter. The experiment consists of several runs (at least six) with 40 trials each [2]. DS3 (BCI Competition III dataset IVa [2]) is a two-class cued dataset recorded from five subjects using 118 electrodes. This dataset was developed to reduce calibration times. While the number of training samples is high for subjects “aa” and “al”, there are very few samples available for subjects “av”, “aw”, and “ay”. It addresses the challenge of getting along with only a small amount of training data. The signals were band pass-filtered between 0.05 – 200 Hz and digitized at 1000 Hz. There are 280 trials for each subject [2].

DS4 (BCI Competition IV dataset 1 [3]) is an asynchronous dataset. This dataset presents the challenge of applying classifiers to continuous MI-EEG signals without cue. This dataset includes two different sub-datasets. The first one involved four healthy participants. In the whole session, the MI was performed without feedback. For each participant, two classes of MI were selected from the three classes left hand, right hand, and foot. The signals were recorded from 59 electrode locations. They were band-pass filtered between 0.05 and 200 Hz and then digitized at 1000 Hz. The session was divided into two parts: training and testing recordings. In the first two runs, arrows pointing left, right, or down were presented as visual cues and training data was recorded. Each run consists of 50 trials of each of the chosen two classes that have been presented, resulting in 200 trials. In the next four runs, the MI tasks were cued by acoustic stimuli and test data was recorded. Each run consists of 30 trials for each class that has been recorded. Thus, a total of 240 trials were performed. The second sub-data set was artificially generated, designed to test machine learning methods to see if MI-EEG signals with specific characteristics can be artificially generated. The artificial MI-EEG signals were constructed using the linear combination of background noise, baseline drifts, etc. DS5 (BCI Competition IV dataset 2a [3]) is composed of four classes (left-hand, right-hand, both feet and tongue) of data collected from nine subjects. It was recorded in two sessions on separate days using a cue based BCI paradigm, and no feedback was provided during the execution of the MI. The training and test sets were created with data from these distinct sessions. It includes the challenge of session-to-session transfer. The classifier model constructed using the training session should generalize on unseen data recorded on different day. Each session contains about 288 trials (artefacted trials were marked by an expert). EEG was recorded monopolarly from 22 electrode locations, sampled with 250 Hz and bandpass filtered in the 0.5–100 Hz. A 50 Hz notch filter was also used to reduce line noise. Approximately five minutes of EEG recording (2 min with eyes open, 1 min with eyes closed, and 1 min with eye movements) was also made at the beginning of each session to assess the effects of EOG. The other challenge is how eye movement artefacts could affect classification performance. Due to this, data from 3 EOG channels were recorded in addition to the EEG channels [3].

DS6 (BCI Competition IV dataset 2b [3]) includes two-class (left hand, right hand) data from nine subjects. The whole dataset for each subject contains five sessions, whereby the first two sessions contain MI-EEG data without feedback, and the last three sessions were recorded with feedback. All the training and test sets were recorded on five separate days. Due to this, the session-to-session differences must be taken into consideration. As for DS5, a recording of about 5 minutes was performed at the start of each session to gauge the effect of EOG. The bipolar recording was made from C3, Cz, and C4 positions with a sampling frequency of 250 Hz. EEG signals were bandpass filtered between 0.5 Hz to 100 Hz, and a notch filter at 50 Hz was enabled. EOG data was recorded during all sessions with three monopolar electrodes to facilitate artefact processing. This data set focuses on classifying MI-EEG data affected by eye movement artefacts. The first two sessions (six runs per session) were performed without feedback on two days. Each session consisted of six runs with ten trials each and two classes of imagery, which resulted in 120 trials per session (60 for the left hand and 60 for the right hand). The three other sessions with smiley feedback consist of four runs. Each type of MI task has twenty trials in each run. The first three sessions are for training, and the last two for testing [3].

The other datasets are as follows. Most of these are freely available benchmark datasets. DS7, DS09, and DS10 are another open dataset. DS8 and DS11 are used only in specific studies. DS7 (PhysioNet - EEG Motor Movement/Imagery Dataset [4, 5]) is another benchmark dataset with the largest number of subjects. It contains MI-EEG signals for opening and closing the left fist, right fist, both fists, and both feet MI tasks. This dataset also includes baselines and the signals generated when tasks are performed using the motor system. EEG was recorded from 64 electrodes with a rate of 160 samples per second [4, 5]). DS8 [6] comprises two-class (left hand, right hand) data collected from five subjects. EEG was recorded from C3, C1, Cz, C2, and C4 electrodes, sampled with 250 Hz. There are a total of 400 trials for each subject [6]. DS9 [7] is a two-class (left hand, right hand) dataset recorded from 52 subjects using 64 EEG positions with a sampling frequency of 512 Hz. EMG signals were simultaneously recorded to check hand movements [7]. DS10 [8] is a large dataset collected during the development of a BCI. Data was collected from 13 subjects using 19 EEG inputs. The dataset involves 60 hours of data, 75 recording sessions, 201 individual EEG BCI interaction session segments, and motor imageries of the left hand, right hand, left leg, right leg, and tongue [8]. DS11 [9] is a four-class (left hand, right hand, feet, tongue) dataset. MI-EEG and monocular vision were used to realize the UAV indoor space target searching for a BCI system. The data set consists of recordings from twelve subjects. It contains 15-channel data sampled with 250 Hz [9].

Table 1. Summary of the datasets used for evaluating MI-EEG BCIs

Dataset	Study	# of Subjects	Channels	MI Tasks
DS1	[10, 11]	1	3 channels (C3, Cz, C4)	left hand, right hand
DS2	[12, 13]	3	64 channels	left hand, right hand, foot, and tongue
DS3	[14]	5	118 channels	right hand, foot
DS4	[15-17]	7	59 channels	left hand, right hand, foot (side chosen by the subject; optionally also both feet)
DS5	[12, 18-22]	9	22 channels	left hand, right hand, both feet and tongue
DS6	[6, 10, 15, 16, 18, 21, 23, 24]	9	3 channels (C3, Cz, C4)	left hand, right hand
DS7	[21, 25-27]	109	64 channels	opening and closing left fist, right fist, both fists, both feet
DS8	[6]	5	5 channels (C3, C1, Cz, C2, C4)	left hand, right hand
DS9	[12]	52	64 channels	left hand, right hand
DS10	[12]	13	19 channels	left hand, right hand, left leg, right leg, tongue
DS11	[9]	12	15 channels	left hand, right hand, feet, tongue

3. SIGNAL-TO-IMAGE CONVERSION METHODS AND CLASSIFICATION

The conversion of MI-EEG signals to 2D images has usually performed by time-frequency representation approaches. They transform MI-EEG signals into 2D time-frequency images, i.e., spectrograms or scalograms. Many other characteristic methods have also been researched extensively. Earlier studies on the classifications of MI-EEG signal images were mostly conducted with deep learning. It is mainly due to their recent success, where they work well for automatic feature extraction and classification. Apart from deep learning, several traditional machine learning approaches have also been engaged. Deep learning models need large sample data sets to achieve good performance. Nevertheless, collecting large-scale and high-quality MI-EEG data is difficult due to the strict requirements for subjects and experimental environments. Deep transfer learning and data augmentation are promising candidates to solve these problems. They have much potential; nonetheless, the number of studies in this field is limited.

A comprehensive review of the classification and signal-to-image conversion methods is shown in Table 2. Studies are first grouped according to the data sets. In some studies, specific MI categories were classified, while in others, additional categories were constructed. Consequently, the classified MI tasks are given in the third column. The next column gives the signal-to-image conversion and classification methods, respectively. Important features specific to the application of these methods are also emphasised. In the last column, performance results are given regarding accuracy or Kappa. It also explains how the results are obtained. For example, some studies use the original training and test sets, some use only the training or test set, and some use k-fold cross-validation.

Table 2. An extensive review of the classification and signal-to-image conversion methods

Dataset	Study	MI Tasks	Methods	Results and Experiment Details
DS1	[10]	2 (left/right hand)	STFT + CNN-SAE	Acc: 90.0% and K: 0.8; using original training/test sets
	[11]	2 (left/right hand)	CWT + AlexNet	Acc: 96.43% and K: 0.93; using original training/sets
DS2	[12]	4 (left/right hand, foot, and tongue)	STFT + EEGNet + VMD	Acc: 91.37%±2.12%; 10-fold CV
	[13]	4 (left/right hand, foot, and tongue) 2 (left/right hand)	AAG + SIFT+ BoW + kNN	Acc: 97.99%; using original training/test sets Acc: 96.50%; using original training/ test sets
DS3	[14]	2 (right hand and right foot)	CWT + CNN STFT + CNN	Acc: 99.35% and K: 0.9869 Acc: 98.7% and K: 0.9798
DS4	[15]	2 (left/right hand)	STFT + CNN without DA	Acc: 74.5±4.0% and K: 0.4018±0.048; 10-fold CV; data of b, d, e, g subj
			STFT + CNN-DCGAN	Acc: 83.2±3.5% and K: 0.4679 ± 0.050; 10-fold CV; data of b, d, e, g subj
	[16]	2 (left hand and both feet for subj a and f, left hand and right hand for subj b and g)	STFT + PCNN	Acc: 84.5±1.5% and K: 0.690±0.029; 5×10-fold CV; data of a, b, f, g subj
	[17]	2 (left hand and foot for subj a and f, left hand and right hand for subj b and g)	Morlet Wavelet + 3DCNNs + Bi-GRUs	Acc: 64.93%; 8x8 CV; data of a, b, f, g subj
DS5	[18]	4 (left/right hand, both feet and tongue)	STFT + Deep CNN and CutCat DA	Acc: 75.81% and K: 0.678±0.198
	[21]	4 (left/right hand, both feet and tongue)	TPCT + mVGG	Acc: 88.87% and K: 0.57; using original training/test sets Acc: 92.13% and K: 0.90; training/test sets are merged and evaluated with 10-fold CV
	[12]	4 (left/right hand, both feet and tongue)	STFT + EEGNet + VMD	Acc: 94.41±2.74%; 10-fold CV

Table 2 (continued). An extensive review of the classification and signal-to-image conversion methods

Dataset	Study	MI Tasks	Methods	Results and Experiment Details
DS5	[20]	4 (left/right hand, both feet and tongue)	Morlet wavelet + cubic spline interpolation + IncepCNN-BGRU	Acc: 76.62% and K: 0.688; 10×10-fold CV
	[22]	4 (left/right hand, both feet and tongue)	ETR + CNN	Acc: 87.66±5.34% and K: 0.82±0.081
		5 (left/right hand, both feet and tongue tasks, and others)	ETR + CNN	Acc: 85.57±7.08 and K: 0.801±0.088
	[19]	2 (left/right hand)	CWT without spatial dropping + CNN	Acc: 87.6±5.7% and K: 0.75; 10-fold CV
		3 (left/right hand, and foot)	CWT with spatial dropping + CNN	Acc: 71.2±7.0% and K: 0.56; 10-fold CV
	[25]	4 (left/right hand, both feet and tongue)	STFT + KBIM + PMMCL [Case 1]	Acc: 90.13%; 10-fold CV
			STFT + KBIM + PMMCL [Case 2]	Acc: 77.33%; 10-fold CV
DS6	[6]	2 (left/right hand)	STFT + CNN-VAE	K: 0.564±0.065; 10×10-fold CV
	[15]	2 (left/right hand)	STFT + CNN without DA	Acc: 80.6±3.2% and K: 0.4789±0.077; 10-fold CV
			STFT + CNN-DCGAN	Acc: 93.2±2.8% and K: 0.671±0.067; 10-fold CV
	[23]	2 (left/right hand)	STFT + pre-trained VGG-16 CNN + target CNN	Acc: 74.2%; ten runs on training/testing sets with 80/20% splits using only first three sessions
	[24]	2 (left/right hand)	STFT + CapsNet	Acc: 78.44%; using original training/test sets
	[16]	2 (left/right hand)	STFT + PCNN	Acc: 83.0±3.4% (with 2.4 inter subj std. dev) and K: 0.659±0.067 (with 0.048 inter subj std. dev.); 5×10-fold CV
	[18]	2 (left/right hand)	STFT + Shallow CNN + CutCat DA	Acc: 78.44% and K: 0.569; 10-fold CV
	[10]	2 (left/right hand)	STFT + CNN-SAE	Acc: 77.6±2.1% (with 8.1 inter subj std. dev.) and K: 0.547±0.083 (with 0.161 inter subj std. dev.); 10×10-fold CV on 3 sessions in training set
	[21]	2 (left/right hand)	TPCT + mVGG	Acc: 96.82% and K: 0.94; training and test sets are merged and evaluated with 10-fold CV Acc: 96.48% and K: 0.93; 10-fold CV on training set
DS7	[25]	5 (opening & closing left fist, right fist, both feet, both fists, and eye closed)	STFT + KBIM + PMMCL [Case 1]	Acc: 99.64%;10-fold CV
			STFT + KBIM + PMMCL [Case 2]	Acc: 97.42%; 10-fold CV
	[26]	3 (left/right hand or rest classes)	ST + triplet network	Acc: 0.647
	[27]	2 (open & close left or right fist)	FFT + azimuthal equidistant projection + Deep CNN	Acc: 90.52% and K: 0.81; 10-fold CV
	[21]	2 (both fists and both feet)	TPCT + mVGG	Acc: 88.62% and K: 0.77; 10-fold CV; data from all subj are merged
DS8	[6]	2 (left/right hand)	STFT + CNN-VAE	K: 0.568±0.068 with 3 electrodes; 10×10-fold CV K: 0.603±0.067 with 5 electrodes; 10×10-fold CV
DS9	[12]	2 (left/right hand)	STFT + DeepConvNet + VMD	Acc: 88.51±10.64%; 10-fold CV
DS10	[12]	5 (left/right hand, left/right foot, and tongue)	STFT + EEGNet + VMD	Acc: 90.20±4.34%; 10-fold CV
DS11	[9]	4 (left/right hand, feet, and tongue)	STFT + CNN	Acc: 88.06% for calibration, 88.95% for indoor 3D space target searching

Some acronyms are as follow: STFT (Short-time Fourier transform), CWT (continuous wavelet transform), FFT (fast Fourier transform), ETR (EEG topographical representations), TPCT (Clough-Tocher interpolation-based imaging), ST (Stockwell Transform), AAG (angle-amplitude graph), CNN (convolutional neural network), SAE (Stacked Autoencoders), VAE (Automatic Variation Encoder), VMD (variational mode decomposition), PCNN (parallel CNN), 3DCNNs (three-dimensional CNNs), modified visual geometry group network (mVGG), BGRU/Bi-GRUs (bidirectional gated recurrent units), (PMMCL) parallel multimodule CNN and long short-term memory network, DCGAN (deep convolutional generative adversarial network), SIFT (scale-invariant feature transform), BOW (bag-of-words), kNN (k-nearest neighbour algorithm), DA (data augmentation), KBIM (key band imaging method), CV (cross-validation), (K) Kappa, subj (subject).

3.1 Time-frequency image representation methods

The human brain is a complex nonlinear system that produces non-stationary EEG oscillations [28]. Conventional Fourier-based methods are successful in investigating the spectrum of stationary signals. However, they often fail in MI-EEG, where the spectrum changes with time [29]. Time-frequency image representation (TFIR) methods have been used in the literature to comprehend which frequency components exist in different time intervals and how they vary over time. They transform 1D signals into 2D time-frequency representations and provide information about local frequency variations [30]. Short-time Fourier transform (STFT), and Continuous wavelet transform (CWT) are widespread examples applied to numerous problems [14]. TFIR especially has constraints in dealing with fast varying instantaneous frequencies [30], and one of the problems is finding suitable time-frequency resolution [29]. For instance, traditional STFT with invariable window width provides a constant resolution over the entire spectrum [14], and therefore, it suffers poor resolution [29]. Compared to STFT's constant resolution over the entire spectrum, CWT can provide a higher resolution [14]. On the other hand, CWT can also fail because its window dilation rule is not signal-dependent [29]. In the literature, STFT, CWT and Stockwell transform has been employed as TFIR for MI-EEG signals.

3.1.1 Short-time Fourier transform-based methods

STFT have been one of the most used methods to convert MI-EEG signals into spectrums. In some studies, these spectrums were given to the classifiers as image input directly and without processing, while in others, sub-images of the spectrum were extracted and combined. Thus, useful MI-EEG information was used together. In many studies, to capture event-related desynchronization and synchronization, μ and β bands were extracted and combined. In such combinations, sub-images related to frequency bands were usually resized to the same size to ensure similar effects for all bands. If STFTs from different electrodes were to be combined, the electrodes neighbouring information were also considered. Thus, STFT-based time-frequency and electrode location information was fused. STFT-based approaches are summarized in Table 3. The table shows the data sets, EEG channels, STFT parameter values (signal length, window size, time-lapse, overlap), and frequency bands from which MI information was obtained.

Tabar and Halici [10] combined time, frequency, and electrode location information to create a new input form for MI-EEG signals. They used it in CNNs, with one 1D convolutional and one max-pooling layer. First, they applied STFT to MI-EEG signals and calculated 257×32-sized spectrum images. After that, to capture and use event-related desynchronizations and synchronizations, they extracted 16×32 and 23×32-sized sub-images related to μ and β frequency bands, respectively. To ensure similar effects of both bands, they resized the 23×32-sized sub-images to 15×32. After that, they vertically combined sub-images of β and μ bands, respectively, and constructed 31×32-sized images for a one-channel MI-EEG signal. Lastly, they vertically combined images of C3, Cz, and C4 channels while preserving neighbouring information and formed a 93×32-sized final input image, as shown in Figure 1. Thus, they used MI signals' time and frequency properties to design network input. The experiments were conducted on DS1 and DS6 to classify left/right-hand tasks. This study combined the CNN and stacked autoencoders (SAE), and the features extracted in CNN were classified through the deep network SAE. The STFT + CNN-SAE approach showed 90.0% accuracy with a kappa of 0.8 on DS1 and 77.6±2.1% with a kappa of 0.547±0.083 on DS6. The results for DS6 are from nine subjects and are slightly lower as they include data from different sessions. However, the proposed method achieved a 9% improvement over the competition winner.

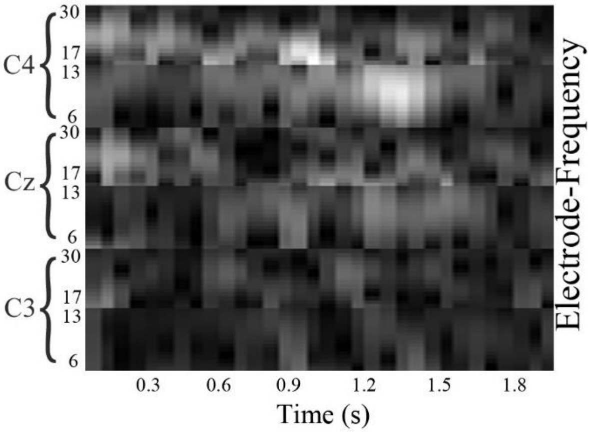

Figure 1. A sample input MI-EEG image where STFT spectrums are extracted, resized, and fused [10]

The same paradigm was used differently, where spectrum images were extracted, resized, and combined. For example, Dai et al. [6] used the electrode location, time, and frequency information and defined new multidimensional input features. They extracted sub-images for β and μ frequency bands from 257×32-sized spectrums. The sub-images were resized to 15×32 to establish the similar effects of both bands. After that, these sub-images were combined vertically, and 31×32-sized images were constructed per channel. In the last stage, images from different channels were vertically combined while preserving the electrodes' neighbouring information. They presented a new classification framework that uses an Automatic Variation Encoder (VAE) after CNN to deal with artefacts, noise, channel correlation, and high-dimensionality problems. The STFT + CNN-VAE method showed an average kappa of 0.564 on DS6 with a 3% improvement. Furthermore, their dataset DS8 achieved the best average kappa of 0.568 with three electrodes and 0.603 with five electrodes.

Xu et al. [23] investigated an STFT-based preprocessing procedure based on time-frequency images. After converting MI-EEG signals to spectrums, they extracted 20×32 and 33×32-sized sub-images for 4–14 Hz μ and 16–32 Hz β bands, respectively, and resized each of these sub-images to 112×224. Then, they combined them and generated 224×224-sized spectrums. Finally, a 224×224×3-sized input image was created by combining spectrum images converted from three EEG channels. After that, to solve the insufficient labelled data problem and enhance the training efficiency of the CNN, they adopted a VGG-16-based deep transfer CNN framework. The model was pre-trained on ImageNet, and the parameters of the VGG-16 CNN were directly transferred to the target CNN. It shares the same structure as VGG-16 except for the softmax output layer. In the target model, the front-layer parameters were frozen, while later-layer parameters were fine-tuned by the target dataset. For the classification of left/right-hand tasks on DS6, the proposed framework achieved 74.2% accuracy, better than most known approaches like standard CNN and support vector machines.

Chaudhary et al. [14] introduced a new deep convolutional neural network for classifying right-hand and right-foot tasks. They applied time-frequency techniques to convert MI-EEG signals into images. In this paper, they obtained spectrograms using the Hamming window, resized them to 227×227×3 and used them as input. After that, they employed a transfer learning strategy to fine-tune the pre-trained Alexnet DCNN. In this network, the last two layers were replaced with new ones to discover features unique to the dataset. The proposed STFT + CNN approach acquired 98.7% accuracy with a kappa of 0.98. In the same study, CWT showed a slightly better result than STFT. Ha and Jeong [24] transformed MI-EEG signals into 2D images and gave them to the capsule network (CapsNet). They formed 3×65×14-sized images by separately applying STFT to three channels. After selecting β and μ bands, they constructed three channel 14×14-sized 2D spectrum images as the initial input. They mainly focused on solving the problems of distorted and inconsistent MI-EEG signals by automatic feature learning and determined the optimal network architecture by analysing various CapsNet configurations. Experiments were conducted on DS6 to classify two-class left/right-hand MI tasks. The STFT + CapsNet approach showed an average accuracy of 78.44%, higher than CNN-based methods (ShallowNet et al.) and various conventional machine learning techniques (SVM et al.). On the other hand, for some subjects, the network architecture failed to capture better features and patterns, and due to this, some of the CNN-based approaches outperformed the proposed method.

Zhang et al. [15] decoded MI-EEG signals based on deep neural networks. They created STFT images of C3, Cz, and C4 channels and mosaicked them into an image while preserving the channel's neighbouring information. They investigated data augmentation methods for classifying spectrogram images. For this purpose, they generated artificial MI-EEG data using five data augmentation methods, i.e., geometric transformation, noise addition, generative model, autoencoder, variational autoencoder, and deep convolutional generative adversarial network (DCGAN). Among them, DCGAN performed better than traditional methods. It showed 17% and 21% accuracy improvements for DS4 and DS6, respectively, for the classification of left/right-hand tasks. The proposed CNN-DCGAN showed high consistency and outperformed the other methods, with average kappa values of 0.564 and 0.677 for these datasets. The results indicated that the data augmentation using operations like rotation may have adversely affected the MI-EEG information.

Al-Saegh et al. [18] employed the raw time-series MI-EEG signals to train a deep end-to-end CNN. They used STFT and converted MI-EEG signals to 2D images in this approach. These images preserve event-related desynchronisation and synchronisation-related information. They applied STFT to capture power spectral density and constructed 250×67-sized images per channel. After that, as with previous studies, μ and β band-related sub-images were extracted, resized, and concatenated. Finally, 96×67-sized final images were formed by combining the electrodes' location information. This paper also presented an augmentation method and enlarged the MI-EEG dataset to solve the same data scarcity problem. It generated trials from inter and intra- subjects and trials. The generated images were then input to the shallow and deep CNNs. The STFT + Deep CNN and CutCat data augmentation showed 75.81% accuracy with a kappa of 0.68 for a four-class (left/right hand, both feet and tongue) problem on DS5. Besides, the STFT + Shallow CNN + CutCat data augmentation acquired 78.44% accuracy with a kappa of 0.57 for a two-class (left/right hand) problem on DS6. Thanks to the favourable augmentation method for small-scale datasets, the improved MI-EEG decoding and implemented networks achieved good results compared to the state-of-the-art.

Table 3. Literature review of STFT-based TFIR approaches

Study	Dataset	Length of Signal	Channels	STFT Parameters	Frequency Bands
[10]	DS6	2 s (500 samples)	C3, Cz and C4	window size: 64 / time lapse: 14	μ: 6–13 Hz
	DS1	2s (256 samples)	C3, Cz and C4	window size: 32 / time lapse: 7	β: 17–30 Hz
[6]	DS6	2 s (500 samples)	C3, Cz, and C4	window size: 64 / time lapse: 14	μ: 6–13 Hz
	DS8	2 s (500 samples)	C3, C1, Cz C2, and C4		β: 17–30 Hz
[23]	DS6	2 s (500 samples)	C3, C4, and Cz	window size: 64 / overlap size: 50	μ: 4–14 Hz β: 16–32 Hz
[14]	DS3	-	-	window size: 120 / overlap: 100	-
[24]	DS6	2 s (500 samples)	C3, Cz, and C4	window size: 140 / overlap size: 100	8–31 Hz
[15]	DS4	4 s (400 samples)	C3, Cz, and C4	window size: 128	8–30 Hz
	DS6	4 s (1000 samples)	C3, Cz, and C4	window size: 256	
[18]	DS5 DS6	4 s (1000 samples)	C3, Cz, C4	window size: 64 / overlap size: 50	μ: 6–13 Hz β: 17–30 Hz
[25]	DS7	2 s (320 samples)	64 channels	window size: 128 / time lapse: 50% of the window	α: 8–13 Hz
	DS5	2 s (500 samples)	22 channels		β: 13–30 Hz
[12]	DS2	5 s (1250 samples)	C3, Cz, C4	window size: 200 ms / overlap: 75% of the window	delta, theta, alpha and beta bands
	DS5	5 s (1250 samples)			
	DS9	4 s (2048 samples)			
	DS10	1 s (200 samples)			
[16]	DS6	2 s (500 samples)	C3, Cz, C4	window size: 64 / time lapse: 14	μ: 6–13 Hz
	DS4	5 s (500 samples)	59 channels		β: 17–30 Hz
[9]	DS11	3 s (750 samples)	15 channels	windows size: 0.5 s / overlap size: 97% of the window	1–100 Hz and 8–30 Hz

Some other studies, e.g., [9, 12, 16, 25], have sought to increase the performance of MI-EEG systems using different STFT-based approaches. In these studies, STFT was generally used with different techniques, e.g., common spatial patterns or spectrum fusion methods. For instance, Han et al. [16] combined spatial filtering and frequency band extracting to create a new image form for the raw MI-EEG signal representation. They used STFT both to obtain MI-related frequency bands and describe MI-EEG signals. They applied STFT after a regularised common spatial pattern (RCSP) and proposed a new method called RCSP-STFT. It has the advantage of combining spatial projection and time–frequency analysis. As in previous studies, sub-images of μ and β bands were extracted, resized (using neighbour interpolation method), and combined vertically in this study. For a single trial, 30 representation images were calculated with the help of 30 pairs of two regularisation parameters. These images were fed to a parallel convolutional neural network (PCNN) architecture. In addition to the usual 2D kernel, 1D kernels for both frequency channel and time were added to optimise the PCNN and sufficiently capture features from the 2D images. Batch normalisation and dropout strategies were also engaged to optimise the network's performance. The proposed STFT + PCNN method yielded an average accuracy of 83.0% with a kappa of 0.66 on DS6 to classify left/right-hand tasks. It outperformed the state-of-the-art by at least 5.2% accuracy and 20.5% kappa improvements.

MI-EEG has high time resolution and frequency-spatial characteristics. Because of the nature of the MI-EEG, the features of α and β frequency band features might not fully be extracted by neural networks. Considering these and for automatic classification, Li et al. [25] constructed two key band image sequences containing temporal-spectral-spatial features. They constructed time-frequency images with STFT and then extracted sub-images of α and β bands. Afterwards, they fused the sub-images from all electrodes for α and β bands. Lastly, each band was arranged to the electrode coordinate using the nearest neighbour interpolation. In this paper, they further investigated the performance of different interpolation methods. They applied α and β band image sequences to a hybrid deep neural network, i.e., parallel multimodule CNN and long short-term memory network (PMMCL) network, to extract and fuse spatial-spectral and temporal features of key band image sequences. In the hybrid DNN, parallel multimodule CNNs were used to simultaneously acquire the global features of frequency band image sequences. LSTM was used to extract the temporal features among key band image sequences. A sliding window technique was also employed to expand the dataset. The proposed STFT + KBIM + PMMCL method showed 77.33% accuracy on DS5 for a four-class MI problem and 97.42% on DS7 for a five-class MI problem, better than the state-of-the-art. However, computational costs due to data augmentation and networking should be reduced. Keerthi Krishnan and Soman [12] aimed to generate spectrum images for preprocessing and inputting to CNNs. For this purpose, they decomposed MI-EEG signals into four variational mode decomposition (VMD) modes and applied STFT to each. Then, they combined the STFT of each VMD mode by stacking and formed the final spectrum image. This study used ConvNet, EEGNet, AlexNet, and LeNet CNN architectures. ConvNet and EEGNet were modified with three channel layers as the RGB spectrum images. AlexNet and LeNet adopted a network without layer modifications and fine-tuned these CNN architectures for spectral image input. The STFT + EEGNet + VMD framework showed average accuracies of 91.37% and 94.41% for the four-class problems on DS2, DS5, respectively, and 90.20% for the five-class problem on DS10. Furthermore, the STFT + DeepConvNet + VMD acquired an accuracy of 88.51% for a two-class problem on DS9. These results reveal the potential of VMD-STFT in the recognition of MI-EEG signals.

Shi et al. [9] investigated a BCI that uses monocular vision and MI-EEG for unmanned aerial vehicle (UAV) indoor space target searching. In this BCI, the navigation subsystem offers the precise, 3D-space-feasible flying direction to the decision subsystem. The decision subsystem first filtered raw MI-EEG signals with two fifth-order Butterworth band-pass filters. After that, common spatial patterns were applied to realise the spatial transformation. Then, spatially filtered signals were converted to images using stacked spectrograms and data augmentation. These spectrograms were inputted into a single convolutional layer CNN for feature extraction and classification. The proposed method yielded a calibration accuracy of 88.06% and an indoor 3D space target searching accuracy of 88.95% for a four-class (left/right hand, feet, and tongue) problem on DS11. The BCI system showed good adaptability and control stability for indoor 3D space target searching.

3.1.2 Wavelet-based methods

Wavelet-based methods are another means of extracting time-frequency patterns and converting MI-EEG signals to images. As in STFT-based methods, Wavelet-based TFIR has been constructed in different ways. In some studies, these images were given as input directly and without processing to the classifiers, while in others, some MI-related parts were extracted and fused. In most of them, electrode neighbouring information was considered, and topographic interpolation approaches were used. Chaudhary et al. [14] analysed non-stationary MI-EEG signals in time-frequency domains using CWT because of higher time-frequency resolution compared to the constant resolution of STFT. They applied CWT to MI-EEG signals, constructed scalograms, resized them to the size of 227×227×3, and used them as DCNN inputs. This paper also investigated a deep convolutional neural network (DCNN) based technique for classifying two-class tasks (right-hand and right-foot). A transfer learning strategy was employed to fine-tune the pre-trained Alexnet DCNN. The proposed STFT+CWT approach achieved 99.35% accuracy with a kappa of 0.99. In the same study, STFT showed slightly worse results. Collazos-Huertas et al. [19] created an input image set for CNNs. They calculated CWT coefficients and TFIR with the help of a complex Morlet wavelet. The representations were extracted within each time window, generating an image containing temporal, spectral, and spatial information. Besides, a topographic interpolation technique was used to project multi-channel data and preserve the spatial knowledge of electrodes. The performance was tested using MI-related different spectral bandwidths. Four combining scenarios of μ and β rhythms were further evaluated, showing that CWT is more suitable for the non-stationary data decomposition compared to power spectral density. In this paper, a CNN approach based on non-sequential Wide&Deep neural networks was proposed. A spatial dropping technique was also employed to eliminate the learning weights reflecting the not engaged localities. This paper focused on the improved interpolation of spatial neural patterns and acquired good performance values on DS5. The CNN architecture was tuned for discriminative MI, and the spatial dropping algorithm was evaluated. CWT without spatial dropping + CNN approach showed 87.6% accuracy with a kappa of 0.75 for a two-class problem, and CWT with spatial dropping + CNN showed 71.2% accuracy with a kappa of 0.56 for a three-class problem. The results showed that the CWT-based vectors were preferable for interpretation, learned weights were less sensitive to overtraining, and the CWT-based weights smoothly changed over time. Wei et al. [11] used a wavelet transform threshold denoising and a finite impulse response filter to eliminate extraneous signals and artefacts from MI-EEG recordings. They used CWT to create 2D images containing time, scale, and amplitude. They adopted the Morlet wavelet as the base function, which has the feature of equal variance in time and frequency, and employed the wavelet transform threshold denoising. Ultimately, they horizontally combined CWT images of the C3 and C4 electrodes. In experiments, a deep transfer learning strategy was developed for a small MI-EEG data set. The final layers of Alexnet were fine-tuned, and transfer learning was used to train the pre-trained AlexNet CNN. The experiments were conducted on DS1, and the proposed CWT + AlexNet approach achieved 96.43% accuracy with a kappa of 0.93 for the classification of the left/right hand MI tasks. Qiao et al. [20] offered a novel feature learning and preprocessing approach. They studied multi-channel MI-EEG signals and represented them while preserving spatial and temporal information. They used Morlet wavelet and cubic spline interpolation in preprocessing and constructed spectral frames. Lastly, they used an interpolation technique to transform each spectral irregularity map into a rectangle, which was given as input to the CNNs. They proposed a spatial-temporal hybrid deep learning model combining IncepCNN and bidirectional gated recurrent unit (BGRU) in classification. The proposed CNN structure was abbreviated as IncepCNN and used to learn and extract spatial features. After that, the output of IncepCNN was used as the input of BGRU to learn the temporal information existing in MI-EEG. The proposed method achieved 76.62% accuracy with a kappa of 0.69 for the classification four-class MI task on DS5, encouraging the problem of inter- and intra-class variations. Wei and Lin [17] sought to solve the problems of poor performance, low efficiency and weak robustness using a multi-dimensional fusion features-based classification. They used an improved Morlet wavelet algorithm to extract stable features from the frequency spectrum, especially for nonlinear and non-stationary signals. After applying Morlet, spectral power maps from 1 Hz to 40 Hz were obtained with frequency bands of 1 Hz. After that, a cubic spline interpolation method was used to convert each spectral map to the corresponding rectangular spectrum energy map. Finally, the entire EEG trial was sampled into time slices, and the temporal-spatial-frequency event-related spectral power features were extracted due to the changes in event-related desynchronisation/synchronisation over time. This paper used a three-dimensional convolutional neural network (3DCNNs) model to extract features. They were subsequently put into the bidirectional gated recurrent units (Bi-GRUs) models to extract the spatial-frequency-sequential multi-dimensional fusion features. The proposed Morlet Wavelet + 3DCNNs + Bi-GRUs method achieved 64.93% accuracy on DS4 for some two-class MI problems. The experiment was also conducted for action observation and action execution tasks. The proposed method also achieved stable classification results in dealing with these tasks.

3.1.3 Stockwell transform

Stockwell Transform (ST) is another method to transform MI-EEG signals from time to the frequency domain. Due to this, it has been used to convert 1D EEG signals to 2D images. However, there are very few studies with ST compared to STFT and CWT. For example, Alwasiti et al. [26] introduced a triplet network to categorise MI-EEG signals for the first time. Unlike the state-of-the-art approaches, in this paper, log-scaling provided better spectrograms and emphasised the μ and β rhythms more. Due to this, they used the log-scaled frequency range of 2-78 Hz in the TFIR. Besides, according to experiments, incorporating a larger frequency range improved the accuracy of deep metric learning. In this paper, ST was applied to each channel of single-epoch signals, and 64 spectrograms were generated, as shown in Figure 2.

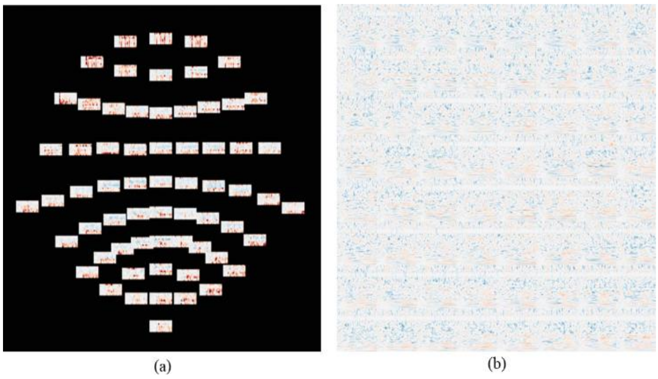

Figure 2. (a) ST spectrograms of 64 electrodes placed according to the 10-10 system, (b) 512×512-sized fused image generated from all spectrograms [26]

After that, the 64 spectrograms were fused on a 512×512-sized rectangular topogram area. Finally, the image dataset was normalised using the mean and standard deviation of the pixel values. Using only a small training set, the authors developed a triplet deep metric network (DML) to solve a three-class problem (left/right hand or rest classes). Thanks to this network, the distance between the embeddings of different labelled images was increased while the distance between the embeddings of spectrograms of the same class was minimised. The experiments were conducted on DS7, which has 109 untrained subjects, and an average accuracy of 0.647 was achieved. It meant the DML could converge with and obtain good results even using a small number of training data. Furthermore, the Stockwell transform consistently outperformed STFT in almost all subjects. Due to these successful results, the techniques used with STFT and CWT for classifying MI-EEG signals should also be evaluated with Stockwell Transform. At the same time, more successful MI-EEG systems can be developed by combining the outputs of all these TFIR methods.

3.2 Other characteristic methods

Converting MI-EEG signals to images has not been carried out with only TFIR approaches. Unlike the TFIR methods discussed above, several studies have used different techniques. For example, the true electrode location information can be lost when converting MI-EEG signals to images, decreasing the classification performance. For this purpose, Li et al. [21] investigated the time domain power and Clough-Tocher interpolation-based imaging (TPCT) to construct 64x64-sized time-frequency-space features-based TPCT images, where each electrode's real positions were used to locate time-frequency characteristics. In this method, a fast Fourier transform (FFT) was first applied to each time window for each channel. The 8-30 Hz frequency range was divided into three sub-bands, i.e., 8-13 Hz, 13-21 Hz and 21-30 Hz, respectively. Then, an inverse FFT was applied, and the average power of the time domain was calculated independently for each of the three sub-bands. After that, a 64×64-sized grid system was established, time-frequency features were placed in a 2D space using the Clough-Tocher algorithm, and final MI-EEG input images were constructed, respectively. This study also modified the visual geometry group network (VGG) to MI-EEG BCIs and proposed a new deep CNN abbreviated mVGG. In this network, the convolution layer was used instead of max pooling, and average pooling was used instead of the fully connected layer for efficient recognition. The convolution kernel size was also modified, and the layers were deepened. The proposed TPCT + mVGG approach achieved 92.13% accuracy on DS5 for a four-class problem and 96.82% on DS6 for a two-class problem when evaluating with a 10-fold CV. The effectiveness of the imaging method, which had lower class skew and error costs, was also demonstrated by kappa values and ROC curves. Wang and Li [27] aimed to decode MI-EEG tasks using a deep parallel CNN (pDCNN) in an efficient way. They used 8-13Hz, 13-21 Hz and 21-30 Hz sub-bands by redividing μ and β bands. They divided the raw data into segments, transformed each time segment into a frequency domain using FFT, and calculated the average power of three sub-bands. They calculated the 2D positions of 64 3D electrode locations using an azimuthal equidistant projection and constructed three MI-EEG spatial-frequency images using the Clough-Tocher interpolation. In this paper, the authors aimed to overcome the insufficient spatial-frequency feature extraction for MI-EEG signals. Due to this, they extracted three groups of features corresponding to the three sub-bands and fused them using pDCNN. The methods were evaluated on DS7, including 109 subjects, and 90.25% average accuracy with a kappa value of 0.81 was achieved for classifying two-class MI tasks (open & close left or right fist).

Finding a reliable approach to support high-dimensional MI-EEG data with poor signal-to-noise ratios is a difficult problem. Xu et al. [22] investigated a new EEG topographical representations (ETR) approach to address this problem and efficiently learn brain activities. They produced ETR topologies that could be functional and correct for spatial location, temporal onset, and stability. The proposed ETR data structure was also designed to be useful for dimension reduction and reflecting intrinsic brain activity connections. In the last stage, ETR data-based spectral-spatial inputs were given to an ETR-based CNN (ETRCNN) learning framework. The performance of ETR + CNN was evaluated on DS5, and the framework achieved 87.66% accuracy (with a kappa of 0.82) for a four-class and 85.57% (with a kappa of 0.801) for a five-class problem. These results showed that the framework could accomplish multi-period and multi-object recognition. Yilmaz et al. [13] used an angle-amplitude transformation technique, a simple signal-to-image transformation approach for the MI-EEG and MEG signals and formed angle-amplitude graph (AAG) images. Then, they employed scale-invariant feature transform (SIFT)-based bag-of-words (BOW) features to extract image features. These features were then inputted into the k-nearest neighbour algorithm for classification. The experiments were conducted on DS2, and the proposed AAG + SIFT+ BoW + kNN approach showed an average accuracy of 97.99% for four-class and 96.50% for two-class MI problems. This study used a straightforward method for classification compared to other methods using deep learning.

4. CHALLENGES AND FUTURE DIRECTIONS

Several problems and challenges have been encountered in classifying images of MI-EEG data. Some important ones are discussed in this section to provide an overview of the potential ability of systems. They can be categorised into three main groups: (i) MI and EEG-related, (ii) signal-to-image conversion-related, and (iii) deep learning-related factors.

4.1 Motor imagery and EEG-related factors

Non-invasive MI-EEG signals are recorded from electrodes attached to the scalp surface, and EEG is a weak signal in a sea of noise and artefacts. For example, eye-blinking, muscle movement, and cardiac pulsation artefacts affect and suppress the useful information in MI-EEG signals [6, 18]. Therefore, the signals must be cleaned of all such artefacts in the preprocessing. However, EOG or EMG signal information would be needed. The human body has a complex nervous system, so the human brain produces non-stationary and nonlinear MI-EEG signals. In particular, the non-stationary nature is a significant problem that needs to be solved.

EEG signals are simultaneously recorded from many scalp electrode locations. It causes problems like channel correlation and high dimensionality that complicate the design of MI-EEG systems [6]. In addition, volume conduction through the scalp, skull, and other brain layers compromises spatial resolution [31]. In the literature, C3, CZ, and C4 channels have been primarily used in classifying MI-EEG signals, but different channels can be successful in different MI tasks. Even neighbouring channels other than these channels can be more successful. Inter- and intra-subject variability is another major challenge. Because MI-EEG signals can be inconsistent and significantly distorted even in the same subject [24], these can happen in the same session or different sessions on the same day. Besides, the MI-EEG data vary subject to subject, leading to poor transfer learning for subject independence [32].

Capturing EEG signals that carry most of the discriminative information for any MI tasks is another big problem [32]. Specific frequency bands, i.e., alpha (μ) and β, hold useful MI activity information, and the performance of these bands can vary depending on the application [18]. For instance, Collazos-Huertas et al. [19] used just one rhythm (μ or β) and achieved unsatisfactory results. Compared to it, Alwasiti et al. [26] achieved best results with more detailed β sub-bands. In the same study, the combination of bands could not considerably improve the performance. On the other hand, Alwasiti et al. [26] introduced a deep metric learning approach incorporating a more extensive frequency range. They improved the performance with a rate of 5% without choosing the frequency of interest. Therefore, further study is needed to investigate the MI application-related frequencies.

Several approaches have been proposed in the literature for extracting useful information and solving the above problems. Deep learning-based approaches can allow end-to-end learning without feature engineering, which are excellent opportunities to eliminate some problems. For example, Ha and Jeong [24] presented a new CapsNet-based architecture in which various features from inconsistent MI-EEG signals were automatically learned, and good decoding results were achieved. Tayeb et al. [33] developed three DL models and decoded MI movements directly from raw EEG signals without manual feature engineering. They used a long short-term memory, a spectrogram-based CNN model, and a recurrent convolutional neural network (RCNN). Similarly, Hwaidi and Chen [34] used a variational autoencoder to remove noise from signals and increase the generalization capacity of MI-EEG classifiers. To overcome such problems, EOG recording, or an eye-blink detection technique is necessary for the majority of artefact reduction technologies. These methods should be able to work in real-life applications. For example, Sawangjai et al. [35] proposed a framework based on generative adversarial networks (GANs) and aimed to remove ocular artifacts using a data-driven assistive tool. EEG signals of subjects with poor MI execution performance should also be improved. To solve this problem, MI performance-based artefacts under poor skill must be removed as in Tobón-Henao et al. [36]. The large number of channels affects both performance and practical applications. Huang and Wei [37] investigated a tensor decomposition-based channel selection method to solve this problem. One of the most critical problems is inter and intra-subject variability, and transfer learning techniques are essential for addressing this. For example, Sun et al. [38] developed a subject transfer neural network to transfer the data distribution from BCI-friendly subjects to the data from more typical BCI-illiterate users and achieved good results for inter-subject variations.

4.2 Signal-to-image conversion-related factors

As the problems related to MI and EEG-related factors are solved, recognition and classification of MI-EEG signals become easier. However, classifying clean MI-EEG data is a complex problem in itself. In the literature, most signal-to-image conversion-based studies have used TFIR methods, enabling us to evaluate MI-EEG signals in both time and frequency domains [14]. As exhibiting a non-stationary nature, when dealing with rapidly varying instantaneous frequencies [30], TFIR have limitations in finding applicable time-frequency resolutions. Therefore, it is necessary to find a suitable time-frequency resolution. Besides, the proposed algorithms must be efficient regarding computation time and complexity. It is becoming more likely thanks to advances in computing power and freely available algorithms. Additionally, the parameters of TFIR methods, i.e., frequently used STFT and CWT for their remarkable properties, must be well adjusted. For instance, window and overlapping size, a window function (Hanning, Hamming, Kaiser, etc.), and the length of the FFT are important for STFT; and Wavelet type and scale, etc., are for CWT. Many studies have achieved the best results with different parameters, even for the same MI-EEG dataset. Therefore, it will be more reasonable if they are adjusted automatically. Zhong and Huang [29] examined this problem and proposed an adaptive short-time Fourier transform (ASTFT). They calculated the window width without prior signal knowledge and developed a signal-independent time-frequency analysis technique. Since the method is adaptive, it should be designed to achieve more accurate results with less computation.

For handling images differently, novel techniques must be investigated to make signal-to-image transformation methods more successful. For instance, TFIR approaches can be constructed after selecting appropriate frequency bands (or sub-bands) for specific MI applications, and then they can be combined, etc. Besides, they can be integrated or interpolated with other information (i.e., electrode locations). For example, combining the TFIR constructed using the electrodes from the motor cortex while preserving neighbouring information is a good approach. Because the electrode positions are lost in combining TFIR and the activation area of MI should be considered [21]. Different TFIR methods can also be fused because of their distinctive features. By way of illustration, CWT provides a higher time-frequency resolution compared to the constant resolution of STFT [14]. WT provides an alternative to the STFT because WT is of interest for analysing non-stationary signals. However, STFT takes the least time compared with CWT in processing time [39]. Therefore, better results can be achieved by combining the best aspects of different methods. In addition to TFIR, new techniques (as in Sect. 3.2) must be investigated to convert MI-EEG signals to images. In the literature, only a few studies have employed deep learning to model the MI-EEG feature representations, and nowadays, how to extract more in-depth features and abstract representations has become a research topic [20].

4.3 Deep learning and classification-related factors

The size of the datasets directly affects the performance of deep learning models [15]. Therefore, a large amount of data is needed to train deep learning models, especially when considering practical applications [23]. It limits the use of deep learning in the MI-EEG research field due to the need for more data [18]. Because collecting large-scale and high-quality MI-EEG data is very difficult for reasons such as the placement of the EEG device, successful acquisition of EEG recordings, strict requirements for experimental conditions, insufficient number of subjects, subject-to-subject, and session-to-session differences, etc. Also, data are usually unlabelled and labelling data manually is challenging. To eliminate these problems, apart from classical approaches, such as trying to increase performance by focusing on classification accuracy, different techniques in learning may lead to new solutions. Deep transfer learning and data augmentation are the most promising candidates, with great potential for MI-based BCIs.

In transfer learning, the model is not created from scratch. Instead, it uses the previously trained networks as a reference to create a new model for the current problem [14], i.e., transfer learning models from one domain to another. During training, it also automatically extracts richer and more expressive features [11]. Here, in comparison to the previously trained datasets, the new model is trained using the new dataset with fewer training images [14]. Transfer learning solves the problems encountered in deep learning that are difficult to improve because of the insufficient training samples, and it also significantly saves the time and cost of retraining the model. In the last few years, transfer learning has grown significantly in MI-based BCIs [11]. For example, Xu et al. [23] improved the training efficiency in limited labelled data conditions. The second approach is data augmentation, which enriches training data by creating new samples to increase the size of the dataset with more general training data [18]. By including new samples, data augmentation makes the training model more complex and decreases overfitting [15]. For instance, Al-Saegh et al. [18] attained augmentation by cutting specific time segments from the two trials, which may belong to the same or two different subjects, and then concatenating those separated time segments. Shi et al. [9] flipped and translated the spectrogram's rows and columns for augmentation. However, care should be taken to ensure the augmentation method suits MI-EEG signals. Because Zhang et al. [15] investigated data augmentation methods' performance and showed that augmentations like rotation may have adversely affected the MI-EEG information.

As mentioned before, MI-EEG data can be distorted and inconsistent even from the same subject [24]. Under these conditions, the classification performance is compromised at some point, and it is not easy to achieve high performance. These problems have been attempted to be solved by applying artefact removal methods or decoding with deep learning methods. However, these processes require extra costs and decoding raw MI-EEG signals is demanding [15]. In addition, training a deep learning model from the ground up is incredibly time-consuming and computationally expensive [23]. Transfer learning can solve these problems, but building models from scratch will produce better results. Deep learning methods such as CNN work well for image understanding and classification, especially in computer vision. However, the classification of MI-EEG signals is much more complicated, and the search for efficient methods is essential today.

5. CONCLUSIONS

This paper reviews the applications of signal-to-image conversion for MI-EEG data. It is essential because, with deep neural networks' growing popularity and success, converting time-series signals into 2D images has attracted considerable attention. Thus, we have searched the studies between 2019 and 2023 within the scope of non-invasive EEG-based MI applications. We first introduced and discussed the use of present signal-to-image encoding techniques. After that, the datasets, classification methods, challenges, and future directions were all discussed to obtain a deeper understanding. It was observed that only a few diverse approaches had been employed for signal-to-image conversions, where only electrode location, time, and frequency information were used to diversify methods. STFT is by far the most used method. After that, CWT and TPCT have been used a lot recently. As can be seen, most of them are TRIR-based approaches. Apart from them, a few methods have been reported, such as ETR and AAG. Available conversion approaches, although few, have great potential to be a solution that can either be used directly as input to the deep neural networks or with structural modifications. Regarding feature extraction and classification, almost all the reviewed articles have employed deep learning techniques like CNN. CNN has been used both on its and with different variants/combinations such as CNN-SAE, CNN-VAE, CNN-DCGAN, PCNN, 3DCNNs, IncepCNN, deep CNN, Shallow CNN, PMMCL, etc. Besides, some studies have used various data augmentation and pre-trained deep transfer learning techniques. EEGNet, DeepConvNet, and CapsNet are other widely used CNNs to recognise MI-EEG data. All these proposed methods have addressed many MI-EEG problems and achieved successful results. They can also be applied to other time series classification problems when 2D inputs are needed. This paper solely focuses on MI-EEG signal-to-image conversion. Further studies are needed to investigate the processing of MI-EEG, especially in the signal-to-image conversion and deep neural networks aspect. In addition, future studies should be carried out for systems that can operate in real-world scenarios, which should be the goal.

REFERENCES

[1] Blankertz, B., Muller, K.R., Curio, G., (2004). The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials. IEEE Transactions on Biomedical Engineering, 51(6): 1044-1051. https://doi.org/10.1109/TBME.2004.826692

[2] Blankertz, B., Muller, K.R., Krusienski, D.J., et al. (2006). The BCI competition III: Validating alternative approaches to actual BCI problems. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14(2): 153-159. https://doi.org/10.1109/TNSRE.2006.875642

[3] Tangermann, M., Müller, K.R., Aertsen, A., et al. (2012). Review of the BCI competition IV. Frontiers in Neuroscience, 6: 1. https://doi.org/10.3389/fnins.2012.00055

[4] Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R. (2004). BCI2000: A general-purpose brain-computer interface (BCI) system. IEEE Transactions on Biomedical Engineering, 51(6): 1034-1043. https://doi.org/10.1109/TBME.2004.827072

[5] Goldberger, A.L., Amaral, L.A., Glass, L., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation, 101(23): e215-e220. https://doi.org/10.1161/01.CIR.101.23.e215

[6] Dai, M., Zheng, D., Na, R., Wang, S., Zhang, S. (2019). EEG classification of motor imagery using a novel deep learning framework. Sensors, 19(3): 551. https://doi.org/10.3390/s19030551

[7] Cho, H., Ahn, M., Ahn, S., Kwon, M., Jun, S.C. (2017). EEG datasets for motor imagery brain–computer interface. GigaScience, 6(7): gix034. https://doi.org/10.1093/gigascience/gix034

[8] Kaya, M., Binli, M.K., Ozbay, E., Yanar, H., Mishchenko, Y. (2018). A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. Scientific Data, 5(1): 180211. https://doi.org/10.1038/sdata.2018.211

[9] Shi, T.W., Chang, G. M., Qiang, J.F., Ren, L., Cui, W.H. (2023). Brain computer interface system based on monocular vision and motor imagery for UAV indoor space target searching. Biomedical Signal Processing and Control, 79: 104114. https://doi.org/10.1016/j.bspc.2022.104114

[10] Tabar, Y.R., Halici, U. (2016). A novel deep learning approach for classification of EEG motor imagery signals. Journal of Neural Engineering, 14(1): 016003. https://doi.org/10.1088/1741-2560/14/1/016003

[11] Wei, M., Yang, R., Huang, M. (2021). Motor imagery EEG signal classification based on deep transfer learning. In 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), Aveiro, Portugal, pp. 85-90. https://doi.org/10.1109/CBMS52027.2021.00083

[12] Keerthi Krishnan, K., Soman, K.P. (2021). CNN based classification of motor imaginary using variational mode decomposed EEG-spectrum image. Biomedical Engineering Letters, 11(3): 235-247. https://doi.org/10.1007/s13534-021-00190-z

[13] Yilmaz, B.H., Yilmaz, C.M., Kose, C. (2020). Diversity in a signal-to-image transformation approach for EEG-based motor imagery task classification. Medical & Biological Engineering & Computing, 58: 443-459. https://doi.org/10.1007/s11517-019-02075-x

[14] Chaudhary, S., Taran, S., Bajaj, V., Sengur, A. (2019). Convolutional neural network based approach towards motor imagery tasks EEG signals classification. IEEE Sensors Journal, 19(12): 4494-4500. https://doi.org/10.1109/JSEN.2019.2899645

[15] Zhang, K., Xu, G., Han, Z., et al. (2020). Data augmentation for motor imagery signal classification based on a hybrid neural network. Sensors, 20(16): 4485. https://doi.org/10.3390/s20164485

[16] Han, Y., Wang, B., Luo, J., Li, L., Li, X. (2022). A classification method for EEG motor imagery signals based on parallel convolutional neural network. Biomedical Signal Processing and Control, 71: 103190. https://doi.org/10.1016/j.bspc.2021.103190

[17] Wei, M., Lin, F. (2020). A novel multi-dimensional features fusion algorithm for the EEG signal recognition of brain's sensorimotor region activated tasks. International Journal of Intelligent Computing and Cybernetics, 13(2): 239-260. https://doi.org/10.1108/IJICC-02-2020-0019

[18] Al-Saegh, A., Dawwd, S.A., Abdul-Jabbar, J.M. (2021). CutCat: An augmentation method for EEG classification. Neural Networks, 141: 433-443. https://doi.org/10.1016/j.neunet.2021.05.032

[19] Collazos-Huertas, D.F., Álvarez-Meza, A.M., Acosta-

Medina, C.D. et al. CNN-based framework using spatial dropping for enhanced interpretation of neural activity in motor imagery classification. Brain Informatics, 7: 8. https://doi.org/10.1186/s40708-020-00110-4

[20] Qiao, W., Bi, X. (2019). Deep spatial-temporal neural network for classification of EEG-based motor imagery. In Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science, Wuhan, China, pp. 265-272. https://doi.org/10.1145/3349341.3349414

[21] Li, M.A., Han, J.F., Duan, L.J. (2019). A novel MI-EEG imaging with the location information of electrodes. IEEE Access, 8: 3197-3211. https://doi.org/10.1109/ACCESS.2019.2962740

[22] Xu, M., Yao, J., Zhang, Z., et al. (2020). Learning EEG topographical representation for classification via convolutional neural network. Pattern Recognition, 105: 107390. https://doi.org/10.1016/j.patcog.2020.107390

[23] Xu, G., Shen, X., Chen, S., et al. (2019). A deep transfer convolutional neural network framework for EEG signal classification. IEEE Access, 7: 112767-112776. https://doi.org/10.1109/ACCESS.2019.2930958

[24] Ha, K.W., Jeong, J.W. (2019). Motor imagery EEG classification using capsule networks. Sensors, 19(13): 2854. https://doi.org/10.3390/s19132854

[25] Li, M.A., Peng, W.M., Yang, J.F. (2021). Key Band Image Sequences and A Hybrid Deep Neural Network for Recognition of Motor Imagery EEG. IEEE Access, 9: 86994-87006. https://doi.org/10.1109/ACCESS.2021.3085865

[26] Alwasiti, H., Yusoff, M.Z., Raza, K. (2020). Motor imagery classification for brain computer interface using deep metric learning. IEEE Access, 8: 109949-109963. https://doi.org/10.1109/ACCESS.2020.3002459

[27] Wang, L., Li, M. (2021). A novel DCNN based MI-EEG classification method using spatio-frequency information. In 2021 China Automation Congress (CAC), Beijing, China, pp. 532-537. https://doi.org/10.1109/CAC53003.2021.9727921

[28] Klonowski, W. (2009). Everything you wanted to ask about EEG but were afraid to get the right answer. Nonlinear Biomedical Physics, 3(1): 4. https://doi.org/10.1186/1753-4631-3-2

[29] Zhong, J., Huang, Y. (2010). Time-frequency representation based on an adaptive short-time Fourier transform. IEEE Transactions on Signal Processing, 58(10): 5118-5128. https://doi.org/10.1109/TSP.2010.2053028

[30] Wang, P., Gao, J., Wang, Z. (2014). Time-frequency analysis of seismic data using synchrosqueezing transform. IEEE Geoscience and Remote Sensing Letters, 11(12): 2042-2044. https://doi.org/10.1109/LGRS.2014.2317578

[31] Brunner, C., Naeem, M., Leeb, R., Graimann, B., Pfurtscheller, G. (2007). Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis. Pattern Recognition Letters, 28(8): 957-964. https://doi.org/10.1016/j.patrec.2007.01.002

[32] Liu, X., Lv, L., Shen, Y., Xiong, P., Yang, J., Liu, J. (2021). Multiscale space-time-frequency feature-guided multitask learning CNN for motor imagery EEG classification. Journal of Neural Engineering, 18(2): 026003. https://doi.org/10.1088/1741-2552/abd82b

[33] Tayeb, Z., Fedjaev, J., Ghaboosi, N., et al. (2019). Validating deep neural networks for online decoding of motor imagery movements from EEG signals. Sensors, 19(1): 210. https://doi.org/10.3390/s19010210

[34] Hwaidi, J.F., Chen, T. M. (2022). Classification of motor imagery EEG signals based on deep autoencoder and convolutional neural network approach. IEEE Access, 10: 48071-48081. https://doi.org/10.1109/ACCESS.2022.3171906

[35] Sawangjai, P., Trakulruangroj, M., Boonnag, C., et al. (2021). EEGANet: Removal of ocular artifacts from the EEG signal using generative adversarial networks. IEEE Journal of Biomedical and Health Informatics, 26(10): 4913-4924. https://doi.org/10.1109/JBHI.2021.3131104

[36] Tobón-Henao, M., Álvarez-Meza, A., Castellanos-Domínguez, G. (2022). Subject-dependent artifact removal for enhancing motor imagery classifier performance under poor skills. Sensors, 22(15): 5771. https://doi.org/10.3390/s22155771

[37] Huang, Z., Wei, Q. (2023). Tensor decomposition-based channel selection for motor imagery-based brain-computer interfaces. Cognitive Neurodynamics. https://doi.org/10.1007/s11571-023-09940-4

[38] Sun, B., Wu, Z., Hu, Y., Li, T. (2022). Golden subject is everyone: A subject transfer neural network for motor imagery-based brain computer interfaces. Neural Networks, 151: 111-120. https://doi.org/10.1016/j.neunet.2022.03.025

[39] Kıymık, M.K., Güler, İ., Dizibüyük, A., Akın, M. (2005). Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application. Computers in Biology and Medicine, 35(7): 603-616. https://doi.org/10.1016/j.compbiomed.2004.05.001