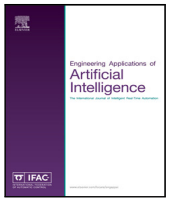
Engineering Applications of Artificial Intelligence 116 (2022) 105347

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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



Adaptive transfer learning-based multiscale feature fused deep convolutional neural network for EEG MI multiclassification in brain–computer interface

Arunabha M. Roy

University of Michigan, Aerospace Engineering, Ann Arbor, MI 48109, USA

ARTICLE INFO

Keywords:
Brain–computer interfaces (BCIs)
Transfer learning (TL)
Motor imagery (MI)
Electroencephalogram (EEG) signal classification
Convolutional neural network (CNN)
Deep learning (DL)

ABSTRACT

Objective. Deep learning (DL)-based brain–computer interface (BCI) in motor imagery (MI) has emerged as a powerful method for establishing direct communication between the brain and external electronic devices. However, due to inter-subject variability, inherent complex properties, and low signal-to-noise ratio (SNR) in electroencephalogram (EEG) signals are major challenges that significantly hinder the accuracy of the MI classifier.

Approach. To overcome this, the present work proposes an efficient transfer learning (TL)-based multi-scale feature fused CNN (MSFFCNN) which can capture the distinguishable features of various non-overlapping canonical frequency bands of EEG signals from different convolutional scales for multi-class MI classification.

Significance. In order to account for inter-subject variability from different subjects, the current work presents 4 different model variants including subject-independent and subject-adaptive classification models considering different adaptation configurations to exploit the full learning capacity of the classifier. Each adaptation configuration has been fine-tuned in an extensively trained pre-trained model and the performance of the classifier has been studied for a vast range of learning rates and degrees of adaptation which illustrates the advantages of using an adaptive transfer learning-based model.

Results. The model achieves an average classification accuracy of 94.06% (±0.70%) and the kappa value of 0.88 outperforming several baseline and current state-of-the-art EEG-based MI classification models with fewer training samples. The present research provides an effective and efficient transfer learning-based end-to-end MI classification framework for designing a high-performance robust MI-BCI system.

1. Introduction

Electroencephalogram (EEG) based motor imagery (MI) classification is a critical aspect of brain–computer interfaces (BCIs) which translate brain activities into recognizable machine commands to control the external electronic devices (Netzer et al., 2020; Parija et al., 2021). The typical workflow of an EEG-based MI BCI system has been illustrated in Fig. 1 that generally consists of the following phases: brain signal acquisition, data prepossessing, feature extraction, feature classification, and translation algorithm for various BCI applications (Zhang et al., 2019b; Rashid et al., 2020). The BCI allows rehabilitation of neuromotor disorders (Zhang et al., 2015), robotic control (Gupta et al., 2020), speech communication (Makin et al., 2020), etc. In BCI paradigms, MI classification is the most critical part in which brain signals can be translated into control signals (Zhang et al., 2019b; Rashid et al., 2020). During MI, desynchronization of neural activities triggers in the primary motor cortex contralateral to the movement which results in a decrease in $\mu - \beta$ rhythm known as event-related desynchronization (ERD) (Pfurtscheller et al., 2006; Yu et al., 2015). It is usually followed by an increase in the β rhythm called event-related synchronization (ERS) when MI ceases to exist. Thus, the main goal is to classify different MI tasks according to ERS and ERD characterized by changing the power spectrum of μ (8–14 Hz) and β (14–30 Hz) bands (Zhang et al., 2019b; Rashid et al., 2020).

In this regard, common spatial pattern (CSP) (Saa and Çetin, 2012) and filter bank CSP (FBCSP) (Ang et al., 2008) are efficient methods in extracting discriminative features that represent ERD and ERS for MI classification. FBCSP finds a set of linear projections that maximize the differences in the variance of the MI classes by employing filtered signals from various frequency bands which have shown to be effective in improving accuracy. However, brain signal from EEG has various distinct characteristics (i.e., non-linearity, uniqueness, and non-stationary behaviors) which significantly vary with the human brain and depend on the mental state of the individual subjects (Xing et al., 2020). Moreover, due to the presence of noise from various muscle artifacts, fatigue, change of environment, and internal body states may significantly alter the characteristic of EEG signals (Zhang et al., 2019b; Rashid et al., 2020). Hence, the aforementioned factors

E-mail address: arunabhr.umich@gmail.com.

https://doi.org/10.1016/j.engappai.2022.105347
Received 1 March 2022; Received in revised form 25 July 2022; Accepted 16 August 2022
Available online 30 August 2022

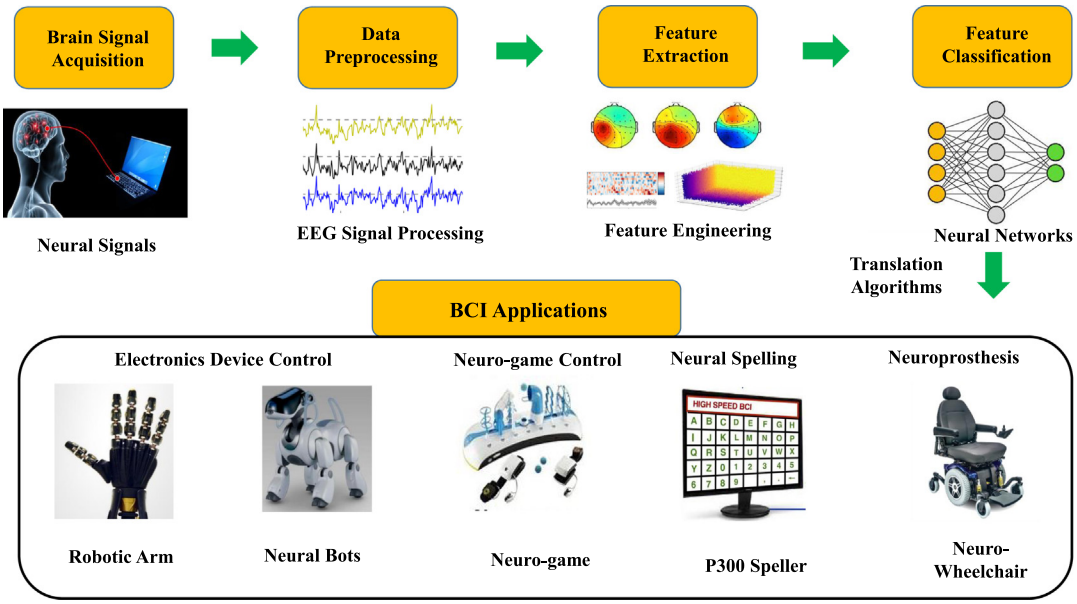


Fig. 1. The general workflow of a typical EEG-based MI-BCI system consists of brain signal acquisition, data prepossessing, feature extraction, feature classification, and translation algorithm for various BCI applications.

make it challenging to improve the signal-to-noise ratio (SNR) for better performance in MI classification.

Related works: With the recent advancement of deep learning (DL), it demonstrates superior performance in MI-BCI classification due to the inherent capability of adapting non-linear EEG signals and extracting important feature information automatically (Tabar and Halici, 2016; Schirrmeister et al., 2017). Thus, several studies are geared towards EEG signal classification employing convolutional neural network (CNN) (Chu et al., 2018; Dose et al., 2018; Sun et al., 2019; Zhu et al., 2019). More recently, there are various CNN based methods (Tang et al., 2020; Nour et al., 2021; Dai et al., 2020; Zhang et al., 2019a; Liu et al., 2021) have been developed which demonstrate excellent performance in EEG MI classification. More recent works indicate the growing popularity of various machine learning (ML) techniques, in particular DL, in EEG-based MI classification (Zheng et al., 2022; Yu et al., 2022; Zhao et al., 2022; Li et al., 2022).

However, DL-based algorithms contain large numbers of trainable model parameters which require a significant amount of training data and lead to an increase in computation time to train the classifier (Sakhavi and Guan, 2017; Kwon et al., 2019). In this regard, transfer learning (TL) based approach is an effective strategy utilizing pre-trained weights from different subject cases (Kwon et al., 2019; Jayaram et al., 2016; Fahimi et al., 2019; Zhang et al., 2021a,b; Sun et al., 2022; Mattioli et al., 2022; Cai et al., 2022; Khademi et al., 2022). In a recent development, the TL method has been employed DL based in the system, in which, an FBCSP based EEG representation utilizing knowledge distillation techniques have been used for MI-BCI classification with a fine-tuned CNN model (Sakhavi and Guan, 2017; Sakhavi et al., 2018). Additionally, a subject-independent deep CNN model has been developed using spectral-spatial input generation for MI-BCI system (Kwon et al., 2019). A hybrid deep neural network utilizing transfer learning has been applied for multi-class MI decoding for better performance (Zhang et al., 2021b). Furthermore, an adaptive transfer learning-based deep CNN has been employed for EEG MI classification which has demonstrated significant improvement in classification accuracy compared to the subject-dependent model (Zhang et al., 2021a).

Motivations: Although, the CNN-based models have achieved better results for EEG-based MI classification, there are various issues that have caused to hinder the performance of the classifier. Firstly, current CNN-based models only consider a single convolution scale for extracting features from MI EEG signal. Such a strategy may not be suitable to capture distinguishable features of various non-overlapping canonical frequency bands of EEG signals efficiently (Roy, 2022). Secondly, an important discriminative feature extraction accounting for ERD/ERS from MI EEG signal has often been ignored which limits the classifier's ability to learn important semantic features from the raw EEG data. Thirdly, only a handful of research has been geared towards establishing the efficient transfer learning framework to address the challenge of inter-subject variability between different subjects in the BCI system which requires fine-tuning the model during target MI subject classification (Wronkiewicz et al., 2015). Furthermore, feature extraction from multiscale CNN considering adaptation-based transfer learning is yet to be designed for full integration into end-to-end DL workflow which is the main bottleneck for the deployment of robust BCI applications with good classification accuracy. Therefore, the goal of the current work is to develop a robust deep learning framework accounting for inter-subject variability of different subjects and further improve the performance of the model by employing the subject-adaptive transfer learning strategy to achieve better accuracy with fewer training samples.

Motivated by the aforementioned challenges and shortcomings, the current work proposes a transfer learning-based multi-scale feature fused CNN (MSFFCNN) for EEG-based multi-class MI classification. The major contributions and findings of the present research work can be summarized as follows.

- The current work designs an efficient multiscale CNN (MSCNN) architecture to capture the semantic features of EEG signals from multiple convolutional scales for four distinguishable frequency bands α, β, δ, and θ from raw EEG signal to enhance the performance of MI classifier.
- The present study integrates an FBCSP with the one-vs-rest (OVR) CNN block (OVR-FBCSP CNN) for extracting the discriminative spatiotemporal CSP features of event-related desynchronization and synchronization (ERD/ERS) suitable for multiclass MI classification tasks.
- In the current work, 4 different model variants of MSCNN including subject-specific, subject-independent, and subject-adaptive classification models considering two different adaptation configurations have been proposed to exploit the full learning capacity of the classifier.

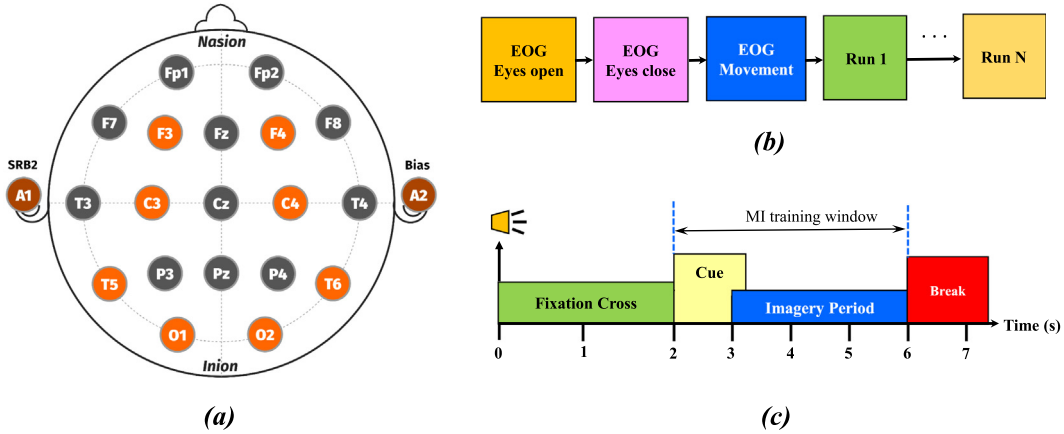


Fig. 2. (a) Schematics of EEG electrodes positioning in international 10–20 standardized electrode system; (b) timing scheme of each session; (c) timing scheme of each trial with 4 s of MI activity.

- The performance of two different subject-adaptive models has been extensively studied for a vast range of learning rates and degrees of adaption to explore the adaptation capability of the proposed model accounting for inter-subject variability from different subjects.
- Current study illustrates the advantages of using an adaptive transfer learning-based model over the subject-specific and subject-independent models in terms of the overall performance of the classifier by achieving the best average classification accuracy outperforming several state-of-the-art models with considerably fewer training samples.

The proposed framework requires less training data and computation time suitable for designing efficient and robust real-time human–robot interactions. The current study effectively addresses the shortcoming of existing CNN-based EEG-MI classification models and significantly improves the classification accuracy. The paper is organized as follows: Section 2 describes the dataset and proposed EEG data representation; OVR-FBCSP has been described in Section 3; Section 4 introduces the proposed MSFFCNN framework; Section 5 describes the transfer learning methods; Sections 6 and 7 deal with the relevant finding and discussion of the proposed classifier. Finally, the conclusions and prospects of the current work have been discussed in Section 8.

2. Dataset and EEG input data representation

2.1. Dataset

For the current study, the performance of the MSFFNN model has been evaluated on the BCI competition IV-2a dataset which contains EEG data from 9 subjects for 4 different MI classes (i.e., left hand, right hand, feet, and tongue) (Tangermann et al., 2012). The recorded EEG data comprises two sessions from each subject with the first and second session consisting of training and test data, respectively. Each session was recorded for 288 trials (72 trials per class) from 22 EEG channels with a sampling frequency of 250 Hz according to a standardized international 10–20 electrode system as shown in Fig. 2-(a). In a single trial, there was a cue followed by 4 s of MI activity from each subject for each of the four classes as illustrated in Fig. 2-(b, c).

2.2. Proposed EEG data representation

Since DL methods in MI-BCI systems requires relatively large numbers of EEG training data to achieve good classification accuracy, current work adopts the representation of 1D EEG segment as input data (Liu et al., 2021). Generally, in MI-BCI systems, inputted data are considered as a 2D array combining spatial information of muti-channels (electrodes) and corresponding time-series data for each trial. Such representation ignores the positional electrode distribution in the actual acquisition device and may lead to complex multi-channel correlations instead of simple adjacent relationships (Schirrmeister et al., 2017; Liu et al., 2021). To circumvent such issues, in the present work, 1D EEG segment from each electrode of fix time window of 4 s has been considered for a particular subject to segment the signals related to MI task as illustrated in Fig. 3. In the proposed method, input data from each channel can be represented by 1D vector $\boldsymbol{S}_m := \left[S_1^m, S_2^m, \ldots, S_P^m\right] \in \mathbb{R}^{1\times P}$ where m is the number of channels; $P = 1000$ is the number of sampling points for a single MI task. Such simple and effective data representation results in a substantial increase in the number of training samples containing the dependence of the signals related to the MI activity. The proposed method can effectively illuminate unnecessary features such as the channel-related spatial information and correlations between electrodes (Schirrmeister et al., 2017; Liu et al., 2021).

3. One-vs-Rest Filter Bank Common Spatial Pattern

In the present work, FBCSP with the one-vs-rest (OVR) method has been utilized for extracting the spatiotemporal-frequency features of event-related desynchronization and synchronization (ERD/ERS) (Blankertz et al., 2008). The OVR-FBCSP network consists of several FBCSP blocks where the segment of EEG signal can be decomposed through a filter bank that contains an array of multiple types II Chebyshev sub-bandpass filters to extract the discriminative CSP features, for multi-class MI-BCI system (Zhang et al., 2019a, 2021b; Wu et al., 2006).

3.1. Spatial feature extraction

By employing OVR-CSP, spatial filtering has been performed by linearly transforming EEG signal to obtain feature information using

$$\boldsymbol{Z}_{b,t} = \boldsymbol{W}_b^T \boldsymbol{X}_{b,t} \tag{1}$$

where t is numbers of EEG sample per channel; T denotes transpose operation; $\boldsymbol{X}_{b,t} \in \mathbb{R}^{1\times t}$ represents the single-trial EEG segment from bth band-pass filter of tth trial; $\boldsymbol{Z}_{b,t} \in \mathbb{R}^{1\times t}$ denotes OVR-FBCSP features; $\boldsymbol{W}_b^T = \left[\boldsymbol{W}_{b,1}^T, \boldsymbol{W}_{b,2}^T, \boldsymbol{W}_{b,3}^T, \boldsymbol{W}_{b,4}^T\right]$ is the weight of OVR-FBCSP filter in which $\boldsymbol{W}_{b,j}^T$ $(j = 1, 2, 3, 4)$ represents CSP projection matrix. Following eigenvalue decomposition problem (Blankertz et al., 2008), the transformation matrix $\boldsymbol{W}_{b,j}^T$ can be obtained to yield optimal discriminating feature variances for multi-class MI as

$$\boldsymbol{C}_{b,j}\boldsymbol{W}_{b,j} = \left(\sum_{j=1}^{N} \boldsymbol{C}_{b,j}\right) \boldsymbol{W}_{b,j}\boldsymbol{E}_{b,j} \tag{2}$$

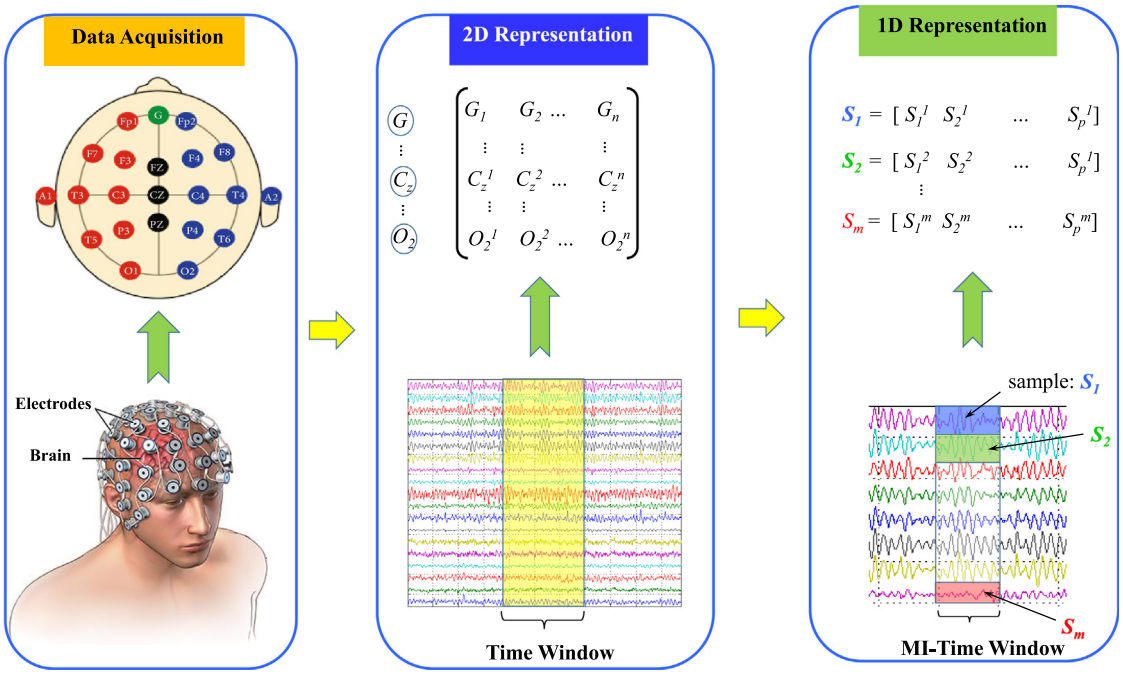


Fig. 3. Illustration of proposed EEG input data representation where acquired 2D data from the electrode distribution has been segmented into 1D vector during MI-activity time window.

where N is a number of MI classes; $\boldsymbol{C}_{b,j}$ is the covariance matrix of bth band-pass filtered EEG signal of respective jth MI class; $\boldsymbol{E}_{b,j}$ is the diagonal matrix containing eigenvalues of $\boldsymbol{C}_{b,j}$. Utilizing $\boldsymbol{W}_{b,j}$ from Eq. (2), spatial filtered signal $\boldsymbol{Z}_{b,t}$ maximizes the differences in the variance of MI classes. For each class, m pairs of CSP features of the tth time window for bth band-pass filtered EEG signal can be expressed as:

$$\boldsymbol{f}_{b,t} = \log \frac{\operatorname{diag}\left(\hat{\boldsymbol{W}}_b^{\,T} \boldsymbol{X}_{b,t} \boldsymbol{X}_{b,t}^{T} \hat{\boldsymbol{W}}_b\right)}{\operatorname{tr}\left(\hat{\boldsymbol{W}}_b^{\,T} \boldsymbol{X}_{b,t} \boldsymbol{X}_{b,t}^{T} \hat{\boldsymbol{W}}_b\right)} \tag{3}$$

where $\boldsymbol{f}_{b,t} \in \mathbb{R}^{2m}$ is the OVR-FBCSP output; $\hat{\boldsymbol{W}}_b$ denotes the first m and last m columns of $\boldsymbol{W}_{b,j}$ $(j = 1, 2, 3, 4)$; diag$(\cdot)$ is diagonal elements of the square matrix; tr$(\cdot)$ is the trace of the matrix. Note that $m = 2$ has been used for BCI IV 2a dataset. The OVR-FBCSP feature vector corresponds to tth time window can be expressed as

$$\boldsymbol{f}_t = \left[\boldsymbol{f}_{1,t}, \boldsymbol{f}_{2,t}, \ldots, \boldsymbol{f}_{k,t}\right] \tag{4}$$

where $\boldsymbol{f}_t \in \mathbb{R}^{1\times 2mk} (t = 1, 2, \ldots, n)$; n total numbers of time windows (i.e., trials); k is the total number of band-pass filters. Training data that comprises extracted feature $\bar{\boldsymbol{f}}_t \in \mathbb{R}^{n_t \times 2mk}$ and corresponding true class labels $\hat{\boldsymbol{f}}_t \in \mathbb{R}^{n_t \times 1}$ $(t = 1, 2, \ldots, n_t)$; n_t total numbers of trials in training data to make a distinction from evaluation data. Finally, the output matrix from OVR-FBCSP for each 1s time windows each has been attached to CNN layer called OVR-FBCSP CNN for spatial feature extraction (see Section 4.3).

3.2. Signal preprocessing

Each MI trial of 4 s long EEG signal has been segmented into 4 parts of 1 s time windows and fed into OVR-FBCSP to obtain spatial features as illustrated in Fig. 7. In OVR-FBCSP, the inputted EEG signal data have been passed through a filter bank that contains an array of a total of 12 bandpass filters including 2–6, 4–8, …, 24–28 Hz. Each filter with a bandwidth of 4 Hz and an overlap of 2 Hz has been employed covering the frequency range 2–32 Hz.

4. Proposed multi-scale feature fused CNN (MSFFCNN)

Achieving good accuracy in MI-BCI subject classification can be challenging due to the high variability and non-stationary characteristic of the raw EEG data. In this regard, CNN has demonstrated the effectiveness of extracting relevant features. Additionally, it is efficient in learning the hierarchical representations of complex high dimensional EEG data. CNN is a feed-forward artificial neural network generally consisting of alternating convolutional and sub-sampling layers to extract important local feature information between adjacent elements of the feature vector, and a fully connected (FC) layer at the end for final classification. Present work proposed an efficient end-to-end MSFFCNN framework in order to improve the accuracy and performance of multi-class MI classification tasks. The MSFFCNN model consists of multi-scale CNN (MSCNN), One-vs-Rest Filter Bank Common Spatial Pattern CNN (OVER-FBCSP CNN), Conv-pool layer, and fully connected layers as illustrated in Fig. 4. Each network block and the corresponding algorithm have been detailed in the subsequent sections.

4.1. Multi-scale CNN block (MSCNN)

In Inputted EEG data, there exist multiple non-overlapping canonical frequency bands where each band signify distinct behavioral states (Zhang et al., 2018; Donoghue et al., 2020). For MI-BCI systems, α (8–13 Hz) and β (13–30) bands are the most important (Malan and Sharma, 2019), since the increase/decrease of the power spectrum of these bands results in ERS/ERD, respectively (Shahid et al., 2010; Djemal et al., 2016). Whereas, low frequency δ- bands (2–4 Hz) carry important class-related information (Zhang et al., 2019b; Vuckovic and Sepulveda, 2008; Abiri et al., 2019). Additionally, θ-band (4–8 Hz) differs during the left/right-hand MI which is helpful during the MI-BCI classification process (Dai et al., 2020; Liu et al., 2019; Weber and Doppelmayr, 2016). Hence, these four non-overlapping bands (i.e., α, β, δ, and θ) have been considered for feature extraction by employing filter banks of 2–4 Hz, 4–8 Hz, 8–13 Hz, and 13–30 Hz in corresponding frequency bands. In the present study, a multi-scale CNN (MSCNN) with multi-scale convolution blocks (MCBs) has been proposed to extract

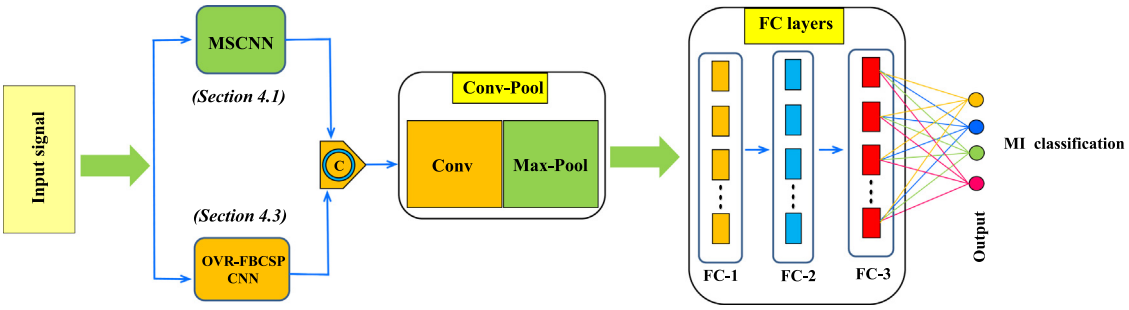


Fig. 4. The overall network architecture of the proposed end-to-end MSFFCNN consists of (a) multi-scale CNN (MSCNN), (b) One-vs-Rest Filter Bank Common Spatial Pattern CNN (OVER-FBCSP CNN) (c) conv-pool layer (d) fully connected (FC) layers for multiclass MI classification.

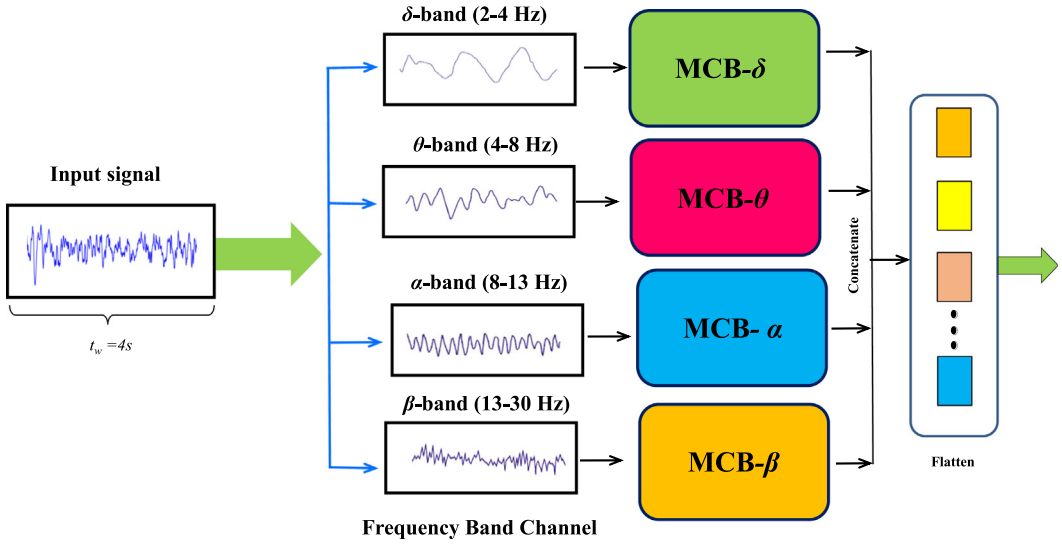


Fig. 5. The architecture of multi-scale CNN (MSCNN) network architecture consists of multi-scale convolution blocks (MCBs) as shown in Fig. 6.

feature information from each canonical frequency band efficiently as shown in Fig. 5.

Since our input EEG data representation is a 1D time-series signal, 1D-CNN has been employed which is relatively easier to train and offers minimal computational complexity compare to its 2D counterparts whilst achieving state-of-the-art performance (Kiranyaz et al., 2021). The convolution layer consists of 1D-convolution filters of a specified kernel stride which perform convolution operations sliding along the time axis of EEG signal to obtain feature maps and time-frequency information of the time series data (Farahat et al., 2019; Deng et al., 2021). In 1D-CNN, forward propagation can be expressed as follows:

$$x_k^l = b_k^l + \sum_{i=1}^{N_{l-1}} \text{conv1D}\left(w_{ik}^{l-1}, s_{l-1}^i\right) \tag{5}$$

where x_k^l is the input; b_k^l represents the bias of kth feature information in lth layer; w_{ik}^{l-1} is defined as the connecting weight between ith feature of the $l-1$th layer and kth feature of the lth layer; s_{l-1}^i represents output of the ith feature of $l-1$th layer; conv1D denotes convolution operation. The intermediate output y_k^l can be obtained passing x_k^l through the activation function $f(\bullet)$ as

$$y_k^l = f(x_k^l) \quad \in s_k^l = y_k^l \downarrow \Omega \tag{6}$$

Where s_k^l is the output of kth feature information in lth layer; Ω denotes the scalar factor of the down sampling operation $\downarrow$; rectified linear units (ReLU) $f(x) = \max(0, x)$ has been chosen as the activation function. During back-propagation, b_k^l and w_{ik}^{l-1} are updated when weight and bias sensitivities are determined by minimizing the error value E. The learning rate η can be defined as

$$w_{ik}^{l-1}(t+1) = w_{ik}^{l-1}(t) - \eta \frac{\partial E}{\partial w_{ik}^{l-1}}; \quad b_k^l(t+1) = b_k^l(t) - \eta \frac{\partial E}{\partial b_k^l} \tag{7}$$

The convolutional layer is followed by the pooling layer or down-sampling layer to reduce the spatial size of the representation and the number of network parameters while preserving important and relevant feature information. It aggregates neighboring values in the feature map by taking the average (average pooling) or maximum (maximum pooling) of the feature information. However, max pooling is more efficient in preserving the main features of the previous layer after down-sampling an input representation. Thus max-pooling has been adopted in the present study.

4.2. Multi-scale convolution block (MCB)

During the convolution operation, different kernel sizes can extract different spatial feature maps. For example, a relatively large kernel size captures the overall feature. However, it may miss the relevant and important fine-grain feature information. In such a case, a relatively small kernel size can effectively retain fine-grain information (Szegedy et al., 2017, 2015). Thus, in order to improve the performance of the MI classifier, a multi-scale convolution consisting of various kernel sizes can be an effective strategy to preserve both fine-grain high-frequency localized information as well as low-frequency overall representations for the various frequency bands of EEG signal (Tang et al., 2020; Roy, 2022).

Thus, a multi-scale convolution block (MCB) has been designed consists of three different kernel sizes in the convolution process. The

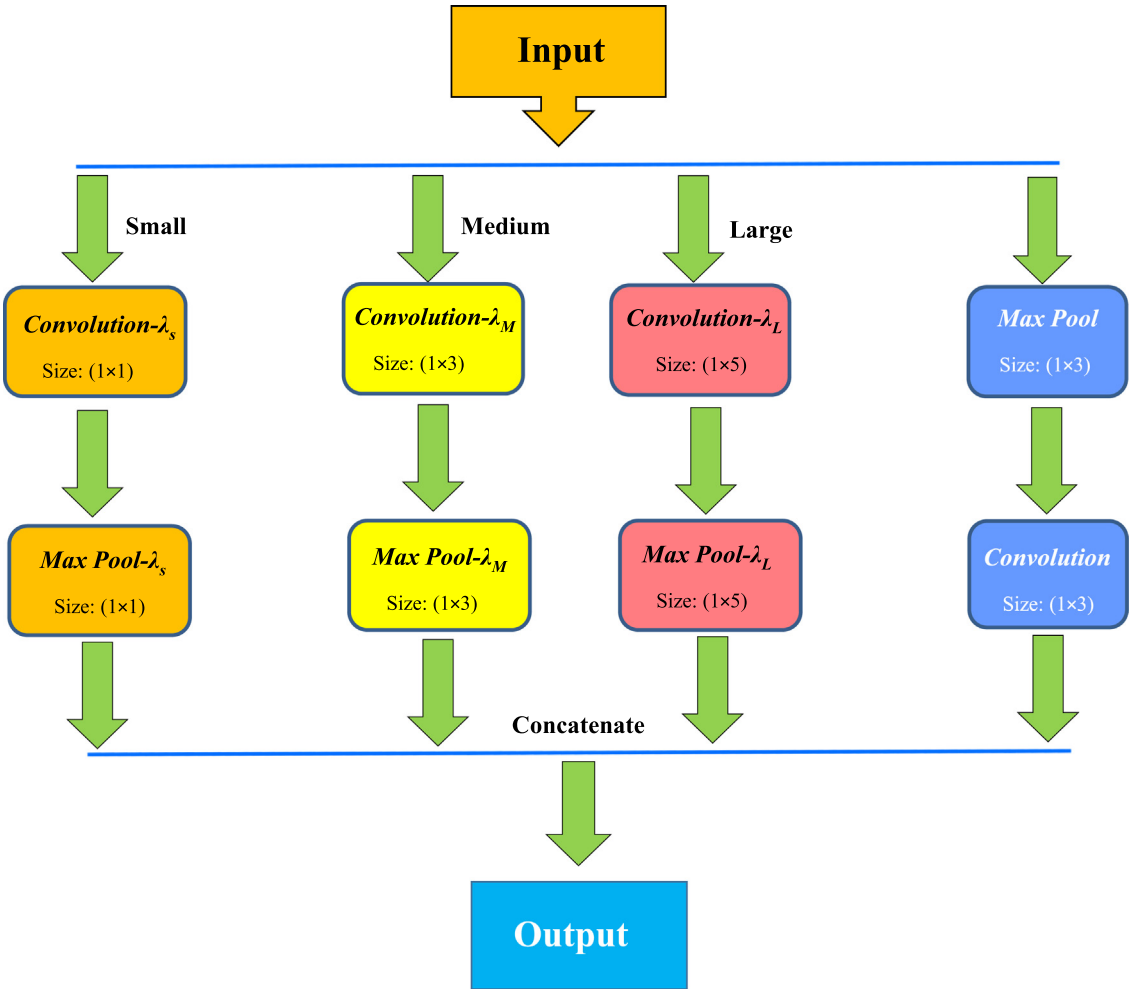


Fig. 6. Network architecture of proposed MCB comprise of large (λ_L), medium (λ_M), and small (λ_S) convolution scale for multi-scale feature extraction. See Table 1 for network parameters of MCB.

Table 1
Network parameters of MCB architecture as shown in Fig. 6.

Layer	Filter size	No. of Filters	Stride	Padding	Feature map size
Convolution layer-λ_S	1 × 1	2	1	SAME	(1 × 1000) × 2
Max-pooling layer-λ_S	1 × 1	–	4	SAME	(1 × 250) × 2
Convolution layer-λ_M	1 × 3	2	1	SAME	(1 × 1000) × 2
Max-pooling layer-λ_M	1 × 3	–	4	SAME	(1 × 250) × 2
Convolution layer-λ_M	1 × 5	2	1	SAME	(1 × 1000) × 2
Max-pooling layer-λ_M	1 × 5	–	4	SAME	(1 × 250) × 2
Max-pooling layer	1 × 3	–	4	SAME	(1 × 250)
Convolution layer	1 × 3	2	1	SAME	(1 × 250) × 2
Flatten layer (after concat)	–	–	–	–	1 × 2000

network architecture of MCB has been shown in Fig. 6. In MCB, convolution block λ_S represents a relatively small kernel size 1 × 1 which can effectively capture fine-grain localized information of the EEG signal. The convolution block λ_M with medium kernel size 1 × 3 can capture relatively coarse grain feature information. Whereas, λ_L represents large kernel size 1 × 5 which can collect the overall feature map efficiently. These three blocks are then followed by max-pooling layer to further reduce the network parameters. In addition, max-pooling and convolution layer of 1 × 3 has been utilized to preserve important features (Tang et al., 2020; Roy, 2022) as shown in Fig. 6. In MSCNN, EEG signal has been divided into four different frequency bands channels and passed through corresponding MCB blocks (i.e., MCB_i, $i = \delta, \theta, \alpha$, and β) as shown in Fig. 5. Finally, the multi-scale feature information has been obtained by concatenation operation. The network parameters of MCB architecture has been detailed in Table 1. The proposed MSCNN network can extract feature information from EEG signals on multiple scales which can significantly improves the classification accuracy of the MI-BCI classifier.

4.3. OVR-FBCSP CNN

The proposed model includes a CNN layer after OVR-FBCSP output for each 1 s time window called OVR-FBCSP CNN to extract spatial features as illustrated in Fig. 7. Each of the OVR-FBCSP CNN consists of two convolution-pooling layers. The feature output of size 12 × 12 for each time window from OVR-FBCSP has been passed through the 2D-convolution layer. The output of feature map x_k^l after 2D-convolution operation can be expressed as

$$x_k^l = b_k^l + \sum_{i \in M_j} \text{conv2D}\left(w_{ik}^{l-1}, s_{l-1}^i\right) \tag{8}$$

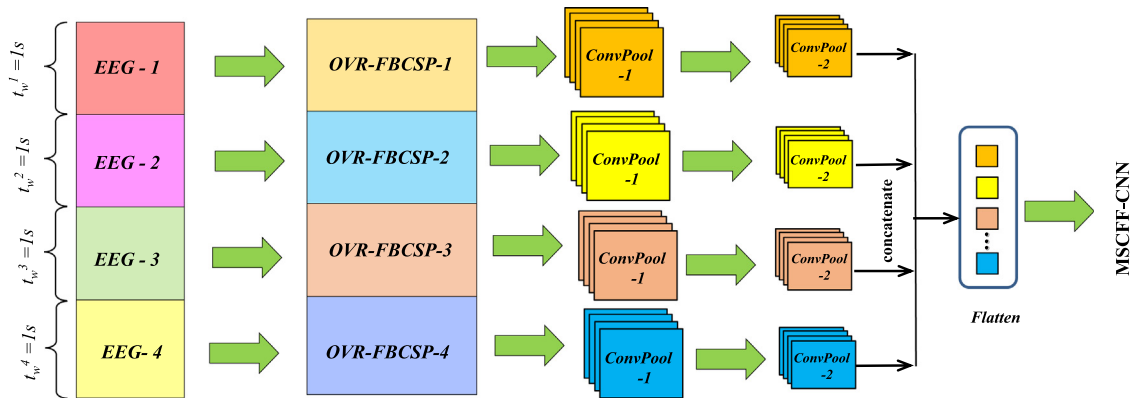


Fig. 7. Schematic of network structure for proposed OVR-FBCSP CNN block. See Table 2 for network parameters of OVR-FBCSP CNN.

Table 2
Network parameters of single branch OVR-FBCSP CNN architecture as shown in Fig. 7.

Layer	Filter size	No. of Filters	Stride	Padding	Feature map size
Convolution layer-1	3 × 3	12	1	SAME	(12 × 12) × 12
Max-pooling layer-1	3 × 3	–	2	SAME	(6 × 6) × 12
Convolution layer-2	2 × 2	12	1	SAME	(6 × 6) × 12
Max-pooling layer-2	2 × 2	–	2	SAME	(3 × 3) × 12
Flatten layer (after concat)	–	–	–	–	1 × 108

Table 3
Network parameters of proposed MSFFCNN model as shown in Fig. 4.

Block/Layer	Kernel size	No. of Kernal	Stride	Padding	Feature map size
MSCNN	–	–	–	–	4 × (1 × 2000)
OVR-FBCSP CNN	–	–	–	–	4 × (1 × 108)
Convolution layer	1 × 3	2	1	SAME	(1 × 8432) × 2
Max-pooling layer	1 × 3	–	4	SAME	(1 × 2108) × 2
FC-1	–	–	–	–	1024
Dropout layer	–	–	–	–	–
FC-2	–	–	–	–	512
FC-3	–	–	–	–	256
Output layer	–	–	–	–	4

where M_j represents input feature collection; conv2D denotes 2D-convolution operation. The activation function ReLU has been used. The convolution layer is followed by max-pooling layer in order to reduce the size of the feature map. The stride of the pooling kernel has been chosen as 2. Additionally, the zero-padding method has been employed to preserve the edge information and size of the spatial feature map. Finally, the output of the max-pooling layer has been resized through flatten layer which produces 1 × 432 array. The main network parameters of OVR-FBCSP CNN architecture have been listed in Table 2. The local features extracted by MSCNN and OVR-FBCSP have been concatenated together to form the global feature. It has been connected to convolution and then max-pooling layer as shown in Fig. 4.

4.4. FC layers

The extracted spatial features obtained from the max-pooling layer are passed through FC layers consisting of by 3-layer feedforward neural network as shown in Fig. 4. Finally, it is connected to the softmax layer containing four neurons to predict the final multi-classification result. The proposed MSFFCNN model has been fine-tuned to increase accuracy of the MI-BCI classifier by continuous experimental adjustment. The overall network parameters of proposed MSFFCNN model has been outlined in Table 3. See the algorithm of the MSFFCNN model in Appendix C.

4.5. Training optimization

The fully connected layer utilizes ReLU as the activation function in hidden layers which helps to accelerate the optimization process of the network. It has shown superior classification accuracy compared to other activation functions in the investigations on MI-BCI applications by Tang et al. (2020). The softmax function has been utilized to obtain exponential probability distribution of 4 different MI-BCI classification tasks in output layer which can be expressed as:

$$y_{p,m} = \frac{e^{y_m}}{\sum_m^T e^{y_m}} \tag{9}$$

where T represents the total number of classes; m represents the index of corresponding classes. Additionally, cross-entropy loss function has been utilized during training to optimize the model. The cross-entropy L_{CE} can be expressed as: $L_{CE}\left(p_i, t_i\right) = -\sum_{i=1}^{n} t_i \log(p_i)$ where n is total number of classes; t_i is the truth label; p_i is the softmax probability for ith class. In addition, Adam (Kingma and Ba, 2014) optimization scheme has been implemented to minimize the difference between probabilistic cross-entropy loss. Moreover, the dropout technique (Srivastava et al., 2014) has been employed to prevent over-fitting and accelerate the training procedure.

5. Transfer learning modeling

Generally, CNN based MI-BCI classification algorithms contain large numbers of trainable model parameters which requires a significant amount of training data and leads to an increase in computation time (Sakhavi and Guan, 2017; Kwon et al., 2019). In order to overcome the aforementioned issues, transfer learning can be an efficient strategy utilizing pre-trained weights from different subject cases. Sakhavi and Guan (2017), Kwon et al. (2019), Fahimi et al. (2019) and Zhang et al. (2021a,b) However, transfer learning can be challenging in the MI-BCI system due to substantial inter-subject variability between different subjects (Wronkiewicz et al., 2015; Zhang et al., 2021a). Therefore, adaptation schemes require special attention for fine-tuning the model parameter prudently to establish an efficient transfer learning-based MI-BCI classifier (Fahimi et al., 2019; Zhang et al., 2021a,b). In the present study, three different classification procedure including subject-specific, subject-independent (Kwon et al., 2019), and subject-adaptive

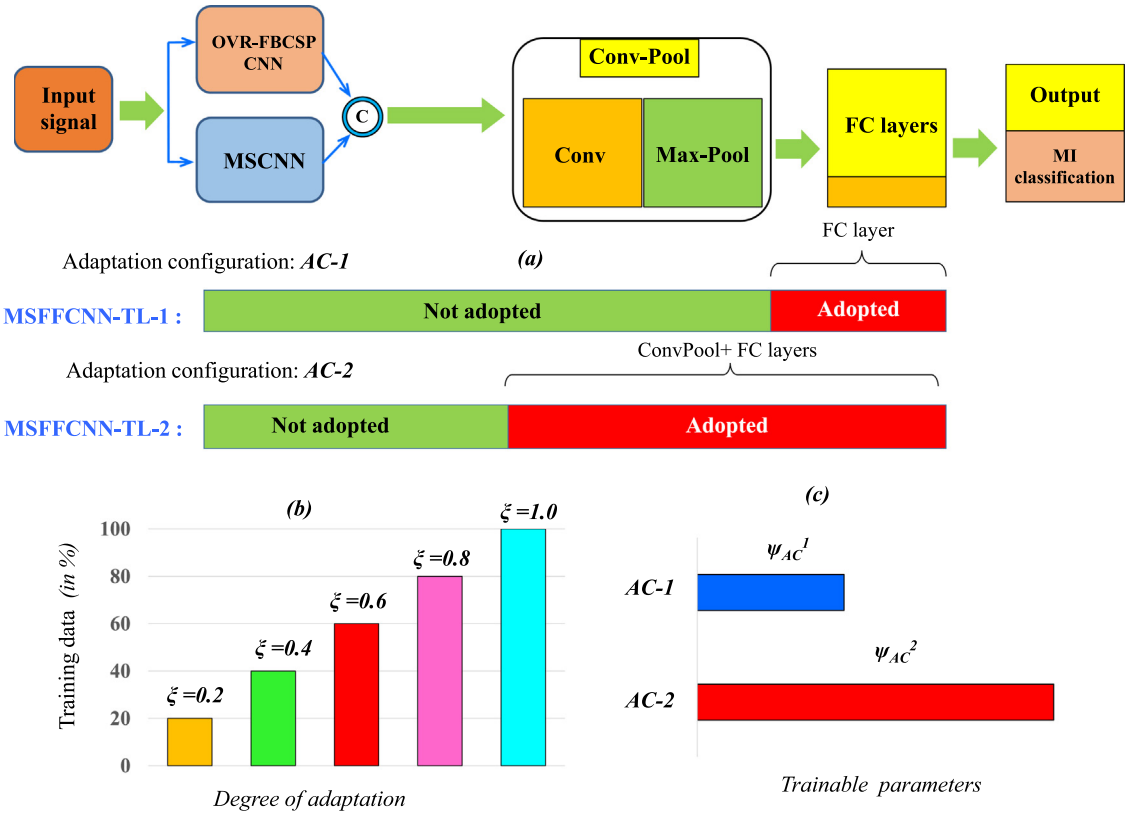


Fig. 8. Illustrations of (a) two different adaption configurations $AC-1$ and $AC-2$ for the proposed subject adaptive classification models MSFFCNN-TL-1 and MSFFCNN-TL-2, respectively; (b) different degree of adaption ξ; (c) numbers of required trainable network parameters Ψ_{AS} for adaption configurations.

classification (Zhang et al., 2021a) method has been employed which have been detailed in subsequent sections.

5.1. Subject-specific classification model (MSFFCNN-1)

In the present study, at first, the conventional subject-specific classification approach has been employed where the proposed model is trained on the target subject and the performance of the model has been evaluated on the same subject. The trained model based on subject-specific classification is the first baseline model for the present study and it has been named MSFFCNN-1.

5.2. Subject-independent classification model (MSFFCNN-2)

For the second approach, Subject-independent classification has been adopted where the model has been trained with all possible available training data except the target subject. During validation, LOSO cross-validation (Kwon et al., 2019) procedure has been utilized. The subject-independent classification-based model has been considered as the second baseline model. In the present work, it has been referred to as MSFFCNN-2.

5.3. Subject-adaptive classification model (MSFFCNN-TL)

Additionally, to further improve the performance of the subject-independent classification model, subject-adaptive classification approach (Zhang et al., 2021a) has been employed where the model has been fine-tuned for a particular subject in a pre-trained model (MSFFCNN-2) with a different fraction of target subject data (see Section 5.4 for datasets division) to study the influence of various degrees of adaptation ξ on the accuracy and performance of the classifier. In this study, the subject-adaptive classification-based model has been termed as MSFFCNN-TL. The proposed MSFFCNN-TL has two variants MSFFCNN-TL-1 and MSFFCNN-TL-2 based on two different adaptation configurations $AC-1$ and $AC-2$, respectively as shown in Fig. 8. In the first adaptation configuration $AC-1$, the fully connected (FC) layers have been adopted and optimized, whereas, the rest of the network parameters have been kept unchanged. In the second adaptation configuration $AC-2$, the last conv-pool block and fully connected layer have been retrained using the adaptation data while the rest of the network parameters have been kept unchanged. These two proposed adaptation configurations have been illustrated in Fig. 8-(a). The degree of adaptation ξ can be defined as the fraction of training data to fine-tune the model for each subject-adaptive configuration as shown in Fig. 8-(b). The numbers of required trainable network parameters Ψ_{AS} for two adaption configurations have been outlined in Fig. 8-(c). Additionally, learning rate η has been scaled down to avoid clobbering the initialization (Girshick et al., 2015). From the Eq. (7), one can assume $\eta = \theta\hat{\eta}$ where $0 < \theta \leq 1$ is the scaling factor of η:

$$b_k^l(t+1) = b_k^l(t) - \theta\hat{\eta}\frac{\partial E}{\partial b_k^l} = (1-\theta)b_k^l(t) + \theta\left(b_k^l - \hat{\eta}\frac{\partial E}{\partial b_k^l}\right) \tag{10}$$

For $\theta < 1$, the model accepts scaled-down weighted adaptation. In the present study, different scaling factors θ has been considered to obtain the optimal choice of the learning rate for efficient adaptation and enhance the performance of the classifier (see Section 6.3.1).

5.4. Dataset arrangement for transfer learning

In this section, the subdivision of EEG data during training and evaluation of the transfer learning models has been detailed. From the BCI competition IV-2a EEG dataset, the first session (identifier: T) is used to train the classifier. The second session (identifier: E) has been strictly utilized for evaluating the corresponding trained classifier. For the subject-specific classification model MSFFCNN-1, the first and

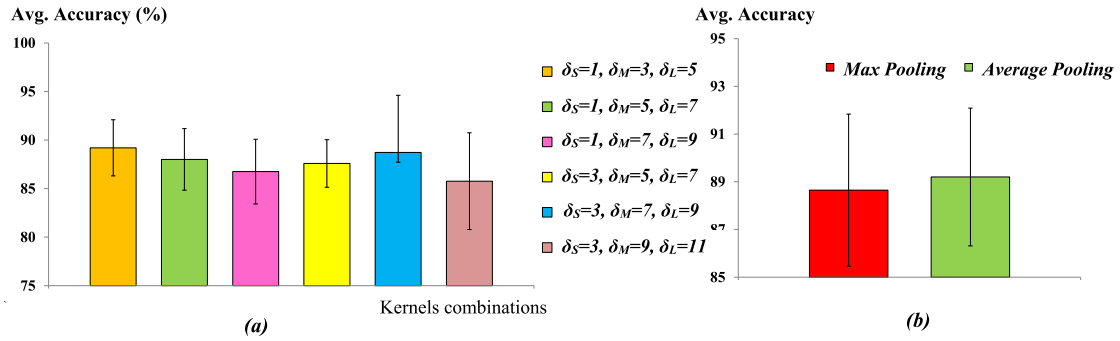


Fig. 9. Comparison of average accuracy with SE (in %) from MSFFCNN-1 for (a) various combinations of kernel sizes; (b) average pooling and max pooling in MCB.

Table 4
Average classification accuracy with SE (in %) from MSFFCNN-1 model for various combinations of kernel sizes in MCB. The bold indicates the best result from the corresponding model.

Kernel size			Avg. accuracy	SE
Small (δ_S)	Medium (δ_M)	Large (δ_L)		
1	3	5	**89.20**	±0.96
1	5	7	88.01	±1.05
1	7	9	86.75	±1.11
3	5	7	87.59	±0.82
3	7	9	88.72	±0.97
3	9	11	85.76	±0.98

Table 5
Average classification accuracy with SE (in %) obtained from MSFFCNN-1 for average pooling and max pooling.

Average accuracy with SE (%)	
Average pooling	Max pooling
88.65 ± 1.50	89.20 ± 0.96

second sessions have been considered for training and evaluation, respectively for a particular subject. In subject-independent classification, first session datasets from all subjects except the target subject have been utilized to train the MSFFCNN-2 model.

6. Results and discussions

In this section, the accuracy and performance of the subject-specific, subject-independent, and subject-adaptive classification models have been discussed and compared with several existing state-of-the-art methods. The performance of the proposed model has been evaluated by measuring classification accuracy $A_c = \sum_Q^{i=1} n_{ii}/N$ which can be obtained from the confusion matrix by the ratio of the sum of diagonal elements $\sum_Q^{i=1} n_{ii}$ to the total number of samples N. For a fair comparison, a similar training protocol has been applied for proposed MSFFCNN-1, MSFFCNN-2, and MSFFCNN-TL models. During training, a batch size of 24 has been considered which provides the highest classification accuracy and optimizes the convergence speed. The network has been trained for 350 epochs. In the second FC layer, a dropout probability value of 0.5 has been prescribed. The models have been implemented in Keras API with TensorFlow as the backend. The model has been trained and tested using an Intel Core i7 CPU with a single NVIDIA GeForce RTX 2080 8 GB GPU. The average A_c value and standard error (SE)= $s/\sqrt{n}$ (s is the sample standard deviation; n is the sample size) are calculated across all subjects.

6.1. Result for subject-specific classification

In this section, various network optimizations and corresponding average classification accuracy have been reported for the subject-specific classification MSFFCNN-1 model.

6.1.1. Effect of convolution kernel size

As mentioned earlier, different kernel sizes of the convolution layer can capture distinct features of the EEG signal. Thus, various combinations of convolution blocks (i.e, λ_S, λ_M, and λ_L in Fig. 6) in MCB are considered to study the effect of kernel sizes on the performance and accuracy of the proposed model as reported in Table 4. With increasing kernel size, in particular, for combinations ($\delta_S = 3$, $\delta_M = 7$, $\delta_L = 9$) and ($\delta_S = 3$, $\delta_M = 9$, $\delta_L = 11$) average classification accuracy values have been achieved as 88.72% and 85.76%, respectively as shown in Fig. 9-(a). For combinations ($\delta_S = 1$, $\delta_M = 7$, $\delta_L = 9$) and ($\delta_S = 3$, $\delta_M = 5$, $\delta_L = 7$) classification accuracy drops down to 86.75% and 87.59%, respectively. However, for the combination of relatively small kernel sizes, classification accuracy improves. After considering all combinations, the best classification result has been obtained corresponds to $\delta_S = 1$, $\delta_M = 3$, $\delta_L = 5$ with an average classification accuracy of 89.20%. Thus, in the present study, combinations of $\Omega_L = 5$, $\Omega_M = 3$, and $\Omega_S = 1$ has been selected. Such combination is found to be computationally most efficient with 1.19% accuracy gain compared to $\Omega_S = 1$, $\Omega_M = 5$, and $\Omega_L = 7$ (see Table 9). Moreover, such combination achieves SE value of 0.96% which is 0.09% lower than $\Omega_S = 1$, $\Omega_M = 5$, and $\Omega_L = 7$. Thus, $\delta_S = 1$, $\delta_M = 3$, $\delta_L = 5$ in MCB provides the best performance in-terms of both classification accuracy and speed from MSFFCNN-1 model.

6.1.2. Influence of pooling type

From the previous section, it has been concluded that the combination of $\delta_S = 1$, $\delta_M = 3$, $\delta_L = 5$ provides the best accuracy. Thus, corresponding pooling window size of $\delta_S^p = 1$, $\delta_M^p = 3$, $\delta_L^p = 5$ has been prescribed in MCB. Two different types of pooling including max pooling and average pooling have been considered to explore the effect of pooling type on the average classification of MSFFCNN-1 as listed in Table 5. From the comparison, one can see that max-pooling provides the best classification accuracy with 0.55% improvement over average-pooling as depicted in Fig. 9-(b). Moreover, max-pooling provides the smallest SE value of 0.96% which is 0.54% lower than average pooling. Thus, max-pooling has been adopted in all four models which provides the better capability of highlighting features conducive to the MI-BCI classification task.

6.2. Result for subject-independent classification

In this section, the classification accuracy for each subject and average classification accuracy across all 9 subjects has been reported for the subject-independent classification model MSFFCNN-2 and then compared with MSFFCNN-1 and other baseline models.

Table 6
Classification accuracy for each subject and average accuracy with SE (in %) from LDA, SVM, standard CNN, subject-specific model MSFFCNN-1, and subject-independent model MSFFCNN-2. The bold indicates the best result from the corresponding model.

Subject	Classification accuracy				
	LDA	SVM	CNN	MSFFCNN-1	MSFFCNN-2
A01	69.16	65.38	71.39	**93.13**	91.57
A02	60.56	59.38	74.16	84.62	**85.67**
A03	61.16	63.49	75.25	91.62	**93.09**
A04	81.17	80.19	83.03	87.89	**91.57**
A05	76.18	74.69	79.23	87.21	**92.56**
A06	64.59	73.16	73.14	85.92	**89.78**
A07	68.13	68.65	71.67	92.03	**94.67**
A08	72.29	75.19	74.64	93.21	**96.67**
A09	68.77	78.15	77.46	87.24	**88.78**
Avg. accuracy with SE	69.12 ± 1.66	70.92 ± 1.16	75.55 ± 1.98	89.20 ± 0.96	**91.61** ± 0.88

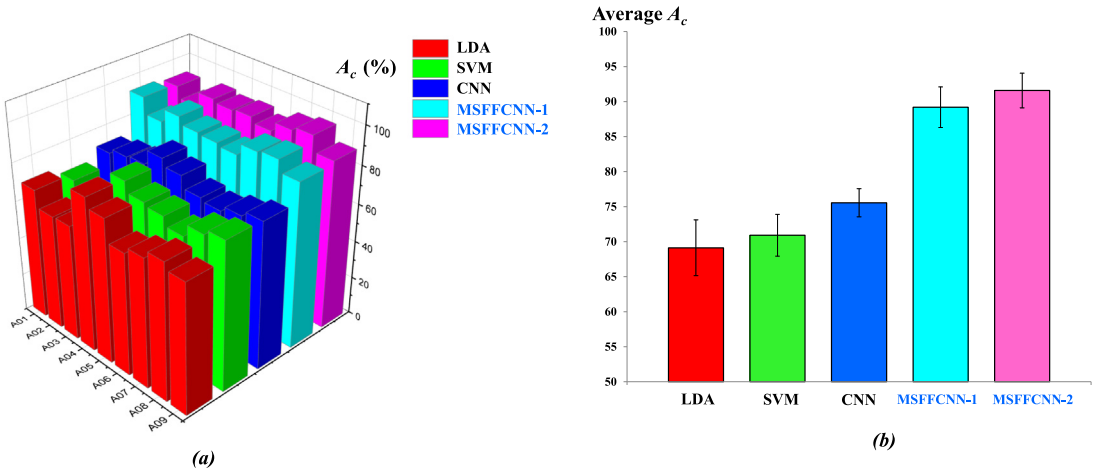


Fig. 10. Comparison bar chart of (a) A_c for each subject; (b) average A_c with SE (in %) between LDA, SVM, CNN, and proposed MSFFCNN models.

6.2.1. Comparison with baseline models

In order to evaluate the performances of the proposed MSFFCNN variants, the classification accuracy has been compared with commonly used traditional ML models as MI-BCI classifiers including SVM and LDA as the baseline models. In addition, a standard CNN has been used as a baseline DL model. The standard CNN network structure consists of five conv-pool layers which are relatively deeper than the proposed MSFFCNN models. For fair comparison, model hyper-parameters were kept as consistent as possible with the proposed models. As outlined in Table 6, both MSFFCNN-1 and MSFFCNN-2 have achieved better accuracy for all subjects with an average accuracy of 89.20% ± 0.96% and 91.61% ± 0.88% compared to baseline models. Compared to LDA and SVM, MSFFCNN-1 has gained 20.43% and 11.05% improvement in average A_c, respectively. As shown in Fig. 10 -(b), both MSFFCNN-1 and MSFFCNN-2 have demonstrated superior performance with accuracy improvement of 13.65% and 14.15% over the standard CNN model. Additionally, the MSFFCNN models have relatively low SE value indicating better capability and generalization in representing EEG data from different subjects compared to other baseline models. The comparison illustrates the advantages of using the DL model for strong feature extraction capability among different subjects in EEG-based MI-BCI systems compared to traditional ML methods.

6.2.2. Comparison of subject-dependent and subject-independent models

Comparing the A_c values, MSFFCNN-2 has achieved better average A_c with 2.41% improvement over MSFFCNN-1 as illustrated in Fig. 10-(b). For subject A01, MSFFCNN-1 has demonstrated better performance with A_c of 93.13% compared to 91.57% achieved by MSFFCNN-2. However, MSFFCNN-2 has demonstrated better accuracy for all other subject classes compared to MSFFCNN-1, in particular there are 5.35%, 3.86%, and 3.46% improvement in A05, A06, and A08, respectively. The comparison indicates that relatively large training data can significantly increase the performance of MSFFCNN-2. Furthermore, MSFFCNN-2 has achieved the lowest SE value of 0.70% indicating better representation and generalization of the classifier than MSFFCNN-1.

6.3. Result for subject-adaptive classification

This section reports the performance of subject-adaptive classification model MSFFCNN-TL for two different adaptation configurations $AC-1$, $AC-2$. Various parametric study considering different learning rate scale factor θ and degree of adaptation ξ on the accuracy of the MSFFCNN-TL classifier has been explored in the subsequent subsections.

6.3.1. Influence of learning rate scale factor θ

As mentioned earlier, different θ can have significant impact on the performance of the MSFFCNN-TL model, thus in the current work six different $\theta = 1, 0.5, 0.1, 0.05, 0.025$, and 0.01 have been studied considering five different degree of adaption $\xi = 0.2, 0.4, 0.6, 0.8$, and 1.0 for each θ as outlined in Tables 7–8 where bold highlights the best adaptation configuration for a particular θ. For $\theta = 1$, the learning rate of the MSFFCNN-TL model is the same as the other two models. In such a case, the best average A_c has been obtained as 91.79% for the full degree of adopted $AC-2$ configuration as shown in Table 7. However, partial degree of adaptation $\xi = 0.8$ in $AC-2$ provides the best $A_c = 92.36\%$ for $\theta = 0.5$. With further reduction of $\theta = 0.1$ there is 0.88% improvement in accuracy compared to $\theta = 0.5$ for $\xi = 1.0$. Similarly, for $\theta = 0.05$ and 0.025, accuracy has been improved to 93.69% and 93.84% for full adaptation $\xi = 1.0$, respectively as shown in Table 8. It is noteworthy to mention, the best accuracy results for

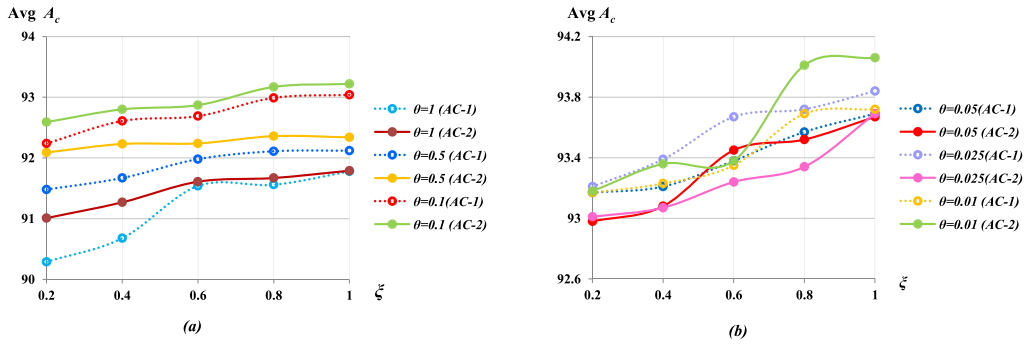


Fig. 11. Influence of θ on the average A_c (in %) of the MSFFCNN-TL model for (a) $\theta = 1, 0.5, 0.1$; (b) $\theta = 0.05, 0.025, 0.01$ in the range $0.2 \leqslant \xi \leqslant 1.0$ considering both $AC-1$ and $AC-2$ configurations.

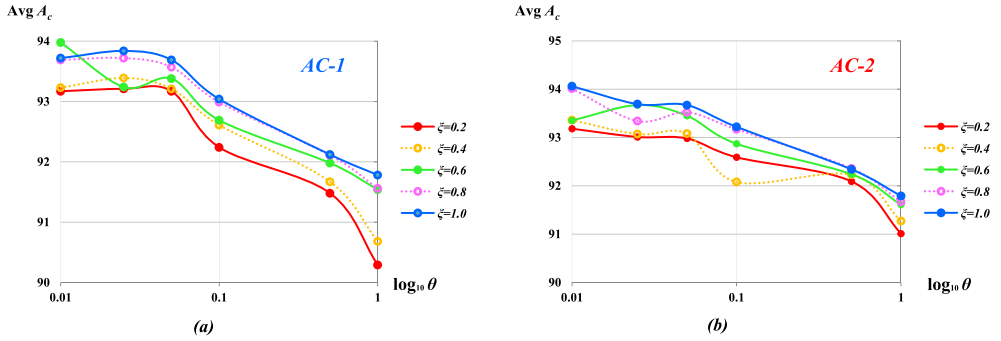


Fig. 12. Influence of degree of adaption ξ on the classification accuracy A_c (in %) of the MSFFCNN-TL model for (a) $AC-1$; (b) $AC-2$ configurations in the range $0.01 \leqslant \theta \leqslant 1.0$.

Table 7
Average A_c obtained from MSFFCNN-TL model for $\theta = 1, 0.5$, and 0.1 considering different ξ for $AC-1$ and $AC-2$ configurations. The bold highlights the best adaptation configuration for a particular θ.

Scale factor	Configuration	$\xi = 0.2$	$\xi = 0.4$	$\xi = 0.6$	$\xi = 0.8$	$\xi = 1.0$
$\theta = 1$	$AC-1$	90.29	90.68	91.54	91.56	91.78
	$AC-2$	91.01	91.27	91.61	91.67	**91.79**
$\theta = 0.5$	$AC-1$	91.48	91.67	91.98	92.11	92.12
	$AC-2$	92.09	92.23	92.24	**92.36**	92.34
$\theta = 0.1$	$AC-1$	92.24	92.61	92.69	92.99	93.04
	$AC-2$	92.59	92.80	92.87	93.17	**93.22**

Table 8
Average A_c obtained from MSFFCNN-TL model for $\theta = 0.05, 0.025$, and 0.01 considering different ξ for $AC-1$ and $AC-2$ configurations. The bold highlights the best adaptation configuration for a particular θ.

Scale factor	Configuration	$\xi = 0.2$	$\xi = 0.4$	$\xi = 0.6$	$\xi = 0.8$	$\xi = 1.0$
$\theta = 0.05$	$AC-1$	93.17	93.21	93.38	93.57	**93.69**
	$AC-2$	92.98	93.08	93.45	93.52	93.67
$\theta = 0.025$	$AC-1$	93.21	93.39	93.67	93.72	**93.84**
	$AC-2$	93.01	93.07	93.24	93.34	93.69
$\theta = 0.01$	$AC-1$	93.17	93.23	93.35	93.69	93.72
	$AC-2$	93.18	93.36	93.98	94.01	**94.06**

$\theta = 0.05$ and 0.025 have been achieved for $AC-2$. However, the overall best average $A_c = 94.06\%$ has been obtained for the lowest $\theta = 0.01$ with 2.27% increase compared to $\theta = 1.0$ indicating importance of tuning down θ to improve the performance of the classifier significantly.

In Fig. 11, average A_c has been plotted as function of ξ for different θ. For $\theta \leq 0.1$, A_c increases with decreasing θ for particular ξ as shown in Fig. 11-(a) with $AC-2$ provides best result. The result indicates that $AC-2$ is suitable for better adaption for relatively high θ. However, for $\theta > 0.1$, $AC-1$ can be efficient, in particular, $0.6 \leqslant \xi \leqslant 1.0$ as illustrated in Fig. 11-(b) indicating better adaption for lower learning rate even for fewer number of adaptation network parameters. Overall, with higher degree of adaption, both configurations tend to perform better.

Table 9
Comparison of average A_c, E_f, and t_c between MSFFCNN-1, MSFFCNN-2, MSFFCNN-TL-1 ($AC-1$), and MSFFCNN-TL-2 ($AC-2$) for some specific θ and ξ. The bold denotes the best result from the corresponding model.

Model	Configuration	θ	ξ	Avg A_c	E_f	t_c (in s)
MSFFCNN-1	–	1	–	89.20	4.34	9.78
MSFFCNN-2	–	1	–	91.61	3.87	138
MSFFCNN-TL-1	$AC-1$	0.5	0.6	91.98	3.43	7.83
MSFFCNN-TL-2	$AC-2$	0.1	0.4	92.80	3.01	6.76
MSFFCNN-TL-1	$AC-1$	0.025	0.8	93.72	2.67	**8.78**
MSFFCNN-TL-2	$AC-2$	0.01	1	**94.06**	**1.43**	9.55

6.3.2. *Influence of degree of adaptation ξ*

In this section, the influence of ξ on AC of the MSFFCNN-TL classifier has been explored in the range $0.01 \leqslant \theta \leqslant 1.0$ as shown in Fig. 12. For $AC-1$ configuration, increasing ξ for a particular θ increases average A_c at least in the range $0.025 \leqslant \theta \leqslant 1.0$ as depicted in Fig. 11-(a). For $\xi = 0.6$, the accuracy the classifier improves for $\theta < 0.05$. Whereas, for $AC-2$, the effect of relatively low $\xi = 0.4$ on the performance of the classifier is more prominent in relatively high $\theta > 0.1$ as shown in Fig. 11-(b). This is because the $AC-2$ configuration has a higher trainable parameter than $AC-1$ which may lead to over-fitting to some degree for higher ξ. The present study reveals that there exist two different regimes of θ for two different adaptation configurations for optimum performance of the MSFFCNN-TL classifier.

6.3.3. *Overall performance comparison of four models*

This section compares the overall performance of MSFFCNN-1, MSFFCNN-2, MSFFCNN-TL-1 ($AC-1$), and MSFFCNN-TL-2 ($AC-2$) by

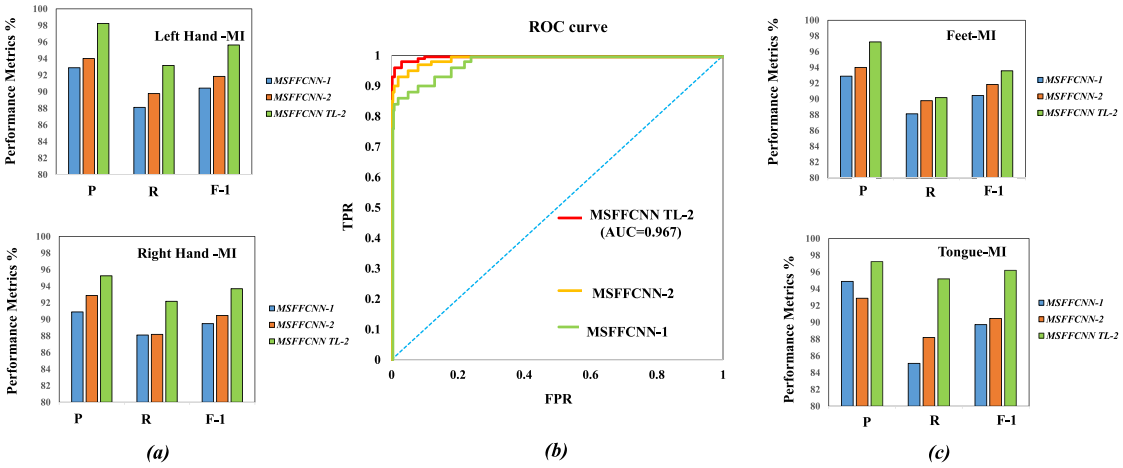


Fig. 13. (a, c) Comparison bar chart of performance metrics: P, R, $F1$ (in %) for the four different MI (left hand, right hand, feet, and tongue); (b) ROC curves and corresponding AUC values obtained from proposed MSFFCNN-1, MSFFCNN-2, and MSFFCNN-TL 2 models, respectively.

evaluating average A_c, final loss E_f, and computation time for training t_c as listed in Table 9. The comparison demonstrates the superiority of subject-adaptive model by achieving the best accuracy compared to subject-dependent and subject-independent counterpart even for $\xi < 1$. It implies effectiveness of the proposed MSFFCNN-TL model by using less training data. MSFFCNN-TL-2 shows better performance than MSFFCNN-TL-1 for a larger number of training samples, in particular, $\xi > 0.6$ since MSFFCNN-TL-2 has a larger number of network parameters to adopt. Overall, MSFFCNN-TL-2 has achieved the highest average $A_c = 94.06\%$ which is a 4.86% and 2.45% improvement over MSFFCNN-1 and MSFFCNN-2, respectively. Comparing E_f, MSFFCNN-1 reaches the maximum value of 4.34, whilst, MSFFCNN-TL for $\theta = 0.01$ attains the lowest value of 1.43 indicating better learning capability between all four model variants. Additionally, a comparison of t_c reveals that MSFFCNN-1 and MSFFCNN-TL variants take a similar time frame (within 10 s) to train, whereas, MSFFCNN-1 takes a slightly higher training time. With the increase of ξ, training time increases. For same ξ and θ, configuration $AC-1$ is faster compared to $AC-2$ due to lower number of trainable parameters. Whereas, MSFFCNN-2 take significantly longer time to train with average t_c of 138 s (or 2.3 min). The comparison demonstrates that the subject-adaptive model has better accuracy, learning capability, and adaptability for different subject classes with significantly less training data and training time.

Furthermore, the performance of MSFFCNN-1, MSFFCNN-2, and MSFFCNN-TL 2 models on the recognition of four different MI classes (left hand, right hand, feet, and tongue) have been compared by measuring performance metrics: precision (P), recall (R), and $F1$ score as shown in Fig. 13-(a, c). For multi-class classification, sample data can be classified into four different categories: true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN), based on the true class and the model-predicted class. The evaluation matrices P, R, $F1$ can be defined as

$$P = \frac{TP}{(TP+FP)}; \qquad R = \frac{TP}{(TP+FN)} \tag{11}$$

$$F1 = \frac{2PR}{(P+R)}. \tag{12}$$

The larger values of P, R, F1 indicate better performance of the model. The MSFFCNN-TL-2 model has achieved the best P, R, $F1$ values of 98.25%, 93.18%, and 95.65% for left hand; 95.23%, 92.23%, and 93.69% for right hand; 97.69%, 89.77%, and 93.53% for feet; 97.31%, 95.18%, and 96.20% tongue MI, respectively. Finally, ROC (receiver operating characteristic) curve has been plotted from the true positive rate (TPR) or R (in the ordinate) and false positive rate (FPR) data (in the abscissa) from MSFFCNN-1, MSFFCNN-2, and MSFFCNN-TL 2 which have been shown in Fig. 13-(b). The area under the ROC curve is expressed in AUC and ranges from 0.5 to 1. The closer the AUC is to 1.0, the higher the performance of the model. As shown in Fig. 13-(b), MSFFCNN-1, MSFFCNN-2, and MSFFCNN-TL-2 have achieved AUC values of 0.917, 0.925, and 0.967, respectively indicating the best performance of MSFFCNN-TL-2 compared to other variants.

6.4. Comparison Cohen's kappa-coefficient

To evaluate the performance of proposed MI-BCI models, average Cohen's kappa-coefficient (κ) has been calculated as Dornhege et al. (2007):

$$\kappa = 1 - \frac{1-P_o}{1-P_e} = \frac{P_o - P_e}{1-P_e} \tag{13}$$

where P_e can be measured by summing the product of predicted numbers for each category and ground truth for each category divided by the square of the total number of samples; P_o denotes the total classification accuracy. For multi-class classification problem, κ value reflects the generalization of the model and consistency of classifying different subjects. In Table 10, κ values obtained from the MSFFCNN-2 and MSFFCNN-TL models have been compared with the state-of-art ML models such as FBCSP (Ang et al., 2012), channel-wise CNN (CW-CNN) (Sakhavi et al., 2018), 3D CNN (Zhao et al., 2019), channel-wise convolution with channel mixing (C2CM) (Sakhavi et al., 2018), sequential updating semi-supervised spectral regression kernel discriminant analysis (SUSS-SRKDA) (Nicolas-Alonso et al., 2015a), deep restricted Boltzmann machine network with t-distributed stochastic neighbor embedding (RBMt-SNE) (Xu et al., 2020a), stacked regularized linear discriminant analysis (SRLDA) (Nicolas-Alonso et al., 2015b), hybrid deep neural network (HDNN) (Zhang et al., 2021b), CNN-LSTM (Zhang et al., 2019a), EEG topographical representative CNN (ETRCNN) (Xu et al., 2020b), and HDNN with transfer learning (HDNN-TL) (Zhang et al., 2021b) for BCI Competition IV 2a datasets. From the overall comparison, it can be seen that the proposed models outshine other ML models significantly in terms of κ values as shown in Fig. 14. Comparing average κ, MSFFCNN-TL has achieved the highest κ value of 0.88 which is 10.00% and 8.00% improvement over the state-of-the-art HDNN and ETRCNN models, respectively.

6.5. Comparison with different state-of-the-art models

This section reports the accuracy comparison between proposed models and several state-of-the-art models such as support vector machine (SVM) (Sakhavi et al., 2018), CW-CNN (Sakhavi et al., 2018), transductive learning with co-variate shift-detection (TLCSD) (Raza et al., 2016), C2CM (Sakhavi et al., 2018), adaptive learning with CSD

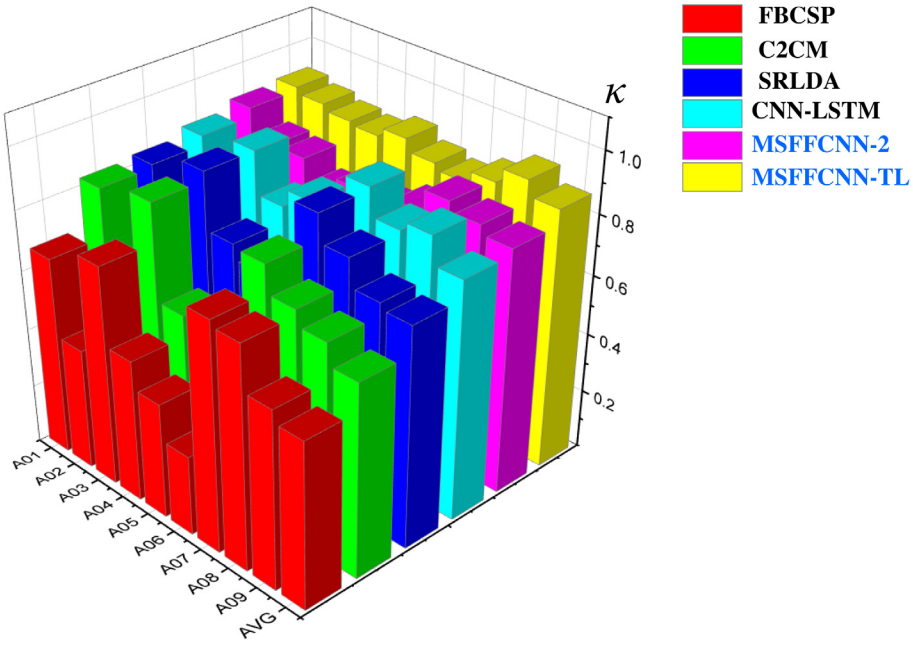


Fig. 14. Comparison of κ across all subjects between state-of-the-art ML models and proposed MSFFCNN models.

Table 10
Comparison of κ with several state-of-the-art ML models across all subjects. The bold denotes the best result from the corresponding model.

Method	Subject									Avg
	A01	A02	A03	A04	A05	A06	A07	A08	A09	
FBCSP	0.67	0.41	0.74	0.48	0.39	0.27	0.77	0.75	0.60	0.56
CW-CNN	0.81	0.47	0.82	0.56	0.50	0.26	0.87	0.75	0.69	0.64
3D CNN	0.69	0.45	0.78	0.59	0.64	0.53	0.65	0.70	0.71	0.64
C2CM	0.83	0.53	0.87	0.55	0.50	0.27	0.86	0.77	0.72	0.65
SUSS-SRKDA	0.83	0.51	0.88	0.68	0.56	0.35	0.90	0.84	0.75	0.70
RBN t−SNE	0.82	0.48	0.76	0.66	0.50	0.53	0.78	0.86	0.89	0.70
SRLDA	0.84	0.55	**0.90**	0.71	0.66	0.44	**0.94**	0.85	0.76	0.74
HDNN	0.82	0.61	0.85	0.64	0.78	0.73	0.84	0.85	0.87	0.78
CNN-LSTM	0.87	0.59	**0.90**	0.76	0.82	0.66	0.95	0.86	0.89	0.80
ETRCNN	0.84	0.66	0.87	0.76	0.57	**0.89**	0.80	**0.89**	0.90	0.80
HDNN-TL	**0.92**	0.63	0.86	0.67	0.81	0.75	0.86	0.87	0.91	0.81
MSFFCNN-2	0.91	0.82	0.81	0.74	0.79	0.77	0.81	0.88	0.85	0.83
MSFFCNN-TL	0.91	**0.89**	0.87	**0.86**	**0.89**	0.85	0.83	0.87	**0.92**	**0.88**

(ALCSD) (Raza et al., 2016), RBN t−SNE (Xu et al., 2020a), densely feature fusion convolutional neural networks (DFFN) (Li et al., 2019), subject specific multivariate empirical mode decomposition based filtering (SS-MEMDBF) (Gaur et al., 2018), ETRCNN (Xu et al., 2020b), Riemannian geometry-based FBCSP (RG-FBCSP) (Majidov and Whangbo, 2019), and hybrid scale CNN (HS-CNN) (Dai et al., 2020) as outlined in Table 11. It can be seen that proposed MSFFCNN-2 and MSFFCNN-TL models attain the highest average A_c of 91.61% and 94.06% among all state-of-the-art models for BCI competition IV-2a dataset. It is noteworthy to mention that MSFFCNN-TL improves the average classification accuracy of 14.12%, 8.51%, and 2.49% over the recent and advanced DL models SS-MEMDBF, RG-FBCSP, and HS-CNN, respectively as shown in Fig. 15. More specifically, MSFFCNN-TL demonstrated significant classification accuracy improvement of 7.70%, 3.67%, 3.03%, and 3.58% in subjects $A02$, $A03$, $A05$, and $A09$, respectively over MSFFCNN-2 which indicates the efficiency and robustness of the proposed subject-adaptive MSFFCNN-TL model.

7. Discussion

The current research demonstrates the efficiency of the proposed MSFFCNN for extracting semantic features of EEG signals from multiple convolutional kernel scales for four different frequency bands to enhance the performance of the MI classifier. Incorporating OVR-FBCSP CNN further improves the accuracy of the classifier emphasizing the prospect of the current framework in adopting the distinguishable feature of MI-EEG signals. Furthermore, the current study illustrates the advantages of using an adaptive transfer learning-based CNN model over the subject-specific and subject-independent model in order to obtain better classification accuracy and performance. Due to inter-subject variability in MI-EEG data, the majority of the traditional ML methods in the MI-BCI system use subject-specific data for better accuracy. However, to train a deep CNN network model containing a large number of trainable parameters requires a significant amount of training data that is not suitable for relatively less available subject-specific data which generally limits the accuracy of the classifier. In this regard, the current study shows the usefulness of subject-independent models which can be trained in a large number of inter-subject samples. In order to further increase the performance of the classifier, different adaptation techniques considering different learning rates, scale factors, degree of adaption, and adaptation configurations have been explored to fine-tune the subject-independent model which provides significant improvement in the accuracy of the classifier. The current study illustrates the importance of lowering learning rates which facilitate effective adaptation and improve classification accuracy. Among all adaptive models, it has been found that for $\theta = 0.01$ with a full degree of adaptation $\xi = 1$ for adaptation configuration $AC - 2$ provides the best average accuracy of 94.06% which is a 4.86% improvement over the subject-specific model. In short, current work effectively addresses training and adaptation strategy in adaptive cross-subject transfer learning considering inter-subject variability for better performance.

Future directions: Future work can be focused on further improving the classification accuracy by incorporating long short-term memory (LSTM) RNN architecture to extract temporal features and employ the proposed framework for classifying spatio-temporal multi-class MI subject classification for various BCI applications. Moreover, additional adaptation schemes in the MSCB depending on the importance of different adaptations of the source distribution to the target domain can be employed to further optimize the model. Another direction could be the implementation of unsupervised or semi-supervised learning (Jia et al., 2014) to circumvent expensive and time-consuming manual labeling in unsupervised learning to perform classification tasks in abundant class

Table 11
A_c for different subjects and average A_c (in %) obtained from existing state-of-the-art MI-BCI classification models. The bold indicates the best result from the corresponding model.

Method	Subject									Avg
	A01	A02	A03	A04	A05	A06	A07	A08	A09	
SVM	82.29	60.42	82.99	72.57	60.07	44.10	86.11	77.08	75.00	71.18
CW-CNN	86.11	60.76	86.81	67.36	62.50	45.14	90.63	81.25	77.08	73.07
ALCSD	90.28	54.17	93.75	64.58	57.64	65.28	62.50	90.97	85.42	73.84
C2CM	87.50	65.28	90.28	66.67	62.50	45.49	89.58	83.33	79.51	74.46
TLCSD	90.28	57.64	95.14	65.97	61.11	65.28	61.11	91.67	86.11	74.92
RBN t−SNE	86.61	61.26	87.27	75.20	64.55	65.91	83.78	89.91	92.08	78.51
DFFN	85.40	69.30	90.29	71.07	65.41	69.45	88.18	86.46	93.54	79.90
SS-MEMDBF	91.49	60.56	94.16	76.72	58.52	68.52	78.67	97.01	**93.85**	79.94
ETRCNN	85.88	75.41	91.32	83.45	72.11	**91.72**	85.71	91.32	93.23	85.57
RG-FBCSP	93.30	84.59	91.68	84.55	86.54	76.92	94.03	93.20	92.24	85.56
HS-CNN	90.07	80.28	**97.08**	89.66	**97.04**	87.04	92.14	**98.51**	92.31	91.57
MSFFCNN-2	91.57	85.67	93.09	91.57	92.56	89.78	**94.67**	96.79	88.78	91.61
MSFFCNN-TL	**93.64**	**93.37**	96.76	**92.17**	95.59	91.07	94.56	97.14	92.36	**94.06**

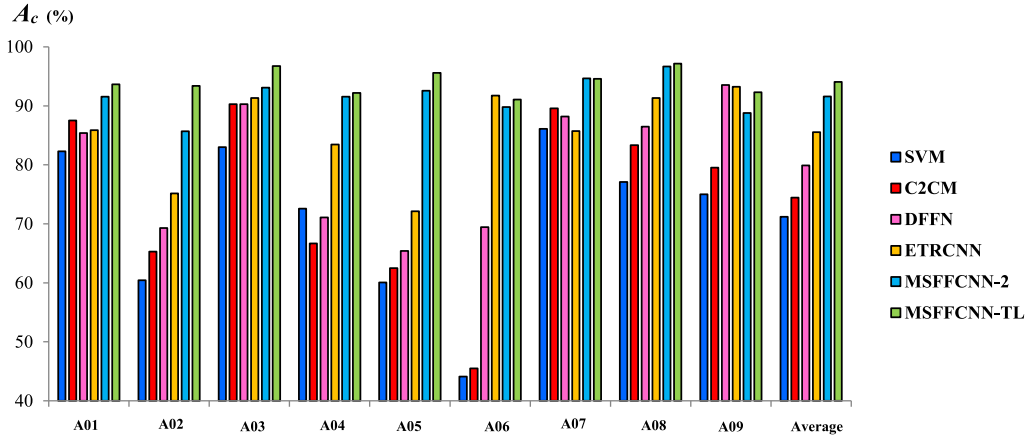


Fig. 15. Comparison bar-chart of A_c for different subjects and average A_c (in %) between proposed MSFFCNN-2, MSFFCNN-TL and existing state-of-the-art MI-BCI models.

labels for a wide range of MI-BCI scenarios. Furthermore, Maximum Mean Discrepancy (MMD) strategy (Rozantsev et al., 2018) can be utilized to further regularize the adaptation of individual CNN layer. Nevertheless, the proposed MSFFCNN model can be employed as a more reliable and robust MI-based real-time BCI applications such as robotic control (Gupta et al., 2020; Liao et al., 2012; Shi et al., 2021), rehabilitation of neuromotor disorders (Zhang et al., 2015), text entry speech communication (Hossain et al., 2019; Makin et al., 2020), and various other applications (Roy et al., 2022; Jamil et al., 2022; Roy and Bhaduri, 2021, 2022; Arozi et al., 2020; Glowacz et al., 2021; Glowacz, 2021; Roy, 2020, 2021).

8. Conclusion

In summarizing, in the current study, a transfer learning-based multi-scale featured fused CNN (MSFFCNN) framework has been presented for multi-class MI classification where the multi-scale convolution block comprises various convolutional kernel sizes that can efficiently extract semantic features for different frequency bands δ, θ, α, and β in multiple scales. Various parametric exploration including the influence of learning rate, scale factor, and degree of adaptions on different adaptation configurations sheds light on effective and optimal adaptation strategies to maximize the performance of the proposed transfer learning model. The current study illustrates the importance of lowering the learning rate which facilitates effective adaptation and improves classification accuracy. Among all adaptive models, it has been found that for relatively low learning rate with a full degree of adaptation and adaptation configuration corresponds to a relatively larger number of adaptable network parameters provides the best average classification accuracy of 94.06% (±0.70%) which is a 4.86% improvement over the subject-specific model. The proposed framework requires considerably less training data and computation time suitable for designing robust and efficient human–robot interactions.

CRediT authorship contribution statement

Arunabha M. Roy: Conceptualization, Investigation, Validation, Methodology, Formal analysis, Writing – original draft, Writing – review and editing.

Declaration of competing interest

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

Data availability

The data that support the findings of this study are available upon reasonable request.

Appendix A. Abbreviations

See Table 12.

Table 12
Abbreviations.

Abbreviations	Words/phrase
AC	Adaptive Configuration
BCI	Brain–Computer Interface
CNN	Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
EEG	Electroencephalography
ERD	Event-Related De-synchronization
ERS	Event-Related Synchronization
FC	Fully Connected
FPR	False Positive Rate
FBCSP	Filter Bank Common Spatial
LSTM	Long Short-Term Memory
MI	Motor Imagery
MCB	Multi-scale Convolutional Block
MSCNN	Multi-scale CNN
MSFFCNN	Multi-scale feature-fused CNN
MSFFCNN-1	Subject-specific classification model
MSFFCNN-2	Subject-independent classification model
MSFFCNN-TL	Subject-adaptive classification model
OVR-FBCSP	One-versus rest filter bank common spatial pattern
ReLU	Rectified Linear Unit
ROC	Receiver Operating Characteristic Curve
SE	Standard Error
SGD	Stochastic Gradient Descendent
SNR	Signal to Noise Ratio
TL	Transfer Learning
TPR	True Positive Rate

Table 13
List of variables.

Variables	Definitions
A_c	Classification accuracy
S_m	Input data from each channels
m	Number of channels
n	Number of classes
P	Number of sampling points
t	Number of EEG sample/channel
t_w	Time window
E	Cost functional
η	Learning rate
θ	Scaling factor of η
ξ	Degree of adaptation
$\lambda_S, \lambda_M, \lambda_L$	Small, medium, large kernel sizes
$\delta_S, \delta_M, \delta_L$	Small, medium, large convolution blocks
Ψ_{AS}	Numbers of required trainable network parameters for AC.
E_f	Final loss
t_c	Computation time for training
κ	Cohen's kappa-coefficient
$T0i^t$	Training dataset for ith subject
$E0i^t$	Evaluation dataset for ith subject
$\bigcup \Theta$	Trained classifier of MSFFCNN-2
p_i	Softmax probability
L_{CE}	Cross-entropy loss
$Z_{b,t}$	OVR-FBCSP features
$W^T_{b,j}$	CSP projection matrix
$C_{b,j}$	Covariance matrix
$f_{b,t}$	OVR-FBCSP output

Appendix B. Variables

See Table 13.

Appendix C. Algorithm

See Algorithm 1.

Algorithm 1: MSFFCNN model Algorithm

Input: $EEG \quad signal \quad (t_w = 4s)$

Concatenate $\leftarrow \begin{cases} 1.\ \text{MSCNN (size: 1 x 8000).} \\ 2.\ \text{OVR- FBCSP CNN (size: 1x 432)} \end{cases}$

MSCFFCNN-1: Training: $\rightarrow T0i^t; \quad (A_c)^i \rightarrow E0i^e \quad (i = 1, ..., 9)$

MSCFFCNN-2: Training: $\rightarrow \sum T0j^t; \quad (A_c)^i \rightarrow E0i^e$
$(i \neq j \quad \forall \quad i, j = 1, ..., 9)$

MSCFFCNN-TL: Training: $\rightarrow \exists \quad \bigcup \Theta \quad \forall j = (1, ..., 9) \in \xi T0j^t;$
$0 \leq \xi \leq 1;$

$(A_c)^i \rightarrow E0i^e \quad (i \neq j \quad \forall \quad i, j = 1, ..., 9)$

References

Abiri, R., Borhani, S., Sellers, E.W., Jiang, Y., Zhao, X., 2019. A comprehensive review of EEG-based brain–computer interface paradigms. J. Neural Eng. 16 (1), 011001.

Ang, K.K., Chin, Z.Y., Wang, C., Guan, C., Zhang, H., 2012. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b. Front. Neurosci. 6, 39.

Ang, K.K., Chin, Z.Y., Zhang, H., Guan, C., 2008. Filter bank common spatial pattern (FBCSP) in brain-computer interface. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). IEEE, pp. 2390–2397.

Arozi, M., Caesarendra, W., Ariyanto, M., Munadi, M., Setiawan, J.D., Glowacz, A., 2020. Pattern recognition of single-channel sEMG signal using PCA and ANN method to classify nine hand movements. Symmetry 12 (4), 541.

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Muller, K., 2008. Signal processing magazine. IEEE 25 (1), 41.

Cai, Y., She, Q., Ji, J., Ma, Y., Zhang, J., Zhang, Y., 2022. Motor imagery EEG decoding using manifold embedded transfer learning. J. Neurosci. Methods 109489.

Chu, Y., Zhao, X., Zou, Y., Xu, W., Han, J., Zhao, Y., 2018. A decoding scheme for incomplete motor imagery EEG with deep belief network. Front. Neurosci. 12, 680.

Dai, G., Zhou, J., Huang, J., Wang, N., 2020. HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification. J. Neural Eng. 17 (1), 016025.

Deng, X., Zhang, B., Yu, N., Liu, K., Sun, K., 2021. Advanced TSGL-EEGNet for motor imagery EEG-based brain-computer interfaces. IEEE Access 9, 25118–25130.

Djemal, R., Bazyed, A.G., Belwafi, K., Gannouni, S., Kaaniche, W., 2016. Three-class EEG-based motor imagery classification using phase-space reconstruction technique. Brain Sci. 6 (3), 36.

Donoghue, T., Haller, M., Peterson, E.J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A.H., Wallis, J.D., Knight, R.T., et al., 2020. Parameterizing neural power spectra into periodic and aperiodic components. Nature Neurosci. 23 (12), 1655–1665.

Dornhege, G., Millán, J.d.R., Hinterberger, T., McFarland, D.J., Muller, K.-R., et al., 2007. Toward Brain-Computer Interfacing, Vol. 63. Citeseer.

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

Fahimi, F., Zhang, Z., Goh, W.B., Lee, T.-S., Ang, K.K., Guan, C., 2019. Inter-subject transfer learning with an end-to-end deep convolutional neural network for EEG-based BCI. J. Neural Eng. 16 (2), 026007.

Farahat, A., Reichert, C., Sweeney-Reed, C.M., Hinrichs, H., 2019. Convolutional neural networks for decoding of covert attention focus and saliency maps for EEG feature visualization. J. Neural Eng. 16 (6), 066010.

Gaur, P., Pachori, R.B., Wang, H., Prasad, G., 2018. A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry. Expert Syst. Appl. 95, 201–211.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2015. Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 38 (1), 142–158.

Glowacz, A., 2021. Thermographic fault diagnosis of ventilation in BLDC motors. Sensors 21 (21), 7245.

Glowacz, A., Tadeusiewicz, R., Legutko, S., Caesarendra, W., Irfan, M., Liu, H., Brumercik, F., Gutten, M., Sulowicz, M., Daviu, J.A.A., et al., 2021. Fault diagnosis of angle grinders and electric impact drills using acoustic signals. Appl. Acoust. 179, 108070.

Gupta, A., Agrawal, R., Kirar, J.S., Kaur, B., Ding, W., Lin, C.-T., Andreu-Perez, J., Prasad, M., 2020. A hierarchical meta-model for multi-class mental task based brain-computer interfaces. Neurocomputing 389, 207–217.

Hossain, M.S., Amin, S.U., Alsulaiman, M., Muhammad, G., 2019. Applying deep learning for epilepsy seizure detection and brain mapping visualization. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) 15 (1s), 1–17.

Jamil, S., Abbas, M.S., Roy, A.M., 2022. Distinguishing malicious drones using vision transformer. AI 3 (2), 260–273.

Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., Grosse-Wentrup, M., 2016. Transfer learning in brain-computer interfaces. IEEE Comput. Intell. Mag. 11 (1), 20–31.

Jia, X., Li, K., Li, X., Zhang, A., 2014. A novel semi-supervised deep learning framework for affective state recognition on EEG signals. In: 2014 Ieee International Conference on Bioinformatics and Bioengineering. IEEE, pp. 30–37.

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

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.J., 2021. 1D convolutional neural networks and applications: A survey. Mech. Syst. Signal Process. 151, 107398.

Kwon, O.-Y., Lee, M.-H., Guan, C., Lee, S.-W., 2019. Subject-independent brain–computer interfaces based on deep convolutional neural networks. IEEE Trans. Neural Netw. Learn. Syst. 31 (10), 3839–3852.

Li, H., Ding, M., Zhang, R., Xiu, C., 2022. Motor imagery EEG classification algorithm based on CNN-LSTM feature fusion network. Biomed. Signal Process. Control 72, 103342.

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

Liao, L.-D., Chen, C.-Y., Wang, I.-J., Chen, S.-F., Li, S.-Y., Chen, B.-W., Chang, J.-Y., Lin, C.-T., 2012. Gaming control using a wearable and wireless EEG-based brain-computer interface device with novel dry foam-based sensors. J. Neuroeng. Rehabil. 9 (1), 1–12.

Liu, Y.-H., Lin, L.-F., Chou, C.-W., Chang, Y., Hsiao, Y.-T., Hsu, W.-C., 2019. Analysis of electroencephalography event-related desynchronisation and synchronisation induced by lower-limb stepping motor imagery. J. Med. Biol. Eng. 39 (1), 54–69.

Liu, J., Ye, F., Xiong, H., 2021. Multi-class motor imagery EEG classification method with high accuracy and low individual differences based on hybrid neural network. J. Neural Eng. 18 (4), 0460f1.

Majidov, I., Whangbo, T., 2019. Efficient classification of motor imagery electroencephalography signals using deep learning methods. Sensors 19 (7), 1736.

Makin, J.G., Moses, D.A., Chang, E.F., 2020. Machine translation of cortical activity to text with an encoder–decoder framework. Nature Neurosci. 23 (4), 575–582.

Malan, N.S., Sharma, S., 2019. Feature selection using regularized neighbourhood component analysis to enhance the classification performance of motor imagery signals. Comput. Biol. Med. 107, 118–126.

Mattioli, F., Porcaro, C., Baldassarre, G., 2022. A 1D CNN for high accuracy classification and transfer learning in motor imagery EEG-based brain-computer interface. J. Neural Eng. 18 (6), 066053.

Netzer, E., Frid, A., Feldman, D., 2020. Real-time EEG classification via coresets for BCI applications. Eng. Appl. Artif. Intell. 89, 103455.

Nicolas-Alonso, L.F., Corralejo, R., Gomez-Pilar, J., Álvarez, D., Hornero, R., 2015a. Adaptive semi-supervised classification to reduce intersession non-stationarity in multiclass motor imagery-based brain–computer interfaces. Neurocomputing 159, 186–196.

Nicolas-Alonso, L.F., Corralejo, R., Gomez-Pilar, J., Alvarez, D., Hornero, R., 2015b. Adaptive stacked generalization for multiclass motor imagery-based brain computer interfaces. IEEE Trans. Neural Syst. Rehabil. Eng. 23 (4), 702–712.

Nour, M., Öztürk, Ş., Polat, K., 2021. A novel classification framework using multiple bandwidth method with optimized CNN for brain–computer interfaces with EEG-fNIRS signals. Neural Comput. Appl. 1–15.

Parija, S., Bisoi, R., Dash, P., Sahani, M., 2021. Deep long short term memory based minimum variance kernel random vector functional link network for epileptic EEG signal classification. Eng. Appl. Artif. Intell. 105, 104426.

Pfurtscheller, G., Brunner, C., Schlögl, A., Da Silva, F.L., 2006. Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. NeuroImage 31 (1), 153–159.

Rashid, M., Sulaiman, N., PP Abdul Majeed, A., Musa, R.M., Bari, B.S., Khatun, S., et al., 2020. Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review. Front. Neurorobot. 14, 25.

Raza, H., Cecotti, H., Li, Y., Prasad, G., 2016. Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface. Soft Comput. 20 (8), 3085–3096.

Roy, A., 2020. Evolution of martensitic nanostructure in NiAl alloys: tip splitting and bending. Mater. Sci. Res. India 17, 3–6.

Roy, A.M., 2021. Finite element framework for efficient design of three dimensional multicomponent composite helicopter rotor blade system. Eng 2 (1), 69–79.

Roy, A.M., 2022. An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces. Biomed. Signal Process. Control 74, 103496.

Roy, A.M., Bhaduri, J., 2021. A deep learning enabled multi-class plant disease detection model based on computer vision. AI 2 (3), 413–428.

Roy, A.M., Bhaduri, J., 2022. Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4. Comput. Electron. Agric. 193, 106694.

Roy, A.M., Bose, R., Bhaduri, J., 2022. A fast accurate fine-grain object detection model based on YOLOv4 deep neural network. Neural Comput. Appl. 1–27.

Rozantsev, A., Salzmann, M., Fua, P., 2018. Beyond sharing weights for deep domain adaptation. IEEE Trans. Pattern Anal. Mach. Intell. 41 (4), 801–814.

Saa, J.F.D., Çetin, M., 2012. A latent discriminative model-based approach for classification of imaginary motor tasks from EEG data. J. Neural Eng. 9 (2), 026020.

Sakhavi, S., Guan, C., 2017. Convolutional neural network-based transfer learning and knowledge distillation using multi-subject data in motor imagery BCI. In: 2017 8th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE, pp. 588–591.

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

Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T., 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapp. 38 (11), 5391–5420.

Shahid, S., Sinha, R.K., Prasad, G., 2010. Mu and beta rhythm modulations in motor imagery related post-stroke EEG: a study under BCI framework for post-stroke rehabilitation. Bmc Neurosci. 11, 1.

Shi, B., Wang, Q., Yin, S., Yue, Z., Huai, Y., Wang, J., 2021. A binary harmony search algorithm as channel selection method for motor imagery-based BCI. Neurocomputing 443, 12–25.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Sun, Y., Lo, F.P.-W., Lo, B., 2019. EEG-based user identification system using 1D-convolutional long short-term memory neural networks. Expert Syst. Appl. 125, 259–267.

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 Netw. 151, 111–120.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-First AAAI Conference on Artificial Intelligence.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9.

Tabar, Y.R., Halici, U., 2016. A novel deep learning approach for classification of EEG motor imagery signals. J. Neural Eng. 14 (1), 016003.

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

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., Leeb, R., Mehring, C., Miller, K.J., Mueller-Putz, G., et al., 2012. Review of the BCI competition IV. Front. Neurosci. 6, 55.

Vuckovic, A., Sepulveda, F., 2008. Delta band contribution in cue based single trial classification of real and imaginary wrist movements. Med. Biol. Eng. Comput. 46 (6), 529–539.

Weber, E., Doppelmayr, M., 2016. Kinesthetic motor imagery training modulates frontal midline theta during imagination of a dart throw. Int. J. Psychophysiol. 110, 137–145.

Wronkiewicz, M., Larson, E., Lee, A.K., 2015. Leveraging anatomical information to improve transfer learning in brain–computer interfaces. J. Neural Eng. 12 (4), 046027.

Wu, W., Gao, X., Gao, S., 2006. One-versus-the-rest (OVR) algorithm: An extension of common spatial patterns (CSP) algorithm to multi-class case. In: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. IEEE, pp. 2387–2390.

Xing, J., Qiu, S., Ma, X., Wu, C., Li, J., Wang, S., He, H., 2020. A CNN-based comparing network for the detection of steady-state visual evoked potential responses. Neurocomputing 403, 452–461.

Xu, M., Yao, J., Zhang, Z., Li, R., Yang, B., Li, C., Li, J., Zhang, J., 2020b. Learning EEG topographical representation for classification via convolutional neural network. Pattern Recognit. 105, 107390.

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

Yu, Z., Chen, W., Zhang, T., 2022. Motor imagery EEG classification algorithm based on improved lightweight feature fusion network. Biomed. Signal Process. Control 75, 103618.

Yu, T., Xiao, J., Wang, F., Zhang, R., Gu, Z., Cichocki, A., Li, Y., 2015. Enhanced motor imagery training using a hybrid BCI with feedback. IEEE Trans. Biomed. Eng. 62 (7), 1706–1717.

Zhang, R., Li, Y., Yan, Y., Zhang, H., Wu, S., Yu, T., Gu, Z., 2015. Control of a wheelchair in an indoor environment based on a brain–computer interface and automated navigation. IEEE Trans. Neural Syst. Rehabil. Eng. 24 (1), 128–139.

Zhang, K., Robinson, N., Lee, S.-W., Guan, C., 2021a. Adaptive transfer learning for EEG motor imagery classification with deep Convolutional Neural Network. Neural Netw. 136, 1–10.

Zhang, X., Yao, L., Sheng, Q.Z., Kanhere, S.S., Gu, T., Zhang, D., 2018. Converting your thoughts to texts: Enabling brain typing via deep feature learning of EEG signals. In: 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom). IEEE, pp. 1–10.

Zhang, X., Yao, L., Wang, X., Monaghan, J., Mcalpine, D., Zhang, Y., 2019b. A survey on deep learning based brain computer interface: Recent advances and new frontiers. p. 66, arXiv preprint arXiv:1905.04149.

Zhang, R., Zong, Q., Dou, L., Zhao, X., 2019a. A novel hybrid deep learning scheme for four-class motor imagery classification. J. Neural Eng. 16 (6), 066004.

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

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

Zhao, X., Zhang, H., Zhu, G., You, F., Kuang, S., Sun, L., 2019. A multi-branch 3D convolutional neural network for EEG-based motor imagery classification. IEEE Trans. Neural Syst. Rehabil. Eng. 27 (10), 2164–2177.

Zheng, L., Feng, W., Ma, Y., Lian, P., Xiao, Y., Yi, Z., Wu, X., 2022. Ensemble learning method based on temporal, spatial features with multi-scale filter banks for motor imagery EEG classification. Biomed. Signal Process. Control 76, 103634.

Zhu, X., Li, P., Li, C., Yao, D., Zhang, R., Xu, P., 2019. Separated channel convolutional neural network to realize the training free motor imagery BCI systems. Biomed. Signal Process. Control 49, 396–403.