MIN2Net: End-to-End Multi-Task Learning for Subject-Independent Motor Imagery EEG Classification

Phairot Autthasan, *Graduate student member, IEEE*, Rattanaphon Chaisaen, *Graduate student member, IEEE*, Thapanun Sudhawiyangkul, Phurin Rangpong, Suktipol Kiatthaveephong, Nat Dilokthanakul, Gun Bhakdisongkhram, Huy Phan, *Member, IEEE*, Cuntai Guan *Fellow, IEEE*, and Theerawit Wilaiprasitporn*, *Member, IEEE*

***Abstract*— *Objective:* Advances in the motor imagery (MI)-based brain-computer interfaces (BCIs) allow control of several applications by decoding neurophysiological phenomena, which are usually recorded by electroencephalography (EEG) using a non-invasive technique. Despite significant advances in MI-based BCI, EEG rhythms are specific to a subject and various changes over time. These issues point to significant challenges to enhance the classification performance, especially in a subject-independent manner. *Methods:* To overcome these challenges, we propose MIN2Net, a novel end-to-end multi-task learning to tackle this task. We integrate deep metric learning into a multi-task autoencoder to learn a compact and discriminative latent representation from EEG and perform classification simultaneously. *Results:* This approach reduces the complexity in pre-processing, results in significant performance improvement on EEG classification. Experimental results in a subject-independent manner show that MIN2Net outperforms the state-of-the-art techniques, achieving an F1-score improvement of 6.72% and 2.23% on the SMR-BCI and OpenBMI datasets, respectively. *Conclusion:* We demonstrate that MIN2Net improves discriminative information in the latent representation. *Significance:* This study indicates the possibility and practicality of using this model to develop MI-based BCI applications for new users without calibration.**

***Index Terms*— Brain-computer interfaces (BCIs), motor imagery (MI), multi-task learning, deep metric learning (DML), autoencoder (AE).**

I. Introduction

BRAIN-computer interface (BCI) systems allow users to non-muscularly communicate with a machine by classifying their neural activity patterns [1]. Recently, electroencephalography (EEG) has been widely used as a brain-activity recording method in BCIs because it provides a non-invasive and relatively cheaper way of measuring neural activity than other neural acquisition techniques. Moreover, EEG offers a higher temporal resolution compared to the other brain measurement techniques [2].

Four main types of neurophysiological patterns are widely used to develop EEG-based BCI applications. These include steady-state visual evoked potential (SSVEP), event-related potentials (ERPs), movement-related cortical potentials (MRCPs), and motor imagery (MI) [3]–[6]. Among these EEG measurements, MI, used in BCI systems, has been gaining more attention because it allows users to generate the suppression of oscillatory neural activity in specific frequency bands over the motor cortex region without external stimuli [7]. The neurophysiological patterns of MI originate from changing brain areas' activations in the sensorimotor cortices similar to limb movements. Furthermore, a recent study has demonstrated MI-based BCIs as an assistive tool in motor rehabilitation in paralyzed patients, such as post-stroke patients [8].

Most MI-based BCI applications rely on a subject-dependent setting. New users have to participate in the calibration process before using a BCI system, which is time-consuming, inconvenient, and exhausting. Recently, numerous zero-calibration methods have been proposed to diminish the number of calibration trials [9], [10]. One prominent method is a calibration-free or subject-independent, where training and testing data are from different subjects. This method exhibits the ability to offer new users to use the BCI system without the calibration phase [11], [12]. It is essential to develop reliable methods based on the subject-independent setting while preserving classification performance in an acceptable range. Thus, it is a challenge to find discriminative MI-EEG features that generalize across subjects.

One conventional hand-crafted feature is the power spectral density (PSD). Event-related desynchronization/synchronization (ERD/ERS) is the brain activity patterns from particular frequency bands (mu (9–13 Hz) and beta (22–29 Hz)) over the sensorimotor cortex region while performing MI [1], [13]. Therefore we can carry out MI classification by considering PSD of EEG [8], [14]. However, there are some limitations of hand-crafted features. The major limitation is the selection of distinguishable information within the raw EEG (e.g., frequency bands and EEG channels), merely depending on the prior knowledge of experts [15].

In recent years, the use of deep learning (DL) has been shown promising results and proven itself to be a successful set of models in the field of computer vision, speech recognition, and natural language processing [16]. In contrast to hand-crafted features, DL methods can simultaneously learn complicated patterns from multiple dimensions of the data. Thus, many BCI researchers have proposed advanced DL

This work was supported by PTT Public Company Limited, The SCB Public Company Limited, Thailand Science Research and Innovation (SRI62W1501), The Office of the Permanent Secretary of the Ministry of Higher Education, Science, Research and Innovation, Thailand (RGNS63-252) and National Research Council of Thailand (N41A640131) *(Phairot Autthasan and Rattanaphon Chaisaen contributed equally to this work) (* Corresponding author: Theerawit Wilaiprasitporn).*

P. Autthasan, R. Chaisaen, T. Sudhawiyangkul, P. Rangpong, S. Kiatthaveephong, N. Dilokthanakul, and T. Wilaiprasitporn are with Bio-inspired Robotics and Neural Engineering (BRAIN) Lab, School of Information Science and Technology (IST), Vidyasirimedhi Institute of Science & Technology (VISTEC), Rayong, Thailand (e-mail: theerawit.w@vistec.ac.th).

H. Phan is with the School of Electronic Engineering and Computer Science, Queen Mary University of London, United Kingdom.

C. Guan is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore.

G. Bhakdisongkhram is with the School of Physical Medicine and Rehabilitation, Institute of Medicine, Suranaree University of Technology, Nakhon Ratchasima, Thailand.

Code examples, and other supporting materials are available on https://github.com/IoBT-VISTEC/MIN2Net

architectures, and significant improvements have been reported for EEG-MI classification. Specifically, the use of convolutional neural networks (CNNs) has been widely applied in EEG-MI classification because it offers the ability to efficiently learn on both temporal and spatial features from EEG signals [17]–[19]. To extract the temporal and spatial connectivity patterns effectively, the combination of 2D-CNNs and long short-term memory units (LSTMs) has been adopted [20], [21]. Even though existing deep learning works have been considerably successful in EEG decoding on several MI datasets, these works perform well only in the subject-dependent task, lacking generalization capabilities on new users.

More recently, the DL-based multi-task autoencoder (multi-task AE) has been employed in the field of EEG-based BCI because it can efficiently learn for data compression and classification tasks simultaneously [22], [23]. However, to our best knowledge, few researchers have adopted the multi-task AE to learn features from EEG to address the MI classification problem because it lacks the ability to maintain discriminative patterns of the original EEG signals. To overcome this issue, we proposed *MIN2Net*, a novel end-to-end neural network architecture and loss function for training multi-task AE in the MI classification task. In this way, the proposed method can learn latent representations that preserve discriminative information of the original EEG data by fusing deep metric learning (DML). MIN2Net is optimized by minimizing three different loss functions simultaneously: reconstruction, cross-entropy, and triplet loss functions.

The three main contributions of this paper can be summarized as follows:

- We propose a novel end-to-end architecture that can effectively extract the meaningful features from EEG data without using high-complexity EEG pre-processing, resulting in an outstanding performance in the subject-independent MI classification. Furthermore, the proposed method demonstrates excellent performance compared to state-of-the-art algorithms in the subject-independent classification over two benchmark datasets.
- To the best of our knowledge, this is the first study proposing deep metric learning to a multi-task AE to improve the MI classification performance. The proposed method indicates the possibility of handling discriminative information in the latent representation.
- Investigation via visualization of the learned latent features is carried out to interpret the proposed method's classification superiority over other state-of-the-art algorithms.

The remainder of this paper is structured as follows. Section II explains some backgrounds and related work. Section III describes the pre-processing of EEG and the structure of the proposed method. Experimental results of the proposed method are presented in Section IV and discussed in Section V. Finally, the conclusion is explained in Section VI.

II. Related Work

In this section, we review the development of EEG-based MI classification and then conclude the limitations on the current MI-BCI research. We also describe the concepts of AE and deep metric learning, which relate to our work. Finally, we give an overview of our proposed method.

A. EEG-based Motor Imagery Classification

With the advance of machine learning, BCI researchers have increasingly proposed intelligent algorithms based on a subject-dependent setting to enhance EEG-based motor imagery decoding performance. Common spatial pattern (CSP) is one of the most popular and commonly used methods in MI-based BCI [24]. Features are effectively extracted via CSP can be achieved by maximizing the differences in the variances for the two classes of EEG signals. Filter bank common spatial pattern (FBCSP) [25] is one of the advanced CSP algorithms, which is based on using multiple frequency bands instead of limiting to a specific band. FBCSP has been proven to be the state-of-the-art method in EEG-based MI classification, owing to its outstanding results [26]. After passing EEG signals through the FBCSP, the meaningful brain features are obtained, and then the most discriminative features are selected using a feature selection method such as mutual information-based best individual feature (MIBIF) [25]. In terms of classifiers, many conventional algorithms such as support vector machine (SVM) and linear discriminant analysis (LDA) can be used to classify these features [12], [26]. In addition, there have been a variety of algorithms extending CSP and leading to good performance [27]–[29]. Some works that utilize minimum training samples have been proposed and validated on several MI datasets [30], [31]. Although these methods have outperformed state-of-the-art methods in the subject-dependent classification task, their performance still needs improvement in the subject-independent case.

One potential direction to improve the performance of EEG-MI classification is to leverage a large-scale EEG-MI dataset using deep learning models [12], [32]. In a more recent paper, Lee et al. [33] provided an OpenBMI dataset, where EEG data is measured with a large number of subjects in multiple sessions using the MI-BCI paradigm. Taking advantage of a large number of training samples, the OpenBMI dataset has become to be one of the benchmark EEG datasets. In a work by Kwon et al. [12], a subject-independent framework based on CNN architectures was proposed using spectral-spatial feature representation to improve subject-independent MI classification, resulting in a state-of-the-art performance over the OpenBMI dataset.

B. Deep Metric Learning Model

Deep metric learning (DML) is a method based on a distance metric concept with the goal of learning representation to measure data similarity, depending on the embedding features learned from a metric learning network [34]. Generally, similarity metric functions such as Euclidean distance, Mahalanobis distance, and cosine distance can be directly employed as the distance metric between two points. In recent years, numerous loss functions such as contrastive loss [35], triplet loss [36], and quadruplet loss [37] have been developed for DML and to enhance feature discrimination. These loss functions are used to compute similarity measures on correlated samples to enforce samples of the same class closer to each other and push samples of different classes apart from each other. Unlike other losses such as cross-entropy loss, where a single sample is used to calculate the gradient, the gradient of a DML loss relies on contrastive pairs, triplets, or quadruplets of samples. Recently, DML has been applied to the field of EEG-BCI studies and achieved promising results [38], [39]

C. Autoencoders

Autoencoder (AE) concept is one of the unsupervised learning algorithms introduced in 1986 [40]. AE is typically employed for data compression, denoising, dimensionality reduction, and feature extraction [11], [41], [42]. This network architecture demonstrates an ability to learn meaningful features from either unlabeled or labeled input data to create latent representation. The learned latent representation is then used to reconstruct the original input. The training objective of this network architecture is to minimize the reconstruction loss of input data. The learned latent representation

TABLE I
DESCRIPTION OF ALL CONSIDERED DATASETS

Datasets	# Subjects	MI Task	# Sessions	# Trials/Session	# Channels	Channel location	Frequency (Hz.)
BCIC IV 2a	9	Left hand vs. Right hand	2	144	20	FC_3, FC_1, FCz, FC_2, FC_4, C_5, C_3, C_1, Cz, C_2, C_4, C_6, CP_3, CP_1, CPz, CP_2, CP_4, P_1, Pz, P_2	250
SMR-BCI	14	Right hand vs. Feet	2	160	15	Small Laplacian electrode were placed at C_3, Cz, and C_4. Distances between neighboring electrodes were 2.5 cm.	512
OpenBMI	54	Left hand vs. Right hand	2	200	20	FC_5, FC_3, FC_1, FC_2, FC_4, FC_6, C_5, C_3, C_1, Cz, C_2, C_4, C_6, CP_5, CP_3, CP_1, CPz, CP_2, CP_4, CP_6	1,000

appears to be more efficient when the reconstruction data is closer to the input data.

In recent years, advanced AE architectures have been developed and adopted for EEG. Denoising sparse autoencoder (DSAE) [43] was proposed to improve the EEG-based epileptic seizure detection. The sparsity constraint of the DSAE makes the reconstruction of the original EEG from the corrupted EEG input more efficient. Furthermore, a compressed sensing (CS) method based on AE was proposed to handle the biopotentials and telemonitoring system [22]. They reveal achievements in both finding the optimal data compression and classifying electrocardiogram (ECG) and EEG signals. However, most studies only focused on using AE as an unsupervised learning method to extract the salient features of original data [43], [44]. Their models were not end-to-end learning paradigms since their classifiers need to be trained separately with the learned features and labeled data to identify the actual class.

To address the aforementioned issue, our previous work [23] presented ERPENet, a multi-task autoencoder-based model (multi-task AE) to jointly learn multi-task deep features from both unsupervised EEG-based ERP reconstruction and supervised EEG-based ERP classification. In particular, we demonstrated that ERPENet obtained an excellent performance in extracting information shared across different datasets for an ERP decoding task. In this study, we propose a new architecture and a loss function for training a multi-task AE, which is capable of handling three tasks simultaneously to deal with the EEG-based MI classification, described in the following sections.

III. METHODS

This section first describes the three benchmark datasets. After that, we describe the design of the proposed method and discuss its loss function. Finally, we elaborate on the EEG-MI classification using the proposed method in a comprehensive study.

A. Data Description

We evaluated the proposed method and other baseline methods on the BCIC IV 2a [45], SMR-BCI [46], and OpenBMI [33] datasets. The first two public datasets are well-known as the benchmark datasets for MI classification offered by the Graz University of Technology. The last one is the largest public MI dataset so far, provided by Korea University. The details of all databases are explained in Table I. Furthermore, the EEG data from all considered datasets was downsampled to have a sampling frequency of 100 Hz. Finally, the time interval of EEG between 0 s and 4 s after stimulus onset was selected as the MI period for all datasets.

B. Time-domain EEG Representation

In this study, the raw EEG signals are time-domain signals that change over time. Since the discriminative features of motor imagery are mainly distributed between 8 Hz and 30 Hz [12], [33], a fifth-order Butterworth band-pass filter is adopted to construct the filtered EEG data in the corresponding frequency bands. Towards this end, the filtered EEG data is used as the input of MIN2Net. Formally, we consider $x \in \mathbb{R}^{C \times T}$ as a single-trial filtered EEG data from k classes, and its corresponding label is defined to be $y \in \{1, 2, ..., k\}$, where C is the number of channels, and T is the number of sampled time points.

C. Proposed Architecture: MIN2Net

An overview of our proposed MIN2Net is illustrated in Fig. 1, and the detailed configuration of each layer is shown in the supplementary materials Table S1[†]. The MIN2Net is composed of three main modules: autoencoder, deep metric learning, and supervised learning.

1) Autoencoder: The autoencoder module in the MIN2Net consists of two major components the encoder $z = q(x)$ and the decoder $\hat{x} = p(z)$ components. In the encoder component, an input signal x is encoded into a latent vector z by reducing the input signal's dimension. For the decoder component, the given latent vector z is decoded back to the input signal $\hat{x}$. The AE module aims to aid the proposed model in extracting meaningful features from the EEG and providing discriminative patterns for different classes. The investigation of input and output EEG signals using the proposed model is summarized in the supplementary materials Fig. S1[†].

The encoder has two CNN blocks; each of them consists of a Conv2D layer, a batch normalization (BN) layer, an exponential linear unit (ELU), and an average pooling layer (AveragePooling2D). The final CNN layer's output is considered as the input of a fully connected layer for mapping the latent representation.

Inspired by CSP, this study utilizes the CNN approach as spatial filtering to effectively learn discriminative features from a set of EEG inputs (x). Each CNN block is operated on the channel mixing CNN concept [18], combining all channels of the input signals. The convolution operation is performed based on the linear combination of all the given channels, convoluted along the time dimension. Therefore, the output is constructed as a new time-series signal, simultaneously extracting spatial information from all the feature channels. Here, the encoder's hidden size is large for the first CNN layer but is gradually decreased in the following CNN layers. More details of the layers' parameters are shown in the supplementary material Table S1[†]. The average pooling layers are applied to extract the important features of the given input signals and reduce the number of parameters. The main benefit of applying the average pooling is to exploit layers with local filters to share weights among all channels of the given input signals. After every CNN layer, a BN layer is used before feeding into the subsequent average pooling layer. The feature maps after the final average pooling are transformed into

[†]https://github.com/IoBT-VISTEC/MIN2Net

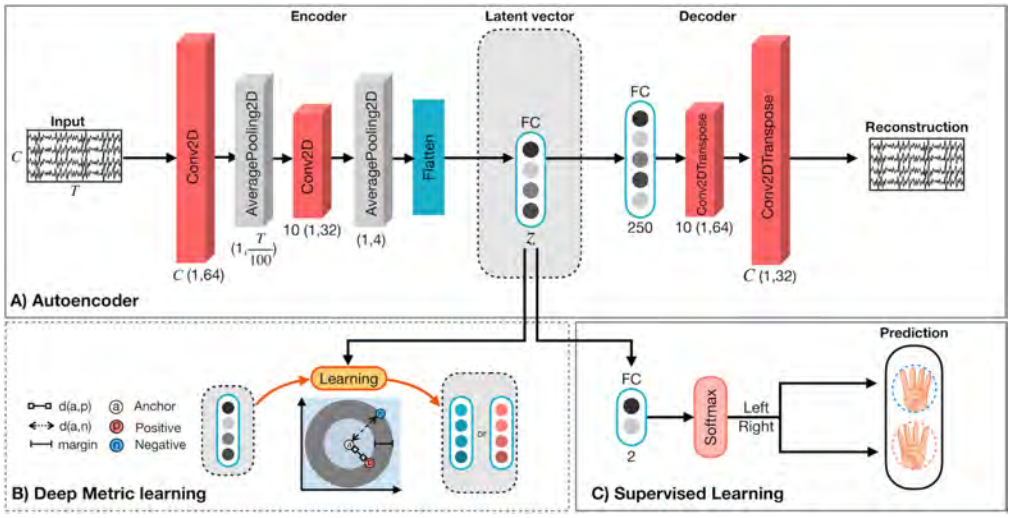

Fig. 1. Overall visualization of MIN2Net Architecture. (a) exhibits the AE network consisting of 3 components: encoder, latent vector, and decoder. The encoder compresses the input data and produces the latent vector, and then the decoder reconstructs the input data from this latent vector. (b) illustrates deep metric learning that learns to minimize the distances of embedding vectors of the same label while maximizing different labels. (c) displays the supervised classifier—the latent vector was fed into the FC layer using softmax activation for classification. Full details about the network architecture can be found in the supplementary materials Table S1†

a vector representation via flattening. Finally, the flattened vector is fed to a fully-connected (FC) layer with the hidden size of z units to embed and produce the latent vector. The latent vector size of z is expected to preserve the meaningful features for high-dimensional EEG-MI signals. Due to the requirement to minimize the size of the latent vector while preserving data representation, the grid search algorithm was performed on the set of $\{8, C, 64, 256\}$, where C is the number of channels in each considered dataset. We found that by setting the size of the latent vector to C and 64 for two- and three-class MI classification, respectively, MIN2Net achieved optimal performance. Note that the average classification performance of MIN2Net on all z values of each dataset is presented in the supplementary materials Table S2†.

For the decoder component, the decoder structure is arranged in a symmetrical way to the encoder component. Since it is essential to match the CNN blocks' input dimension, the latent vector is passed through the FC layer and then fed into a reshape layer to construct the data in a suitable dimension. Each of the two CNN blocks of the decoder component makes use of a transpose convolution layer (Conv2DTranspose) with a stride of 4 and an ELU layer. A stride of 4 is employed to upsample the data's size similarly to an upsampling layer. The transpose convolution can extract meaningful features and reduce useless features, which is beneficial for reconstructing the latent vector. Consequently, the reconstructed input signal is obtained after passing the two CNN blocks.

The training objective of the AE module is to minimize the reconstruction error between the input and the reconstruction. Here, we employ the mean square error (MSE) as the loss function. Given the input signals $x_j = \{x_1, x_2, ..., x_C\}$, the loss function is expressed as:

$$\mathcal{L}_{\text{MSE}}(x, \hat{x}) = \frac{1}{C} \sum_{j=1}^{C} \|x_j - \hat{x_j}\|^2. \quad (1)$$

Where $\hat{x_j}$ is the reconstruction signal of the channel j.

2) Deep Metric Learning: To preserve the distinguishable patterns in the latent representation of the AE, we introduce a deep metric learning module (DML) to the AE, extended from the latent vector. In general, the DML aims to learn a distance metric by improving the learned features' discrimination. This paper employs a triplet loss in the DML module to reflect the relative distances among different classes of the latent vectors. Owing to avoiding the risk of slow convergence and pool local optima, a semi-hard triplet constraint is used throughout all experiments, which was demonstrated good performance in [47]. During training, a set of triplets $\{x^a, x^p, x^n\}$ is randomly sampled from the training data, where the anchor sample x^a is closer to the positive sample x^p than the negative sample x^n. Subsequently, the triplet of three input signals is passed through the encoder component concurrently to obtain their latent vector z^a, z^p, and z^n. Thus, the loss function can be formulated as:

$$\mathcal{L}_{\text{triplet}}(z^a, z^p, z^n) = \frac{1}{2} \left[\|z^a - z^p\|^2 - \|z^a - z^n\|^2 + \alpha \right]_+. \quad (2)$$

where $[z]_+ = max(z, 0)$. The threshold α is the margin parameter that enforces the Euclidean distance $\|z^a - z^p\|^2$ of positive pairs to be shorter than the Euclidean distance $\|z^a - z^n\|^2$ of negative pairs. Importantly, the margin of the triplet loss plays a significant role in training the DML module.

3) Supervised Learning: This module utilizes a standard softmax classifier as a supervised classifier to classify the underlying latent vectors of the input EEG signals. The latent vector z is fed into the FC layer with the softmax activation to obtain the weight of importance for each class, expressed as follows:

$$\hat{y}(z) = softmax(W z + b) \quad (3)$$

Where W and b are the weight matrix and the bias vector, respectively. Then, the model is trained using Adam optimizer to minimize the cross-entropy loss, calculated as:

$$\mathcal{L}_{\text{cross-entropy}}(y, \hat{y}) = - \sum_{k=1}^{|class|} y_k \log \hat{y}_k. \quad (4)$$

where y and $\hat{y}$ are the true label and the classification probabilities respectively. The class with the maximum classification probability is identified as the predicted class of the single-trial EEG signal.

D. Training Procedure for MIN2Net

The training objective of the proposed method is optimized by incorporating the three loss functions: $\mathcal{L}_{\text{MSE}}$ in Equation 1, $\mathcal{L}_{\text{triplet}}$

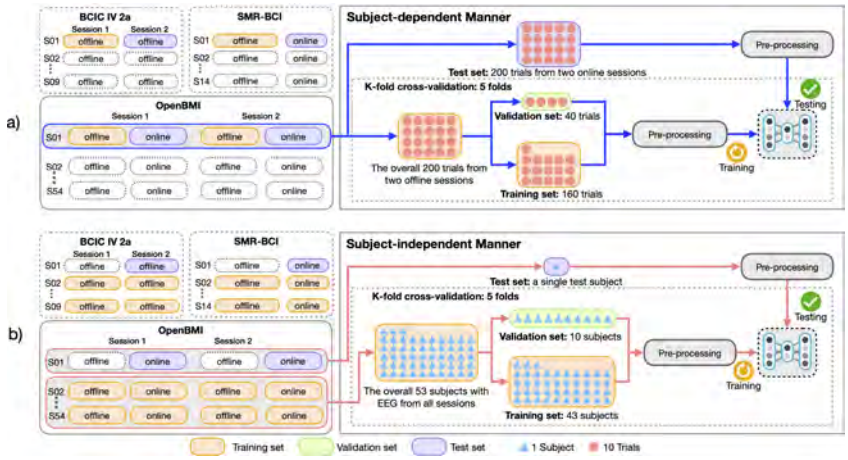

Fig. 2. Framework of a) subject-dependent and b) subject-independent with stratified k-fold cross-validation for the classification models.

in Equation 2, and $\mathcal{L}_{\text{cross-entropy}}$ in Equation 4. The final loss function of our MIN2Net model $\mathcal{L}_{\text{MIN2Net}}$ is expressed as:

$$\mathcal{L}_{\text{MIN2Net}}(x, \hat{x}, z^a, z^p, z^n, y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}\{\beta_1 \mathcal{L}_{\text{MSE}}(x_i, \hat{x_i}) + \beta_2 \mathcal{L}_{\text{triplet}}(z_i^a, z_i^p, z_i^n) + \beta_3 \mathcal{L}_{\text{cross-entropy}}(y_i, \hat{y_i})\}. \quad (5)$$

Where N denotes the total number of input signals. β_1, β_2, and β_3 represent the hyperparameters to weight the contribution of each loss function. As a result of integrating the three loss functions, both the unsupervised and the DML are able to influence the learning process when the supervised learning occurs.

E. Network Training

The proposed method was implemented using the Keras framework (TensorFlow v2.2.0 as backend). The training process was implemented using NVIDIA Tesla v100 GPU with 32GB memory. In each training iteration, the loss function was optimized by utilizing Adam optimizer with the learning rate schedule between [10^{-3}, 10^{-4}] for the binary classification task and the learning rate schedule between [10^{-4}, 10^{-5}] for the multi-class classification task. The learning rate of two considered tasks was decreased with a decay rate of 0.5 when there was no improvement in the validation loss for five consecutive epochs. We set a batch size of 10 samples for the subject-dependent classification setting and 100 samples for the subject-independent classification setting. Finally, the number of training iterations relied on the early stopping strategy so that the training process was stopped if there was no reduction of the validation loss for 20 consecutive epochs.

F. Baseline Methods

To demonstrate the effectiveness of our MIN2Net, we implemented four state-of-the-art methods for comparison. All deep learning approaches were implemented using the Keras framework (TensorFlow v2.2.0 as backend).

1) FBCSP-SVM: FBCSP was developed upon the idea of the original CSP algorithm [25]. Using the FBCSP as the feature extraction method, the distinguishable EEG features were extracted from multiple frequency bands. In this paper, the FBCSP was implemented using MNE-Python package (version 0.20) [48] and then applied with four spatial filters to decompose EEG signals into nine frequency bands with a bandwidth of 4 Hz from 4 to 40 Hz (4–8 Hz, 8–12 Hz, ..., 36–40 Hz). Here, each frequency band was created using bandpass filtering with 5$^{\text{th}}$ order non-causal Butterworth filter. Subsequently, a support vector machine (SVM) was used to classify MI by incorporating a grid search algorithm. For the SVM classifier, the hyperparameters consisted of kernel (linear, radial bias function (RBF), sigmoid), C (0.001, 0.01, 0.1, 1, 10, 100, 1000), and, particularly for RBF kernel, gamma (0.01, 0.001). With respect to the grid search algorithm, the prediction on the validation set of the classification was assessed to obtain the optimal set of hyperparameters. Eventually, the SVM classifier with the optimal parameters was used for testing purposes.

TABLE II
LIST OF THE OPTIMAL SET OF HYPERPARAMETERS FOR MIN2NET

Dataset	Subject-dependent			Subject-independent		
	β_1	β_2	β_3	β_1	β_2	β_3
BCIC IV 2a	1.0	0.1	1.0	0.5	0.1	1.0
SMR-BCI	0.1	0.1	1.0	0.1	1.0	0.1
OpenBMI	0.5	0.5	1.0	0.5	0.5	1.0

2) Deep Convnet: Deep Convnet was introduced as a DL model based on two CNN architectures [17] and proven to be effective for dealing with EEG-MI classification. In this study, the Deep Convnet was implemented in the optimal parameters as done in [17]. Moreover, the raw EEG data was band-pass filtered between 8 and 30 Hz (5$^{\text{th}}$ order non-causal Butterworth filter).

3) EEGNet-8,2: Inspired by the FBCSP method, EEGNet-8,2 was proposed as a compact CNN architecture to capture discriminative EEG features, which achieved outstanding performance in different BCI paradigms [19]. Here, EEGNet-8,2 was reproduced to offer a comparable performance. The network parameters were kept in the optimal set of hyperparameters as recommended in the original publication [19]. Furthermore, the raw EEG data were likewise preprocessed with the same protocol as in the Deep Convnet model to construct the input for the training of EEGNet-8,2.

4) Spectral-spatial CNN: The spectral-spatial CNN framework based on CNN architectures (spectral-spatial CNN) was presented by [12] and demonstrated state-of-the-art performance in a subject–independent MI decoding. The framework learned the spectral-spatial input, capturing discriminative features from the EEG signals' multiple frequency bands. In this paper, the raw EEG signals were similarly constructed the spectral-spatial representation as done in [12]. The spectral-spatial CNN model was implemented in the optimal parameters as defined in the original paper.

TABLE III

CLASSIFICATION PERFORMANCE (ACCURACY ± SD AND F1-SCORE ± SD) IN % OF MIN2NET USING THE SUBJECT-DEPENDENT AND SUBJECT-INDEPENDENT MANNERS COMPARISONS ON SIX DIFFERENT MARGINS (α). BOLD DENOTES THE BEST NUMERICAL VALUES.

Dataset	Margin (α)	Subject-dependent		Subject-independent	
		Accuracy	F1-score	Accuracy	F1-score
BCIC IV 2a	0.1	62.28 ± 13.90	62.61 ± 15.13	58.64 ± 8.57	46.59 ± 23.58
	0.5	62.25 ± 13.89	63.08 ± 14.70	59.27 ± 8.38	49.01 ± 19.29
	1.0	63.46 ± 14.33	64.28 ± 15.27	**60.03 ± 9.24**	**49.09 ± 23.28**
	5.0	63.66 ± 13.65	64.37 ± 14.57	59.61 ± 8.84	49.53 ± 19.56
	10.0	63.87 ± 14.51	64.03 ± 15.66	59.85 ± 8.44	48.39 ± 20.37
	100.0	**65.23 ± 16.14**	**64.72 ± 18.39**	58.78 ± 8.69	46.14 ± 24.29
SMR-BCI	0.1	64.00 ± 15.51	62.47 ± 16.60	56.76 ± 11.19	57.83 ± 20.50
	0.5	64.31 ± 15.70	62.27 ± 17.07	58.45 ± 12.67	58.41 ± 22.40
	1.0	**65.90 ± 16.50**	**64.13 ± 17.66**	**59.79 ± 13.72**	**61.10 ± 23.64**
	5.0	65.14 ± 16.08	62.04 ± 18.20	59.69 ± 13.86	58.88 ± 22.48
	10.0	65.45 ± 15.81	62.01 ± 17.82	58.81 ± 13.50	58.61 ± 23.43
	100.0	66.98 ± 17.22	62.77 ± 20.58	60.79 ± 13.73	60.47 ± 24.31
OpenBMI	0.1	59.25 ± 14.27	61.79 ± 14.26	72.14 ± 14.22	72.07 ± 15.19
	0.5	59.85 ± 13.93	62.17 ± 14.17	71.06 ± 13.91	71.23 ± 14.40
	1.0	**61.03 ± 14.47**	**63.59 ± 14.52**	**72.03 ± 14.04**	**72.62 ± 14.14**
	5.0	59.93 ± 13.77	62.41 ± 14.31	70.43 ± 13.81	71.00 ± 13.90
	10.0	58.97 ± 13.56	61.51 ± 14.14	69.43 ±14.29	69.46 ± 15.32
	100.0	57.14 ± 13.96	59.11 ± 15.98	70.48 ± 14.40	70.95 ± 15.25

TABLE IV

CLASSIFICATION PERFORMANCE (ACCURACY ± SD AND F1-SCORE ± SD) IN % OF MIN2NET COMPARED TO MIN2NET-WITHOUT TRIPLET AND MIN2NET-WITHOUT DECODER USING THE SUBJECT-DEPENDENT AND SUBJECT-INDEPENDENT MANNERS ON ALL DATASETS. BOLD DENOTES THE BEST NUMERICAL VALUES, AND * REPRESENTS THE PERFORMANCE VALUE WHICH WAS SIGNIFICANTLY HIGHER THAN ALL COMPARISON PAIRS, $p < 0.05$.

Dataset	Comparison Model	Subject-dependent		Subject-independent	
		Accuracy	F1-score	Accuracy	F1-score
BCIC IV 2a	MIN2Net-w/o triplet	60.76 ± 11.93	61.09 ± 13.83	58.70 ± 8.91	**49.36 ± 20.25**
	MIN2Net-w/o decoder	**65.71 ± 16.16**	**65.46 ± 18.34**	57.55 ± 9.06	44.24 ± 24.89
	MIN2Net	65.23 ± 16.14	64.72 ± 18.39	**60.03 ± 9.24**	49.09 ± 23.28
SMR-BCI	MIN2Net-w/o triplet	63.86 ± 14.13	61.31 ± 16.19	57.95 ± 12.55	60.53 ± 20.33
	MIN2Net-w/o decoder	64.86 ± 16.21	62.65 ± 17.99	57.38 ± 12.22	55.28 ± 22.17
	MIN2Net	**65.90 ± 16.50**	**64.13 ± 17.60**	**59.79 ± 13.72**	**61.10 ± 23.64**
OpenBMI	MIN2Net-w/o triplet	59.66 ± 14.02	61.64 ± 14.44	71.10 ± 13.58	69.28 ± 16.10
	MIN2Net-w/o decoder	58.76 ± 13.79	61.70 ± 13.64	70.59 ± 14.23	70.69 ± 14.48
	MIN2Net	**61.03 ± 14.47***	**63.59 ± 14.52***	**72.03 ± 14.04***	**72.62 ± 14.14***

G. Experimental Evaluation

To demonstrate our MIN2Net method as a generalized MI classification model, we conducted the experiments with both subject-dependent and subject-independent manners on three benchmark datasets (BCIC IV 2a, SMR-BCI, and OpenBMI). Accuracy and F1-score were used to evaluate the performance of all considered methods.

Fig. 2(a) exhibits an example of how we divided the training and testing sets in the subject-dependent manner. For BCIC IV 2a dataset, the offline data from session 1 was used as the training set, and the offline from session 2 was used as the testing set. Meanwhile, the training and testing sets were obtained from the offline and online sessions for SMR-BCI and OpenBMI datasets. The stratified 5-fold CV was then utilized to split the training set into the new training and validation sets for parameter search. Each fold was performed by preserving 50 percent of samples for each class in the new training and validation sets.

The subject-independent manner was conducted with a leave-one(subject)-out cross-validation (LOSO-CV), as illustrated in Fig. 2(b). Suppose that there are N_s subjects for a specific dataset. In each fold of LOSO-CV, a single subject was used as the testing set, and the remaining $N_s - 1$ subjects were employed as the training set to obtain N_s classification results. The training set was constructed using all data sessions of all $N_s - 1$ training subjects for each dataset. Meanwhile, we chose the offline session 2 of the test subject as the test set of BCIC IV 2a dataset and the online session of the test subject as the test set for both SMR-BCI and OpenBMI datasets. Furthermore, the stratified 5-fold CV scheme was adopted on the training set to find an optimal set of all classifier parameters. Finally, we calculated the average classification accuracy from the $N_s \times 5$ evaluations as the overall performance of MIN2Net and other baseline methods.

The details of the four experiments of the entire study were described as follows:

1) Experiment I: To find the optimal set of hyperparameters of MIN2Net, we conducted initial experiments for parameter search. We first experimented with EEG-MI binary classification to tune β parameters for the loss function of MIN2Net in Equation 5. The grid search algorithm was carried out in the set of $\{0.1, 0.5, 1.0\}$ for β_1, β_2, and β_3. As shown in Equation 2, the margin α plays a significant role during training MIN2Net. To examine the effect of this hyperparameter, we performed experiments with MIN2Net to search for an optimal value of α in the set $\{0.1, 0.5, 1.0, 5.0, 10.0, 100.0\}$.

Furthermore, we conducted an ablation study to examine the effectiveness of each component in MIN2Net. We compared our complete MIN2Net model with two modification models:

- MIN2Net-without triplet: the MIN2Net without the DML module
- MIN2Net-without decoder: the MIN2Net without decoder part of AE module

We then performed one-way repeated measures analysis of variance (ANOVA) with Bonferroni correction to evaluate the significant differences among the classification performance of MIN2Net and the aforementioned modification models.

2) Experiment II: In this study, we compared the EEG-MI classification performance of MIN2Net with the baseline methods. This experiment was conducted based on EEG-MI binary classification to investigate all methods' effectiveness over the three considered datasets in both the subject-dependent and subject-independent scenarios. We evaluated all methods on the same training, validation, and testing sets to make a fair comparison. We carried out a one-way repeated measures analysis of variance (ANOVA) with Bonferroni correction to analyze the classification performance's significant differences between our MIN2Net and all baseline methods.

3) Experiment III: To demonstrate the practicality of MIN2Net in developing real-world applications, we performed a three-class EEG-MI classification task over the OpenBMI dataset (right hand MI vs. left hand MI vs. resting EEG). The purpose of this experiment was to compare the effectiveness of MIN2Net to all other baseline methods in pseudo-online scenarios. This experiment was conducted based on the subject-independent manner because MIN2Net demonstrated superior performance in Experiments I and II. Here, both the right and left hand MI signals were likewise segmented with the same protocol as in subsection III-A. Meanwhile, the resting EEG period was defined as the time interval between 4 and 8 seconds after stimulus onset. Since the resting EEG was obtained from all EEG recordings, the number of trials in each MI class was less than the number of trials in the resting EEG class. To address the imbalance issue, we decided to select half of all resting EEG trials randomly. The detailed experimental procedure is summarized in the supplementary materials Fig. S2[†]. To compare all the used methods, we employed an ANOVA with Bonferroni correction for statistical analysis.

4) Experiment IV: As illustrated in Table V and Fig. 4, MIN2Net performed suboptimally in the subject-dependent scenario on all three benchmark datasets. The primary reason is that MIN2Net was trained using a small number of training samples from a single subject, which

posed an overfitting problem. In general, the deep learning approach requires a large amount of data in order to train a generalized model efficiently and avoid the overfitting problem. As shown in Fig. 6, the result established convincingly that MIN2Net outperformed the others when a large number of training samples was used. As the result, one suggestion for alleviating this overfitting problem is to increase training samples in the subject-dependent scenario via data augmentation. We hypothesized that increasing the number of training samples will aid the model in capturing generalized features and thus improving MI classification performance. To prove this, we examined how the subject-dependent settings combined with varying amounts of data augmentation improved the performance of MIN2Net on all considered datasets. As introduced by [49] for bio-signals, Jittering (Jitter), Magnitude-Warping (MagW), Scaling (Scale), Time-Warping (TimeW), and Permutation (Perm) were used to augment the training data in order to generate new ones. We used an ANOVA with Bonferroni correction for statistical analysis to compare the significant differences in classification performance between MIN2Net with and without data augmentation.

H. Visualization

To compare the capabilities of deep learning approaches for extracting highly discriminative features from EEG signals, we employed the t-SNE method [50] to visualize the generalized brain features learned by different deep learning methods in the two-dimensional embedding space. The t-SNE algorithm was used to visualize the high-dimensional embedding space at the input of the last fully connected layer in all trained models. Additionally, we used the t-SNE directly to raw EEG data as a baseline to visualize the distribution of unlearned brain features. During the t-SNE extraction process, the raw EEG contained a collection of samples $\times$ channels $\times$ time points (EEG amplitude values) is reshaped to be a dimension of samples $\times$ features (all EEG amplitude values from all channels) before computing the t-SNE projection.

IV. Results

This section reports the results and statistical analysis of Experiment I, II, III, and IV to validate the effectiveness of MIN2Net. Furthermore, we visualize the learned EEG features to demonstrate the discriminative power of the features learned by the MIN2Net. The performance of each experiment was reported as accuracy and F1-score with standard deviation (Accuracy $\pm$ SD and F1-score $\pm$ SD).

A. Experiment I: Parameters Adjustment

Table II presents the optimal hyperparameters to adjust each module's weight (β_1, β_2, and β_3) for the loss function of MIN2Net in Equation 5, resulting in the optimal performance of classifying EEG-MI data. Note that the average classification performance of MIN2Net on all β combinations of each dataset is reported in the supplementary materials[†]. Table III illustrates the classification results when different values of the margin parameters (α) are contributed in the DML module of MIN2Net. We observed that the margin parameter's size had a significant impact on the final classification performance, and when marking the margin value as 1.0, the MIN2Net achieved the best performance in a subject-independent manner for all datasets. In a subject-dependent manner, the margin value of 100.0 contributed to the best performance of MIN2Net on the BCIC IV 2a dataset, whereas the best performance of MIN2Net on both SMR-BCI and OpenBMI datasets were obtained by setting the margin value as 1.0. The results of the ablation study are summarized in Table IV. It

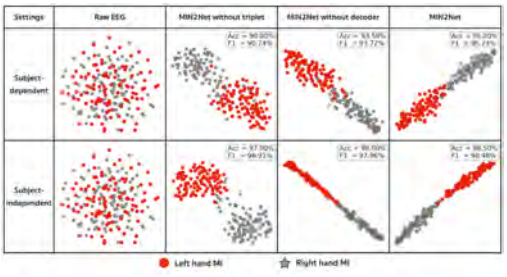

Fig. 3. Comparison of two-dimensional t-SNE projections for a single subject's binary classification in the OpenBMI dataset. The picture depicts the raw EEG features and the latent EEG features generated by the MIN2Net and their modifications in the subject-dependent and subject-independent settings.

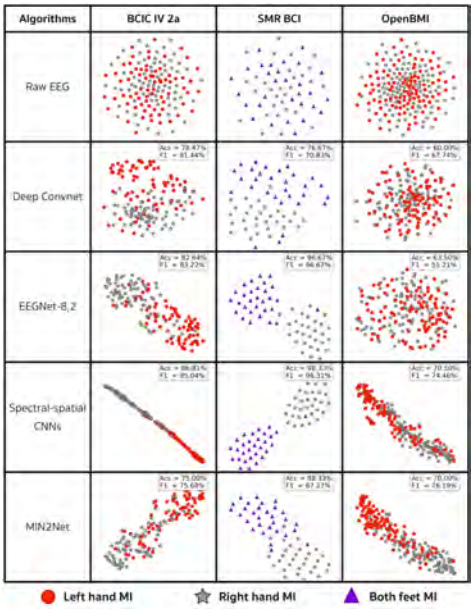

Fig. 4. Visualization of raw and learned EEG features produced by all used approaches for a single subject using t-SNE projection. The picture displays a comparison of two-dimensional t-SNE projections for the subject-dependent binary classification.

can be seen that the classification performance of MIN2Net outperformed both MIN2Net-without triplet and MIN2Net-without decoder models in terms of accuracy and F1-score on both subject-dependent and subject-independent manners for both SMR-BCI and OpenBMI datasets. In a paired t-test, MIN2Net was significantly higher than these modification models in both manners for the OpenBMI dataset, $p < 0.05$. However, on the BCIC IV 2a dataset, the performance improvement of MIN2Net to their modification models was not found in both manners. Furthermore, Fig. 3 illustrates the t-SNE projection of learned embedding features of MIN2Net and their modifications in both scenarios from the OpenBMI dataset.

B. Experiment II: Binary MI classification

Table V presents the overall performance of our MIN2Net and four baseline methods across all subjects for both subject-dependent and subject-independent settings. It is observed that in a subject-independent manner, MIN2Net achieved the highest performance in terms of accuracy on the OpenBMI dataset and terms of F1-score on both SMR-BCI and OpenBMI datasets. Specifically, significant differences were seen among the accuracy and F1-score provided by MIN2Net and the other baseline methods in the OpenBMI

TABLE V

CLASSIFICATION PERFORMANCE (ACCURACY ± SD AND F1-SCORE ± SD) IN % FOR THE SUBJECT-DEPENDENT AND SUBJECT-INDEPENDENT SCHEMES ON BCIC IV 2A, SMR-BCI, AND OPENBMI COMPARED TO FIVE DIFFERENT METHODS. BOLD DENOTES THE BEST NUMERICAL VALUES, AND * REPRESENTS THE PERFORMANCE VALUE WHICH WAS SIGNIFICANTLY HIGHER THAN ALL COMPARISON PAIRS, $p < 0.05$.

Dataset	Comparison Model	End-to-end	Subject-dependent		Subject-independent	
			Accuracy	F1-score	Accuracy	F1-score
BCIC IV 2a (9 subjects, 288 trials/subject)	FBCSP-SVM	No	75.93 ± 14.93	74.49 ± 18.68	58.09 ± 9.91	51.53 ± 24.01
	Deep Convnet	Yes	63.72 ± 17.18	59.85 ± 22.17	56.34 ± 8.86	30.62 ± 28.96
	EEGNet-8,2	Yes	65.93 ± 18.44	64.45 ± 26.23	64.26 ± 11.03	60.19 ± 19.96
	Spectral-Spatial CNN	No	**76.91 ± 13.75**	**77.03 ± 15.41**	**66.05 ± 13.70**	**61.91 ± 20.31**
	MIN2Net	Yes	65.23 ± 16.14	64.72 ± 18.39	60.03 ± 9.24	49.09 ± 23.28
SMR-BCI (14 subjects, 160 trials/subject)	FBCSP-SVM	No	74.50 ± 18.14	**70.65 ± 23.64**	62.64 ± 15.43	45.07 ± 34.93
	Deep Convnet	Yes	61.40 ± 15.66	55.27 ± 22.00	65.26 ± 16.83	54.38 ± 32.58
	EEGNet-8,2	Yes	67.76 ± 18.09	68.05 ± 21.11	58.07 ± 11.45	34.43 ± 31.35
	Spectral-Spatial CNN	No	**76.76 ± 16.66**	69.87 ± 28.15	**66.21 ± 15.15**	54.36 ± 31.21
	MIN2Net	Yes	65.90 ± 16.50	64.13 ± 17.66	59.79 ± 13.72	**61.10 ± 23.64**
OpenBMI (54 subjects, 400 trials/subject)	FBCSP-SVM	No	**66.06 ± 16.58**	64.66 ± 19.47	64.96 ± 12.70	65.25 ± 15.14
	Deep Convnet	Yes	60.31 ± 16.76	61.66 ± 18.17	68.33 ± 15.33	70.20 ± 15.18
	EEGNet-8,2	Yes	60.41 ± 17.12	56.80 ± 23.54	68.84 ± 13.87	70.39 ± 14.30
	Spectral-Spatial CNN	No	65.19 ± 15.94	**66.97 ± 16.71***	68.27 ± 13.56	65.86 ± 17.37
	MIN2Net	Yes	61.03 ± 14.47	63.59 ± 14.52	**72.03 ± 14.04***	**72.62 ± 14.14***

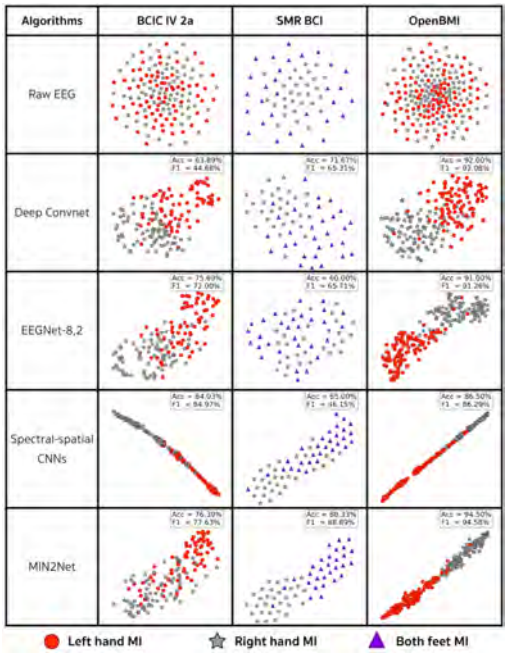

Fig. 5. Visualization of raw and learned EEG features produced by all used approaches for a single subject using t-SNE projection. The picture exhibits a comparison of two-dimensional t-SNE projections for the subject-independent binary classification.

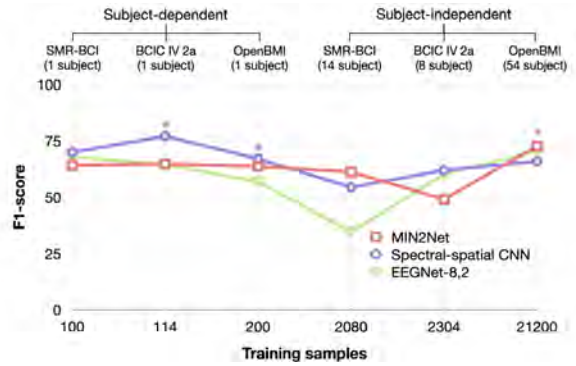

Fig. 6. Effect of the number of training samples on the binary classification performance across three considered methods.

dataset, $p < 0.05$. Considering the SMR-BCI dataset, the F1-score improvement of MIN2Net to the baseline methods was significant ($p < 0.05$) except for the Deep Convnet and Spectal-spatial CNN models. However, the overall performance of MIN2Net was lower than some baseline methods in the subject-independent over the BCIC IV 2a dataset. Furthermore, in the subject-dependent setting, the performance improvement of MIN2Net to the baseline methods was not found on three considered datasets.

Moreover, we reveal the average training and prediction times for all subjects per epoch on all considered datasets as shown in Table VI. Note that the training time was identified as the duration time in each training iteration. Meanwhile, the prediction time was defined as the duration time in each fold to classify all testing samples. The results also present the number of trainable parameters for MIN2Net and all baseline methods.

Fig. 4 and Fig. 5 illustrate the t-SNE projection of the learned embedding features of all the datasets for the subject-dependent and subject-independent manners. The results display the two-dimensional embedding features of MIN2Net and all baseline methods, considering all trials of one testing subject of each dataset.

Fig. 6 displays the variations of the classification F1-score with respect to the number of training samples. The number of training samples in all the datasets is represented on the x-axis, while the y-axis expresses the binary classification F1-score of MIN2Net and the two best baseline methods. It is demonstrated that increasing the number of training samples from 100 to 21200 samples can provide a better classification F1-score, recommending a significant factor to boost the final classification F1-score. Therefore, the classification F1-score of MIN2Net was shown to be higher than the F1-score of the two baseline methods when trained with a large number of training samples.

C. Experiment III: Multi-class MI classification

Table VII exhibits the entire performance of the three-class MI classification on the OpenBMI dataset by comparing MIN2Net, and all baseline approaches. It was found that in a subject-independent manner, MIN2Net outperformed all baseline methods with the ac-

TABLE VI
TIME COMPLEXITY OF TRAINING (T_{train}) AND PREDICTION (T_{pred}) IN SECONDS PER EPOCH FOR ALL METHODS AND NUMBER OF TRAINABLE PARAMETERS FOR ALL DEEP LEARNING APPROACHES.

Dataset	Comparison Model	*# trainable params*	Subject-dependent		Subject-independent	
			T_{train}	T_{pred}	T_{train}	T_{pred}
BCIC IV 2a	FBCSP-SVM	-	-	0.0008	-	0.0112
	Deep Convnet	151,027	0.1709	0.1617	0.2748	0.1739
	EEGNet-8,2	5,162	0.1476	0.1173	0.3735	0.0920
	Spectral-Spatial CNN	77,577,714	10.2031	0.7600	8.1334	0.7444
	MIN2Net	55,232	0.2320	0.1803	0.4724	0.2373
SMR-BCI	FBCSP-SVM	-	-	0.0005	-	0.0047
	Deep Convnet	150,302	0.1352	0.1519	0.2412	0.1906
	EEGNet-8,2	5,082	0.1164	0.1296	0.3210	0.1105
	Spectral-Spatial CNN	54,076,914	2.1321	1.0257	5.8785	0.6688
	MIN2Net	38,297	0.1463	0.2433	0.4948	0.2966
OpenBMI	FBCSP-SVM	-	-	0.0020	-	0.1906
	Deep Convnet	153,427	0.1804	0.1618	1.7497	0.4734
	EEGNet-8,2	5,162	0.1882	0.1439	3.0951	0.1372
	Spectral-Spatial CNN	77,577,714	2.2476	1.0934	11.9067	0.8560
	MIN2Net	55,232	0.3527	0.2851	1.3626	0.1043

curacy and F1-score of 68.81 $\pm$ 12.44% and 68.04 $\pm$ 12.97%, respectively. Moreover, there were significant differences in the accuracy and F1-score between MIN2Net and all baseline methods, $p < 0.05$. Fig. 7 shows the confusion matrix of MIN2Net in the three-class MI classification on OpenBMI dataset. It was observed that in the subject-independent scenario, MIN2Net yielded the highest recall of resting EEG class and the lowest of right hand MI.

Fig. 8 reveals the scatter plots of the learned embedding features using t-SNE of the OpenBMI dataset. Results of MIN2Net and the others are considered the three-class MI classification task from all trials of one representative subject. Similar to the t-SNE projection results in the binary classification, the learned embedding features from three different classes tended to be separated into three compact clusters and made an effort to maintain the relative distances among different clusters.

D. Experiment IV: Data Augmentation

Table VIII reports a comparison of the MIN2Net's classification performance with and without data augmentation on three benchmark datasets. As can be seen, MIN2Net, utilizing 100 percent data augmentation outperformed all other methods in terms of classification accuracy and F-score on all used datasets. In a paired t-test, MIN2Net with 100 percent data augmentation significantly outperformed all other models on the OpenBMI dataset, $p < 0.05$. Using both the BCIC IV 2a and SMR-BCI datasets, only MIN2Net with 100 percent data augmentation and without data augmentation showed significant differences, $p < 0.05$.

The image clusters in Fig. 9 illustrate the learned embedding features for three benchmark datasets using t-SNE. The binary MI classification task from all trials of one representative subject is used to evaluate MIN2Net with and without data augmentation.

V. DISCUSSION

A. Effectiveness of Deep Metric Learning

Deep learning (DL) has made significant contributions to the development of efficient MI-based BCI applications. End-to-end multi-task AE is a powerful DL technique for processing raw EEG data due to the combination of feature extraction and classification. Even though AE can recognize instances rather than distinguish between classes, the current study applies DML-based triplet loss to the multi-task AE to learn the relative distances among different classes of latent representations. According to the results from Table IV, the MIN2Net's classification performance is significantly higher than

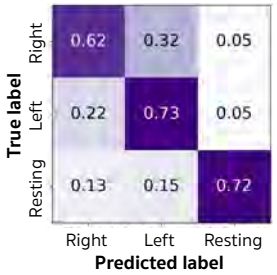

Fig. 7. Confusion matrix of three-class MI classification.

TABLE VII
CLASSIFICATION ACCURACY AND F1-SCORE (IN %, $\pm$ SD) ON THE OPENBMI DATASET FOR THE THREE-CLASS CLASSIFICATION OF MI IN THE SUBJECT-INDEPENDENT MANNER. BOLD DENOTES THE BEST NUMERICAL VALUES, AND * REPRESENTS THE PERFORMANCE VALUE WHICH WAS SIGNIFICANTLY HIGHER THAN ALL COMPARISON PAIRS, $p < 0.05$.

Algorithm	Accuracy	F1-score
FBCSP-SVM	50.71 $\pm$ 9.99	46.29 $\pm$ 12.40
Deep Convnet	54.04 $\pm$ 10.12	49.77 $\pm$ 12.58
EEGNet-8,2	67.93 $\pm$ 11.94	66.41 $\pm$ 13.23
Spectral-spatial CNN	64.67 $\pm$ 11.63	63.16 $\pm$ 12.53
MIN2Net	**68.97±11.84***	**68.07±12.34***

the MIN2Net-without decoder and MIN2Net-without triplet on two benchmark datasets. Fig. 3 illustrates that the embedding features of MIN2Net from different classes were clustered towards a more compact form. By contrast, the embedding features of their modifications from different classes appeared to be less compact and dispersed throughout the projection space. This evidence suggests that the DML based on triplet performs well when incorporating with the multi-task AE architecture, resulting in learning and improving the discriminative pattern of EEG data among different classes.

B. Analysis of the Proposed Method

Recently, the deep learning technique has attained popularity in BCI because it is capable of effectively learning the brain activity patterns from EEG data without using high complexity in EEG pre-processing [51]. In this paper, our MIN2Net method's input is the time-domain EEG signals, filtered using a particular frequency band to eliminate high- and low-frequency artifacts. Based on Experiment I and II results, MIN2Net performs with higher accuracy than FBCSP and spectral-spatial CNN, which use multiple frequency bands to filter out artifacts. These results suggest that MIN2Net is robust to artifacts and offers higher classification results than the other baseline methods, utilizing simplistic EEG pre-processing only once.

To give insight into an internal perspective behind the optimization process of MIN2Net, we examined the changes of both training and validation losses in the binary classification during the training process of the OpenBMI dataset in a subject-independent manner. Four different losses of MIN2Net (MSE, triplet, cross-entropy, and the total losses) were monitored for 60 epochs from all subjects. It was observed that all the four losses converged around 15 epochs, as shown in Fig. 10. Similar to the convergence process on the OpenBMI dataset, we also found signs of the convergence of these four losses within 60 epochs over the BCIC IV 2a and SMR-BCI datasets. As the results in Table VI, MIN2Net has the 2nd smallest size of trainable parameters on all considered datasets. Furthermore, the MIN2Net has a speed of training and prediction similar to the compact baseline models such as EEGNet-8,2 and Deep Convnet.

TABLE VIII

CLASSIFICATION ACCURACY AND F1-SCORE (IN %, ± SD) OF MIN2NET WITH AND WITHOUT DATA AUGMENTATION IN THE SUBJECT-DEPENDENT MANNER. BOLD DENOTES THE BEST NUMERICAL VALUES, AND * REPRESENTS THE PERFORMANCE VALUE WHICH WAS SIGNIFICANTLY HIGHER THAN ALL COMPARISON PAIRS, $p < 0.05$.

Dataset	Percentage of data augmentation	Accuracy	F1-score
BCIC IV 2a	No augmentation	65.23 ± 16.14	64.72 ± 18.39
	25	68.24 ± 16.81	69.22 ± 16.04
	50	69.44 ± 18.31	70.01 ± 17.80
	75	69.01 ± 16.78	69.41 ± 16.40
	100	**70.09 ± 16.87**	**70.44 ± 16.49**
SMR-BCI	No augmentation	65.90 ± 16.50	64.13 ± 17.66
	25	70.55 ± 17.57	66.41 ± 22.37
	50	71.62 ± 17.48	69.30 ± 18.98
	75	72.33 ± 17.48	69.29 ± 21.48
	100	**72.95 ± 15.76**	**69.51 ± 20.00**
OpenBMI	No augmentation	61.03 ± 14.47	63.59 ± 14.52
	25	64.11 ± 15.89	65.93 ± 15.54
	50	65.06 ± 15.19	66.42 ± 14.90
	75	66.10 ± 14.99	67.59 ± 14.52
	100	**66.51 ± 15.53***	**68.47 ± 14.65***

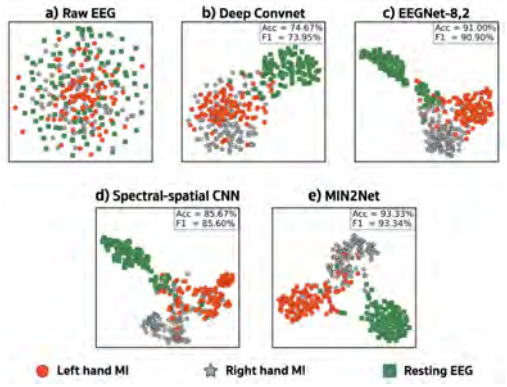

Fig. 8. Visualization of EEG features using two-dimensional t-SNE projection on the three-class classification of EEG-MI data in the subject-independent manner. We picked learned EEG features from one subject on OpenBMI dataset for visualization purpose.

Regarding the module's weight (β_1, β_2, and β_3) for the loss function of MIN2Net shown in Table II, we can observe that the optimal performance obtains from different sets of the module's weight. The reason is related to the difference in the amount of data among all considered datasets, where the small dataset requires the small values in the module's weight to prevent overfitting. Meanwhile, all the module's weight values close to 1 are desirable for the large dataset to achieve optimal performance. We also found that when all three β have the same value, there is a slight difference between the small and the large values on the large dataset. However, using the small dataset shows a considerable difference between the small and the large values, as shown in the supplementary materials†.

C. Analysis of the Comparison Performance

The binary classification results on three benchmark datasets are listed in Table V. It can be observed that on SMR-BCI and OpenBMI datasets, MIN2Net outperforms all baseline methods in a subject-independent setting. Even though the SMR-BCI dataset's accuracy of MIN2Net is lower than some baseline methods, the F1-score of MIN2Net is higher than all baseline methods. According to the claims in [52] that F1-score is more valuable than accuracy because it allows for both false positives and false negatives. Additionally, although the two benchmark datasets have different training samples, MIN2Net still results in the best performance in MI classification for both

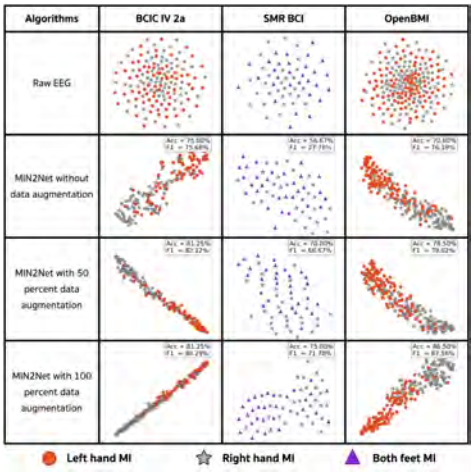

Fig. 9. Comparison of two-dimensional t-SNE projections for a single subject's binary classification. The figure depicts the raw EEG features and latent EEG features generated by the MIN2Net and the MIN2Net with data augmentation in the subject-dependent setting.

datasets. This investigation suggests that incorporating multi-task AE and DML, as done in MIN2Net, plays a vital role in extracting generalized EEG features for MI classification, resulting in excellent generalization performance on new subjects.

However, in a subject-dependent setting, the results indicate that MIN2Net performs suboptimally on all used datasets. The reason for this is that MIN2Net based on the integration of multi-task AE and DML does not perform well when using few training samples from only a single subject, resulting in an overfitting problem. According to the results in Table VIII, it is demonstrated that by incorporating data augmentation techniques into MIN2Net in a subject-dependent setting, an overfitting problem is avoided, and classification performance is improved. Interestingly, when 100 percent data augmentation is used, MIN2Net significantly outperforms MIN2Net without data augmentation on all the used datasets. This finding implies that increasing the number of training samples could help MIN2Net capture generalized features and improve MI classification performance.

D. Visualization of the Learned Latent Representation

According to Fig. 4, when compared to EEGNet-8,2 and Spectral-spatial CNNs approaches, the latent embedding features generated by MIN2Net in a subject-dependent MI classification task are more likely to be less compact and appear to be expanding out the projection space. As a consequence of the low-quality representation of the MI in the learned latent embedding features, this observation reflects that MIN2Net does not perform well in the learning process using a few training samples. Nonetheless, as illustrated in Fig. 9, including data augmentation methods in MIN2Net can provide a higher-quality representation of latent embedding features.

In a subject-independent scenario, as illustrated in Fig. 5, MIN2Net produces highly discriminative patterns over the SMR-BCI and OpenBMI datasets compared to the other baseline methods. Similarly, in multi-class classification, the latent embedding features extracted by MIN2Net generate considerably more discriminative patterns than the other baseline approaches on the OpenBMI dataset, as presented in Fig. 8. As a result, this demonstrates that MIN2Net outperforms others due to the higher quality representation of the MI in the learned latent embedding features.

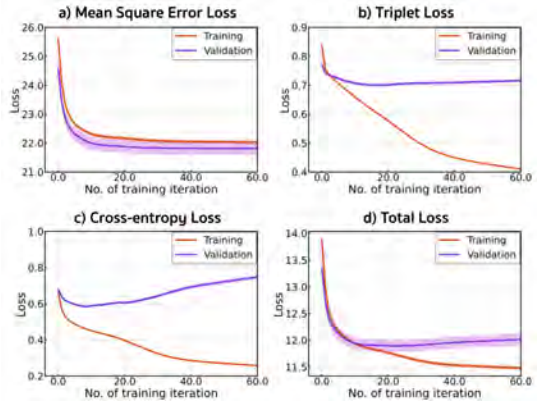

Fig. 10. Training and validation losses of our proposed MIN2Net on OpenBMI dataset. The plots show averaged losses with standard error of 54 subjects while d) total loss was weighed by 0.5, 0.5, 1.0 for a) mean square error b) triplet and c) cross-entropy loss respectively.

E. Feasibility in Online BCI Systems

To further evaluate the feasibility and pseudo-online performance of MIN2Net, we established an experiment on the OpenBMI dataset, exhibiting the classification among resting EEG, left hand MI, and right hand MI. The results demonstrate that MIN2Net significantly outperforms all baseline methods in the subject-independent setting, as depicted in Table VII. Additionally, MIN2Net yields an acceptable misclassification rate in the classification of three-class MI, as depicted in Fig. 7. The present results confirm the possibility of using the MIN2Net in two promising aspects. First, MIN2Net has the capability of classifying multi-class MI-EEG. Second, MIN2Net could integrate with online BCI applications, providing various movement intentions for users, such as stop moving, left-hand grasping, or right-hand grasping.

Moreover, these research findings provide foundations in developing BCI applications based on a calibration-free method. MIN2Net can be used as a pre-trained model, where the model is pre-trained before being applied by a new user. The latency of MIN2Net is considered as a prediction time in the testing session. With the subject-independent setting, the BCI application considers the prediction time rather than the training time. According to Table VI, it is found that the prediction time to classify all testing trials is 0.2373 s, 0.2966 s, and 0.1043 s for BCIC IV 2a, SMR-BCI, and OpenBMI datasets, respectively. As shown in Fig. 6, when few subjects are used for training, we could observe that the proposed model provides suboptimal performance. On the other hand, training with more subjects could improve the overall performance of MIN2Net. Therefore, once we perceive a new user as an outstanding BCI user, we can include that user in our dataset to retrain the proposed model to achieve better classification performance.

F. Future Directions

Although MIN2Net achieves a promising classification result, there are still several rooms for further improvement. Firstly, the DML module is based on triplet loss, and there are numerous new loss functions developed for DML that are more attractive to investigate [35], [37]. Thus, in the future study, we will focus on incorporating these new loss functions into multi-task AE to improve classification performance further. Secondly, we can explore the utilization of MIN2Net on other EEG measurements, such as SSVEP, MRCPs, and ERP. MIN2Nets might be helpful in the extraction of the most discriminative features for classification. Thirdly, a transfer learning framework based on a fast adaptation procedure will be considered in our future work to thoroughly investigate the possibility of our MIN2Net [8], [10]. Finally, since MIN2Net is developed to learn three modules simultaneously, the loss weight in each module plays a significant role in MIN2Net's learning process. This work has determined an optimal set of loss weights using parameter search, which is time-consuming and results in various optimal sets when different datasets are used. Thus, rather than performing a parameter search for the loss weights, we will explore the adaptive gradient blending concept developed in this work [53] to regulate several loss weights and automatically gain the optimal set of all loss weights throughout the MIN2Net's learning process.

VI. Conclusion

This study proposed MIN2Net, a novel end-to-end multi-task learning, for classifying motor imagery EEG signals. MIN2Net is developed by integrating an autoencoder, deep metric learning, and a supervised classifier, which learns to compress, discriminate embedded EEG and classify EEG simultaneously. We compared the binary classification performance of MIN2Net with four different deep learning algorithms on three benchmark datasets. The classification results revealed that MIN2Net significantly outperformed the developed baselines in the subject-independent settings on SMR-BCI and OpenBMI datasets. Moreover, we obtained promising experimental results from three-class EEG-MI classification (left hand MI vs. right hand MI vs. resting EEG). This finding indicates the possibility and practicality of using this model toward developing real-world applications.

References

[1] D. McFarland and J. Wolpaw, "Eeg-based brain–computer interfaces," *Current Opinion in Biomedical Engineering*, vol. 4, pp. 194–200, 2017, ISSN: 468-4511.

[2] H. Cecotti *et al.*, "Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 11, pp. 2030–2042, 2014.

[3] P. Autthasan *et al.*, "A single-channel consumer-grade eeg device for brain–computer interface: Enhancing detection of ssvep and its amplitude modulation," *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3366–3378, 2020.

[4] Y. Zou *et al.*, "Automatic identification of artifact-related independent components for artifact removal in eeg recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 1, pp. 73–81, 2016.

[5] J. Jeong *et al.*, "Decoding movement-related cortical potentials based on subject-dependent and section-wise spectral filtering," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 3, pp. 687–698, 2020.

[6] R. Chaisaen *et al.*, "Decoding eeg rhythms during action observation, motor imagery, and execution for standing and sitting," *IEEE Sensors Journal*, vol. 20, no. 22, pp. 13 776–13 786, 2020.

[7] G. Pfurtscheller and F. Lopes da Silva, "Event-related eeg/meg synchronization and desynchronization: Basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842 –1857, 1999, ISSN: 1388-2457.

[8] K. K. Ang and C. Guan, "Eeg-based strategies to detect motor imagery for control and rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 4, pp. 392–401, 2017.

[9] A. M. Azab *et al.*, "Weighted transfer learning for improving motor imagery-based brain–computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 7, pp. 1352–1359, 2019.

[10] K. Zhang *et al.*, "Adaptive transfer learning for eeg motor imagery classification with deep convolutional neural network," *Neural Networks*, vol. 136, pp. 1 –10, 2021, ISSN: 0893-6080.

[11] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006, ISSN: 0036-8075.

[12] O. Y. Kwon *et al.*, "Subject-independent brain–computer interfaces based on deep convolutional neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 3839–3852, 2020.

[13] D. J. McFarland and J. R. Wolpaw, "Sensorimotor rhythm-based brain-computer interface (bci): Feature selection by regression improves performance," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 3, pp. 372–379, 2005.

[14] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[15] Z. Jiao *et al.*, "Deep convolutional neural networks for mental load classification based on eeg data," *Pattern Recognition*, vol. 76, pp. 582 –595, 2018, ISSN: 0031-3203.

[16] F. Schroff *et al.*, "Facenet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.

[17] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[18] S. Sakhavi *et al.*, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 11, pp. 5619–5629, 2018.

[19] V. J. Lawhern *et al.*, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056 013, 2018.

[20] T. Wilaiprasitporn *et al.*, "Affective eeg-based person identification using the deep learning approach," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 3, pp. 486–496, 2020.

[21] P. Zhang *et al.*, "Learning spatial–spectral–temporal eeg features with recurrent 3d convolutional neural networks for cross-task mental workload assessment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 1, pp. 31–42, 2019.

[22] A. Gogna *et al.*, "Semi-supervised stacked label consistent autoencoder for reconstruction and analysis of biomedical signals," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2196–2205, 2017.

[23] A. Ditthapron *et al.*, "Universal joint feature extraction for p300 eeg classification using multi-task autoencoder," *IEEE Access*, vol. 7, pp. 68 415–68 428, 2019.

[24] H. Ramoser *et al.*, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.

[25] Kai Keng Ang *et al.*, "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)*, 2008, pp. 2390–2397.

[26] K. K. Ang *et al.*, "Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b," *Frontiers in Neuroscience*, vol. 6, p. 39, 2012, ISSN: 1662-453X.

[27] K. P. Thomas *et al.*, "A new discriminative common spatial pattern method for motor imagery brain–computer interfaces," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 11, pp. 2730–2733, 2009.

[28] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.

[29] Y. Zhang *et al.*, "Improving eeg decoding via clustering-based multitask feature learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–11, 2021.

[30] A. Llera *et al.*, "Adaptive Multiclass Classification for Brain Computer Interfaces," *Neural Computation*, vol. 26, no. 6, pp. 1108–1127, Jun. 2014, ISSN: 0899-7667.

[31] Y. Jiao *et al.*, "Sparse group representation model for motor imagery eeg classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 2, pp. 631–641, 2019.

[32] R. Mane *et al.*, "A multi-view cnn with novel variance layer for motor imagery brain computer interface," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2020, pp. 2950–2953.

[33] M.-H. Lee *et al.*, "EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy," *GigaScience*, vol. 8, no. 5, Jan. 2019, ISSN: 2047-217X.

[34] W. Ge, "Deep metric learning with hierarchical triplet loss," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.

[35] K. Prannay *et al.*, "Supervised contrastive learning," *NeurIPS 2020*, 2020.

[36] F. Schroff *et al.*, "Facenet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823.

[37] W. Chen *et al.*, "Beyond triplet loss: A deep quadruplet network for person re-identification," English, in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, IEEE, 2017, pp. 1320–1329, ISBN: 9781538604588.

[38] R. Thiyagarajan *et al.*, "A learned embedding space for eeg signal clustering," in *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2017, pp. 1–4.

[39] H. Alwasiti *et al.*, "Motor imagery classification for brain computer interface using deep metric learning," *IEEE Access*, vol. 8, pp. 109 949–109 963, 2020.

[40] D. E. Rumelhart *et al.*, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. Cambridge, MA, USA: MIT Press, 1986, 318–362, ISBN: 026268053X.

[41] G. Antoniol and P. Tonella, "Eeg data compression techniques," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 2, pp. 105–114, 1997.

[42] P. Vincent *et al.*, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, 3371–3408, Dec. 2010, ISSN: 1532-4435.

[43] Y. Qiu *et al.*, "Denoising sparse autoencoder-based ictal eeg classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 9, pp. 1717–1726, 2018.

[44] M. Wang *et al.*, "Deep gaussian mixture-hidden markov model for classification of eeg signals," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 4, pp. 278–287, 2018.

[45] M. Tangermann *et al.*, "Review of the bci competition iv," *Frontiers in Neuroscience*, vol. 6, p. 55, 2012, ISSN: 1662-453X.

[46] D. Steyrl *et al.*, "Motor imagery brain-computer interfaces: Random forests vs regularized lda - non-linear beats linear," English, in *Proceedings of the 6th International Brain-Computer Interface Conference Graz 2014*, 2014, pp. 061–1–061–4.

[47] F. Schroff *et al.*, "Facenet: A unified embedding for face recognition and clustering," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[48] A. Gramfort *et al.*, "Meg and eeg data analysis with mne-python," *Frontiers in Neuroscience*, vol. 7, p. 267, 2013, ISSN: 1662-453X.

[49] T. T. Um *et al.*, "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 2017, 216–220, ISBN: 9781450355438.

[50] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.

[51] D. Zhang *et al.*, "Making sense of spatio-temporal preserving representations for eeg-based human intention recognition," *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 3033–3044, 2020.

[52] S. Bhattacharyya *et al.*, "A generic transferable eeg decoder for online detection of error potential in target selection," *Frontiers in Neuroscience*, vol. 11, p. 226, 2017, ISSN: 1662-453X.

[53] H. Phan *et al.*, "Xsleepnet: Multi-view sequential model for automatic sleep staging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.