Learning EEG Representations With Weighted Convolutional Siamese Network: A Large Multi-Session Post-Stroke Rehabilitation Study

Shuailei Zhang, *Member, IEEE*, Kai Keng Ang, *Senior Member, IEEE*, Dezhi Zheng, Qianxin Hui, Xinlei Chen, Yang Li, Ning Tang, Effie Chew, Rosary Yuting Lim, and Cuntai Guan, *Fellow, IEEE*

***Abstract*— Although brain-computer interface (BCI) shows promising prospects to help post-stroke patients recover their motor function, its decoding accuracy is still highly dependent on feature extraction methods. Most current feature extractors in BCI are classification-based methods, yet very few works from literature use metric learning based methods to learn representations for BCI. To circumvent this shortage, we propose a deep metric learning based method, Weighted Convolutional Siamese Network (WCSN) to learn representations from electroencephalogram (EEG) signal. This approach can enhance the decoding accuracy by learning a low dimensional embedding to extract distance-based representations from pair-wise EEG data. To enhance training efficiency and algorithm performance, a temporal-spectral distance weighted sampling method is proposed to select more informative input samples. In addition, an adaptive training strategy is adopted to address the session-to-session non-stationarity by progressively updating the subject-specific model. The proposed method is applied on both upper limb and lower limb neurorehabilitation datasets acquired from 33 stroke patients, with a total of 358 sessions. Results indicate that using k-Nearest Neighbor as the classification algorithm, the proposed method yielded 72.8% and 66.0% accuracies for the two datasets respectively, significantly better than the other state-of-the-arts ($p < 0.05$). Without losing generality, we also evaluated the proposed method on two publicly available datasets acquired from healthy subjects, wherein the proposed algorithm demonstrated superior performance at most cases as well. Our results support, for the first time, the use of a metric learning based feature extractor to learn representations from non-stationary EEG signals for BCI-assisted post-stroke rehabilitation.**

***Index Terms*— Brain–computer interface, EEG, weighted convolutional Siamese network, post-stroke rehabilitation.**

Manuscript received 5 July 2022; revised 14 September 2022; accepted 19 September 2022. Date of publication 26 September 2022; date of current version 20 October 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61873021 and Grant 62088101; in part by the China Scholarship Council under Grant 202006020219; in part by the Special Fund for Basic Scientific Research in Central Colleges and Universities-Youth Talent Support Program of Beihang University; in part the Key Laboratory of Precision Opto-Mechatronics Technology, Beihang University, Ministry of Education, China; in part by the "Next-Generation Brain–Computer-Brain Platform—A Holistic Solution for the Restoration and Enhancement of Brain Functions (NOURISH)" under Project EC-2020-040; and in part by the Agency for Science, Technology and Research (A*STAR) Research, Innovation and Enterprise 2020 (RIE2020) Advanced Manufacturing and Engineering (AME) Programmatic Fund, Singapore, under Grant A20G8b0102. *(Corresponding author: Dezhi Zheng.)*

Shuailei Zhang is with the School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China, also with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 138632, and also with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: zhangshuailei@buaa.edu.cn).

Kai Keng Ang is with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 138632, and also with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: kkang@i2r.a-star.edu.sg).

Dezhi Zheng is with the Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Instrumentation and Opto-electronic Engineering, Research Institute for Frontier Science, Beihang University, Beijing 100191, China, and also with the Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education, Beijing 100191, China (e-mail: zhengdezhi@buaa.edu.cn).

Qianxin Hui is with the School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China (e-mail: huiqx2017307@buaa.edu.cn).

Xinlei Chen is with the Shenzhen International Graduate School, Tsinghua University, Beijing 518057, China, and also with Pengcheng Laboratory, Guangdong 518066, China (e-mail: chen.xinle@sz.tsinghua.edu.cn).

Yang Li is with the Shenzhen International Graduate School, Tsinghua University, Beijing 518057, China (e-mail: yangli@sz.tsinghua.edu.cn).

Ning Tang is with the Division of Rehabilitation Medicine, Department of Medicine, National University Hospital, Singapore 119074 (e-mail: ning_tang@nuhs.edu.sg).

Effie Chew is with the Division of Rehabilitation Medicine, Department of Medicine, National University Hospital, Singapore 119074, and also with the Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119077 (e-mail: effie_chew@nuhs.edu.sg).

Rosary Yuting Lim is with the Institute for Infocomm Research, Agency for Science Technology and Research, Singapore 138632 (e-mail: lim_yuting_rosary@i2r.a-star.edu.sg).

Cuntai Guan is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ctguan@ntu.edu.sg).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2022.3209155, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2022.3209155

I. Introduction

STROKE not only results in a high mortality rate, but its high morbidity also leads up to 50% of survivors being chronically disabled [1]. After a stroke, approximately 80% of all cases suffer motor impairments, and at least 30% will become permanently disabled and hence dependent on others on a daily basis [2]. Therefore, rehabilitation towards motor recovery is imperative for post-stroke treatment [3]. Advances in brain-computer interfaces (BCI) provide a platform to augment Active Motor Training (AMT), a common stroke rehabilitation method, at all stages of stroke recovery. Unlike conventional stroke recovery treatments like AMT, BCI is not dependent on residual motor performance. Instead, it routes a direct communication pathway between the brain and an external device [4] which bypasses the conventional neuromuscular pathways. One approach to accessing the

motor system and realizing BCI-based stroke rehabilitation is Motor imagery (MI). MI is a cognitive process in which the subject performs mental rehearsal of physical movement without actually performing the movement [5]. During MI-assisted rehabilitation, the BCI system first collects the patients' electroencephalograph (EEG) signals, then decodes their voluntary motor intent into commands via signal processing methods. The commands are used to drive a robotic device. In turn, the robotic device moves the patient's paralyzed limb to accomplish rehabilitation exercises. Studies have reported that patients undertaking MI-based rehabilitation show cortical activation in their motor cortex during MI and statistically significant improvements in motor assessment after subsequent rehabilitation [6], [7].

Although motor imagery-based BCI (MI-BCI) is considered a promising technique for stroke rehabilitation [9], most of its clinical applications are still limited. One major limitation is its low decoding accuracy from brain signal to motor intent. The cortical signal of interest is extremely weak and highly susceptible to contamination by both physiological artifacts originating from the subject (e.g., ocular activity, muscle movement, cardiac pulse) and non-physiological artifacts originating from the external environment (e.g., powerline noise, radio, and electrical interference). In particular, BCI for post-stroke rehabilitation may be more vulnerable to interference from physiological artifacts because motor disorders such as tremor is a common complication of stroke [9]. Consequently, most systems are unsuccessful in learning effective representations from the acquired EEG.

Therefore, an efficient feature extractor is essential to drive a BCI system for neurorehabilitation. Current feature extractors in MI-BCI try to remove these artifacts by designing various spectral and spatial filters. One classical spatial filter is Common Spatial Pattern (CSP) [10]. This method separates an EEG signal into additive subcomponents with maximum differences in the variance between two different frequency bands to select components with specific frequency patterns. Our group previously proposed a Filter Bank Common Spatial Pattern (FBCSP) to perform an autonomous selection of key temporal-spatial discriminative EEG characteristics [11]. This approach first bandpass filtered EEG with a filter bank, then optimized the CSP filters for each frequency band. Finally, it extracted the most relevant features via a feature selection algorithm. This algorithm provided an effective approach to combining spatial filtering and frequency filtering together and has been regarded as the de facto standard for motor decoding from EEG recordings [12]. Park *et al.* [13] proposed a local region common spatial pattern which maximized the power/variance ratios of two classes from local CSPs generated from individual channels and their neighbors rather than a global CSP generated from all channels. Gaur *et al.* [14] optimized the temporal filter by designing sliding windows and voting the results. Jin *et al.* [15] proposed a regularized common spatial pattern to enhance the noise robustness and generalization capacity of CSP.

Recently, several studies have proved that deep learning methods are promising for feature extraction and classification in BCI. Owing to their end-to-end strategy and highly nonlinear structure, these methods can learn rich features and make predictions automatically. In most deep learning-based works, spatial and spectral information in EEG was learnt by classification-based approaches. In a representative study, Tibor Schirrmeister *et al.* [12] boosted the EEG decoding accuracy by proposing a Convolutional Neural Network (CNN)-based method. In their work, the EEG data was first preprocessed, then fed into a CNN network with several convolutional architectures to learn representations in higher dimensions. Sakhavi *et al.* [16] developed a MI data classification framework by introducing a new temporal data representation. They first performed the FBCSP method on raw EEG data, then obtained the temporal features by extracting the envelope of each signal using the Hilbert transform, and finally utilized the CNN architecture for classification. More recently, Zuo *et al.* [17] proposed an ensemble learning framework to decompose EEG trials into sub-datasets with different distributions by clustering. Then each sub-data set trained a set of heterogeneous classifiers to generate a diversified classifier search space.

When adopting deep learning for EEG decoding, a considerable amount of literature focus on classification-based representation learning. However, too little attention was given to metric learning to learn EEG representations. Henceforth, our paper will be addressing this knowledge gap by proposing a deep metric learning method named Weighted Convolutional Siamese Network (WCSN) to learn representations from MI EEG. This approach learns an embedding to map data into a representation space where intra-class variations are minimized while inter-class variations are maximized. Owing to its nonlinear structure and end-to-end training strategy, this method is able to learn informative feature representations from EEG data effectively and automatically. To realize this approach, we first design a Siamese Network with two branches of convolution sub-networks to learn the temporal and spatial information from the EEG data, then train it with a contrastive loss function. Conventionally, training the classical Siamese networks is a relatively slow process as models operate on pair-wise input samples. Thus we propose a sampling method to select the informative sample pairs according to their relative temporal-spectral distances to speed up convergence. In addition, a simple but effective offline training strategy was integrated to address the session-to-session non-stationary nature of EEG data.

The main contributions of this paper are as follows:

- A Weighted Convolutional Siamese Network to automatically learn low dimensional embeddings from EEG data in both the temporal and spatial domains. This is the first time a metric learning-based method is proposed to analyze non-invasive EEG for BCI-based post-stroke rehabilitation.
- A novel weighted sampling strategy to select informative EEG samples based on their temporal-spectral distances to speed up convergence and stabilize the training procedure.

- An offline adaptive training strategy to address non-stationarity in EEG data.

The rest of this paper is organized as follows. Section II describes the details of our proposed method. Section III introduces the datasets on which our proposed method was applied, the performance metrics, and the parameter settings in the experiments. Section IV presents the experimental results. Section V discusses further on our work. Finally, section VI presents the conclusions as well as future prospects of our work.

II. Method

A. Problem Formulation

In this study, we formulate MI signal decoding as a multi-class classification problem in which we aim to predict the labels of unannotated data. Suppose we are given a training set consisting of n EEG trials from C classes, which is denoted by $(x_i, c_i)_{i=1}^{n}$, where $x_i \in \mathcal{X}^{D\times T}$ is the i-th trial of EEG signal with D denoting the number of EEG channels, T denoting the number of discrete time points in a selected time window within this trial, and c_i is the class label of x_i. For discrimination, we extract a set of features, denoted by $F = y^{(1)}), y^{(1)}), \ldots, y^{(m)}$. Each feature is a function: $y^{(t)} : \mathcal{X} \to \mathcal{R}$, which maps a sample vector to a scalar. For each sample x, all the m feature values constitute a feature vector, denoted by $y(x) = [y^{(1)}(x), y^{(2)}(x), \ldots, y^{(m)(x)}]^T \in \mathcal{Y}$. For succinctness, we denote the features for the i-th training sample by $y_i = [y^{(1)})_i, y^{(2)})_i, \ldots, y_i^{(m)}]^T$. The goal is to learn a feature embedding $f : \mathcal{X} \to \mathcal{Y}$ that enable a classifier $h(\cdot)$ to predict the label c_i of a given testing sample x_i with a highest classification accuracy. This multi-classification problem can be formulated as an optimization problem as follows:

$$\min_{\theta} E_{(x,c)\in\mathcal{D}}[L(h(f_\theta(x)), c)] \tag{1}$$

where E is the empirical risk, L is the loss function, and θ is the parameter learnt in the feature embedding.

To solve this task, we propose a Weighted Convolutional Siamese Network to learn representations from EEG data. This proposed network consists of four sequential phases: (1) a weighted temporal-spectral distance sampling strategy to select the informative EEG pairs, (2) a weighted Convolutional Siamese Network to learn the temporal-spectral representations from EEG pairs, (3) a k-nearest neighbors (k-NN) based classification to predict the label of testing data by majority voting (see Fig. 1).

B. Weighted Convolutional Siamese Network

The main objective of the Weighted Convolutional Siamese Network is to select informative EEG pairs, map them into a higher dimensional space, and return a high similarity score if they belong to the same category or a low score otherwise. To achieve this objective, (1) a temporal-spectral distance weighted sampling method is first proposed. This method ensures the input EEG pair is spread evenly by sampling them according to their intra-pair temporal-spectral distance, (2) the sampled EEG pairs are fed into two parallel convolutional sub-networks to obtain the temporal-spatial features in the higher dimension, (3) similarity between two temporal-spatial feature vectors are calculated by the Euclidean distance function, and (4) the error between the predicted similarity and the real label is calculated using a contrastive loss function. The proposed deep metric learning phase is formulated as

$$f(\hat{x}_i(t), \hat{x}_j(t)) = L(D(\phi(\hat{x}_i(t)), \phi(\hat{x}_j(t)))), \tag{2}$$

where $\hat{x}_i(t)$ and $\hat{x}_j(t)$ are the pair-wise preprocessed EEG; $\phi(\cdot)$ is the feature embedding to be learned; $D(\cdot)$ is the distance function; and $L(\cdot)$ is the loss layer.

1) *Temporal-spectral Distance Weighted Sampling:* From Wu's work [18], when the data points are distributed on a n-dimensional unit sphere $\mathcal{S}^{n-1}$ with $n > 128$, the distribution of pairwise distances q follows:

$$q(d) \propto d^{n-2}\left[1 - \frac{1}{4}d^2\right]^{\frac{n-3}{2}}, \tag{3}$$

where d is the intra-pair distance. When n is large, $q(d)$ approaches $\mathcal{N}\left(\sqrt{2}, 1/2n\right)$. What can be predicted is that if we sample the EEG trials randomly, we are likely to obtain examples that are $\sqrt{2}-$away. The loss function will produce no error if thresholds are less than $\sqrt{2}$, resulting in no improvement in the training procedure. Therefore, training Siamese networks with randomly selected EEG pairs will be relatively slow and inefficient. To circumvent this, we propose a temporal-spectral distance weighted sampling method to ensure that EEG pairs with any distance stand an equal chance for selection. Before feeding EEG pairs into the Siamese network, they are sampled with weights $q(d)^{-1}$. Kindly refer to the supplementary material for proof that the distances of EEG pairs are evenly distributed after weighted sampling.

Considering the nature of the EEG signal, the value of d should be determined by its characteristics in the time and frequency domains. To calculate the distance of pair-wise EEG in the time domain (called the temporal distance d_t), Euclidean distance is adopted as follows

$$d_t = D\left(\hat{x}_i, \hat{x}_j\right) = \sqrt{\sum_{t=1}^{T}\left(\hat{x}_i(t) - \hat{x}_j(t)\right)^2} \tag{4}$$

where T is the number of time points in each EEG trial. To calculate the distance of pair-wise EEG in the frequency domain (called spectral distance d_f), three steps are operated as follows: Firstly, compute a Fourier transform $G(f)$ of input $\hat{x}_i(t)$ and $\hat{x}_j(t)$ along the time axis respectively. Secondly, the average frequency amplitude $\hat{G}(f)$ is computed along the channel axis.

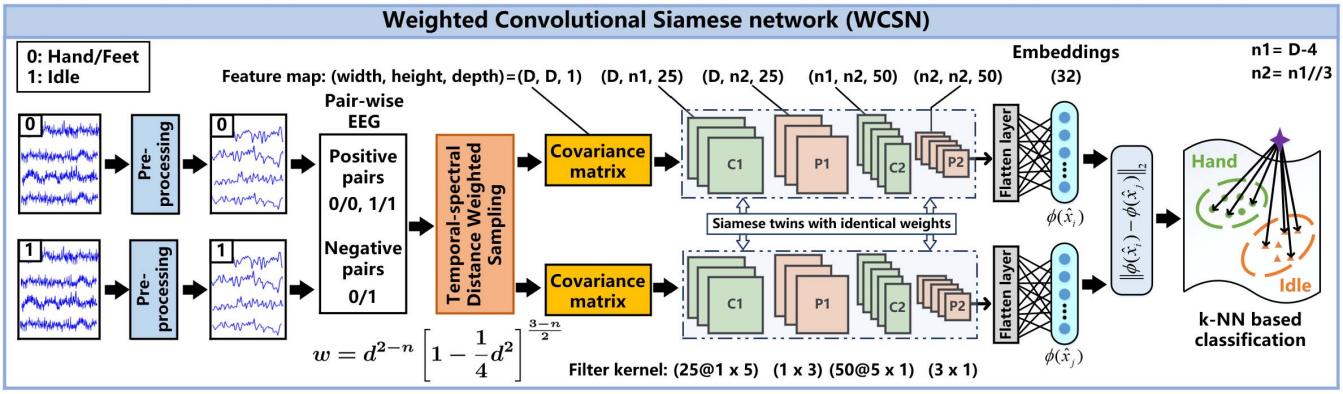

Fig. 1. A schematic diagram of the proposed WCSN framework. This framework includes three main phases: temporal-spectral distance weighted sampling, convolutional Siamese network, and kNN-based classification. In the first phase, the informative EEG pairs are sampled by weighted temporal-spectral distance sampling strategy. In the second phase, a Siamese network sharing identical weights is designed to learn the temporal-spectral representations from EEG pairs. In the final phase, a k-NN predicts the label of testing data by majority voting. "C" is for the convolutional layer, "P" is for the pooling layer, and "D" denotes the number of EEG channels.

Thirdly, the spectral distance d_f between $\hat{x}_i(t)$ and $\hat{x}_j(t)$ is computed by

$$d_f = D\left(\hat{G}_i, \hat{G}_j\right) = \sqrt{\sum_{f=f_L}^{f_H}\left(\hat{G}_i(f) - \hat{G}_j(f)\right)^2} \quad (5)$$

where f_L is the lower cut-off frequency and f_H is the upper cut-off frequency of the bandpass filter in preprocessing phase. After the d_t and d_f are obtained, the temporal-spectral distance d is then computed by

$$d = \sqrt{(d_t)^2 + (d_f)^2}, \quad (6)$$

Finally, the weight of each EEG pair w is computed by

$$w = q(d)^{-1} = d^{2-n}\left[1-\frac{1}{4}d^2\right]^{\frac{3-n}{2}}, \quad (7)$$

The weight serves as the probability of each EEG pair being selected so that EEG pairs with distance centered around $\sqrt{2}$ will have a lower probability of being selected, while others will have greater probabilities. After weighted sampling, the distances of the selected EEG pairs will be evenly distributed. This strategy frees the network from the risk of overfitting the data whose distance centers around $\sqrt{2}$ and allows the network to be more competent in predicting unseen samples.

2) *Learning Temporal-spatial Representations:* A convolutional Siamese network is specifically designed to project the EEG trials onto the higher dimensional space and learn temporal-spatial representations from the EEG trials. Before projecting the EEG data onto the higher dimension, we first calculate the covariance matrix of each selected trial i by:

$$Z = \frac{x_i x_i^T}{tr(x_i x_i^T)}, \quad (8)$$

where $tr(\cdot)$ denotes the trace of a matrix. The covariance matrix provides information that models the interaction between different brain regions, thus capturing their interdependencies during motor imagery. Many feature extractors like FBCSP adopt covariance matrices as representations [11].

Then, a Siamese network with two branches of convolutional sub-networks was designed to learn embeddings from pair-wise EEG data (see Fig. 1). Notably, two sub-networks share the same parameters together to ensure the pair-wise EEG passes through the same embedding function. Each sub-network consists of two convolution blocks, aiming to learn representations in higher dimensions. Each convolution block includes one convolutional layer, one maxpooling layer, one batch normalization layer, and one drop-out layer. The convolutional layer is designed to learn the representations in higher dimensions. The batch normalization layer ensures high learning rates of the model, thereby increasing the training efficiency while reducing the initial weight sensitivity. The max-pooling layer and dropout layer [19] reduce the number of parameters and enhance the generalization capability of the proposed model. The following convolutional block projects the features onto the higher dimension. After the convolution operations, a flatten layer transforms the multi-dimensional matrix of features into a one-dimensional vector. Finally, a dense layer outputs the extracted feature for similarity calculation.

Via the above operations, this structure is able to learn representations with both temporal and spatial information from the multi-channel EEG signal.

3) *Similarity Measurement*: To enhance the training speed, the Euclidean distance function is still employed to measure the similarity S between two temporal-spatial feature vectors:

$$S = D\left(\phi\left(\hat{x}_i\right) - \phi\left(\hat{x}_j\right)\right) = \sqrt{\sum_{m=1}^{M}\left(\phi\left(\hat{x}_i\right)_m - \phi\left(\hat{x}_j\right)_m\right)^2} \quad (9)$$

where $\phi(\hat{x}_i)$ and $\phi(\hat{x}_j)$ represent the temporal-spatial feature embeddings learnt from the previous step, and M is the number of each feature embedding.

4) *Loss Function*: The objective of the loss function is to return a high similarity score if the positive pairs are encoded to similar (closer) representations and negative ones to different (farther) representations. We adopt the contrastive loss function, a widely-used tool for measuring the loss in a Siamese network [20], to evaluate the error between the predicted similarity and the real label. This function ensures that semantically similar examples are in close proximity. The contrastive loss is obtained using:

$$L = \frac{1}{2N}\sum_{i=1}^{N} y_{\text{pair}}\, S^2 + \left(1 - y_{\text{pair}}\right) \max(M - S, 0)^2 \tag{10}$$

where $y_{\text{pair}} = \begin{cases} 1 \text{ if } y_i = y_j \\ 0 \text{ if } y_i \neq y_j \end{cases}$, y_i and y_j are the labels of EEG trial x_i and x_j respectively, N is the number of EEG pairs, and M is a threshold, which is set as 1 in this work.

In one-batch learning, the parameters in two branches of convolution sub-networks are updated by back-propagation and then shared with each other. As the final output, the WCSN returns a low distance value if $x_i(t)$ and $x_j(t)$ belong to the same category and a low score otherwise.

C. k-NN Based Classification

The objective of k-NN based classification is to use the learnt metric to measure the distance between unlabeled test data and labeled training data, and then classify this test data by a plurality vote of its k nearest neighbors. It is a simple but effective classifier in BCI [21]. k-NN is adopted for the following reasons: 1) the output of our proposed WCSN is the distance between EEG trials, which can directly serve as the input of k-NN. 2) k-NN is robust to noise in EEG trials and is easy to implement, which meets the requirement of future online BCI systems. Notably, k-NN is sensitive to high dimensional data because it requires two samples to be proximate in every single dimension. However, this is not a limitation in our work since the representations learnt by our proposed WCSN would have regressed to a single dimension after convolutional operations. In our work, the value of k is optimized by cross-validation.

D. Offline Adaptive Training

The objective of offline adaptive training is to address the session-to-session non-stationarity in the EEG signal [22]. When performing adaptive training, we compute the subject-specific model using EEG data collected from the calibration session and subsequently update the model using the data collected in the evaluation sessions (see Fig. 2).

Let $\mathbf{V}_i$ denote the data used for decoding the EEG data from i-th evaluation session. The initial model $\overline{\mathbf{V}}$ is trained based on the EEG data from the offline calibration session. In the i-th evaluation session ($i > 1$), the label of EEG data collected from the previous $(i-1)$ sessions are readily available. Therefore, these supervised EEG data can be fully utilized to retrain a new model. The offline adaptive training strategy can be formulated as:

$$\mathbf{V}_i = \begin{cases} \overline{\mathbf{V}} & \text{if } i = 1 \\ \overline{\mathbf{V}} \cup \bigcup_{j=1}^{i-1} \mathbf{X}_j & \text{if } i > 1 \end{cases} \tag{11}$$

where $\overline{\mathbf{V}}$ is the data from the calibration session, $\mathbf{X}_j$ is the EEG data from the j-th evaluation session, and $\bigcup$ denotes the union operator.

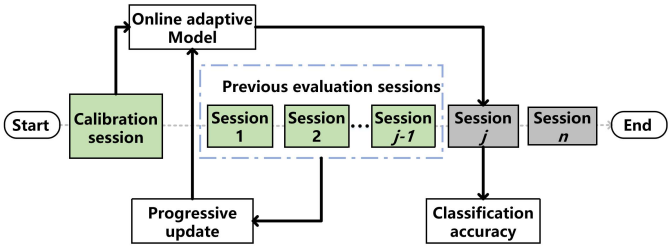

Fig. 2. Offline training strategy addresses the non-stationarity in multi-session neurorehabilitation datasets by using the previous evaluation sessions to update the model.

III. Experiments

A. Data Description

Four datasets (two neurorehabilitation datasets from stroke patients and two publicly available datasets from healthy subjects [5], [23]) are used to assess and compare the proposed WCSN algorithms and baselines. Notably, the two neurorehabilitation datasets were collected from 33 stroke patients with 368 sessions in total. Therefore, our work is a clinical study based on a large population of stroke patients using EEG-based MI BCI. Details for the stroke datasets and other two datasets (BCI Competition IV Dataset 2a [24] and High-gamma dataset [12]) from healthy subjects are reported in the supplementary material.

B. Data Processing

The objective of data preparation is to exclude noise and bad trials from acquired EEG signals and perform normalization before combining them as pair-wise input.

The acquired EEG signal can be extremely weak and contains noise that may bias the analysis [25]. Hence, preprocessing is necessary to filter the noise and normalize the signals. First, a zero-phase Chebyshev Type II filter with a passband of 4 to 30 Hz is employed to remove unwanted frequency noise. We then normalize the signal via the electrode-wise exponential moving average method [12]. After preprocessing, we combine two EEG trials from the same category as a positive pair and label it as "1", while those from different categories are combined as a negative pair and labeled as "0".

C. Implementation Details

Four state-of-arts, FBCSP, EEGNet, ShallowConv, and DeepConv, were selected as baselines to compare with our proposed WCSN. FBCSP is a state-of-the-art feature extraction

TABLE I
CLASSIFICATION ACCURACIES ON THE UPPER-LIMB REHABILITATION DATASET. THE HIGHEST ACCURACIES ARE MARKED IN BOLDFACE. ALL BASELINES ARE REPRODUCED

Method	Avg±SD %	N001	N005	N006	N007	N009	N010	N011	N015	N018
WCSN	**72.8±9.3**	87.3	**66.8**	**82.9**	61.3	**80.9**	63.0	**76.1**	85.8	62.2
FBCSP	70.4±11.5	**91.0**	62.0	81.5	**67.8**	73.3	**66.3**	70.0	84.0	53.0
EEGNet	58.2±4.6	66.8	57.3	56.1	57.1	63.0	52.2	55.3	64.5	56.4
ShallowConv	52.9±3.1	60.1	49.9	53.8	53.4	59.2	49.6	49.0	55.5	52.6
DeepConv	68.4±11.4	88.2	62.7	72.1	54.1	80.1	55.8	68.1	**89.4**	**66.4**
Method	N019	N021	N025	N027	N029	N030	N031	N032	N035	N037
WCSN	**60.5**	**69.9**	**73.9**	**83.9**	**66.0**	**73.6**	75.6	55.4	87.1	70.8
FBCSP	53.0	68.5	73.0	74.3	55.3	69.5	**78.0**	51.0	**89.0**	**77.0**
EEGNet	50.6	57.8	60.5	65.3	55.0	56.4	58.7	53.8	64.8	54.4
ShallowConv	48.9	54.9	53.9	54.9	49.4	50.9	53.3	51.3	50.4	53.9
DeepConv	54.9	58.5	72.9	80.8	62.2	65.9	66.9	**59.1**	85.6	54.5

method in MI-BCI [21] and was the best-performing method in the BCI competition IV-2a as well as in other EEG-decoding competitions [26]. The FBCSP baseline was reproduced following the steps in [11]. EEGNet is another state-of-the-arts to decode EEG signals [27]. It is a discriminant-based classification algorithm and one of the most prevalent deep learning architectures in EEG classification tasks. The EEGNet baseline was reproduced following the steps in [28]. Both ShallowConv and DeepConv are state-of-art CNN based approaches for EEG decoding [12]. Their common structures are one temporal filter along the time axis and one spatial filter along the channel axis to the channel axis. For DeepConv, three additional convolution operations are built to learn representations in higher dimensions. For ShallowConv, it directly makes predictions using a softmax function. Notably, the ShallowConv and DeepConv were proposed with plenty of training strategies like early stopping. To reach a fair comparison, we only adopt the structure of their network.

Offline adaptive training was used to improve the session-to-session transfer accuracy. It subsequently updates a subject-specific model using EEG data collected from the previous evaluation sessions. Take the upper limb dataset as an example, the calibration session comprised of 160 trials, and each evaluation session consisted of 40 trials of EEG data. When performing offline adaptive training on the first evaluation session, only data from the calibration session were used. In the i-th session, $160+40(i-1)$ trials of data were available to train a new model.

All hyperparameters in this work (kindly refer to supplementary material for detail) were empirically chosen during all experiments and were not tuned to the task. To get averaged results, we trained our proposed method and baselines five times before obtaining the average classification accuracies as final results. Classification accuracy (ACC) is selected as the performance metric with the Wilcoxon signed-rank test employed to test if there were significant differences in the performance of algorithms. Results about sensitivity and specificity are reported in the supplementary material.

IV. RESULTS

In this experiment, two large neurorehabilitation datasets were used to evaluate our proposed methods. The upper-limb neurorehabilitation dataset with 190 sessions was acquired from 19 upper-limb stroke-affected patients, while the lower-limb dataset with 168 sessions was recorded from 14 lower-limb stroke-affected patients. To address session-to-session non-stationarity, we adopted the offline adaptive strategy by using data from the previous evaluation sessions to retrain a subject-specific model. Here we mainly report the results on two neurorehabilitation datasets. Please refer to the supplementary material for the results on two datasets from healthy subjects.

A. Upper-Limb Rehabilitation Dataset

In the first experiment, we compared the proposed WCSN algorithm with the FBCSP and EEGNet algorithms on the upper-limb rehabilitation dataset. In this dataset, 19 upper-limb stroke-affected patients went through one calibration session and 10 evaluation sessions. Experiments were repeated five times and then the results were averaged to reach a stable performance. Table I reports the mean classification accuracies averaged across evaluation sessions. Results show that the proposed WCSN algorithm yielded a mean accuracy of 72.8%, whereas the FBCSP, EEGNet, ShallowConv, and DeepConv yielded mean accuracies of 70.4%, 58.2%, 52.9%, and 68.4% respectively. Notably, the proposed WCSN and FBCSP baseline accuracies are higher than the threshold for BCI control (70%) [29]. The advantage of WCSN can be further supported by the observation that the majority of the patients (10/19) were able to operate the BCI with accuracies higher than 70%.

Fig. 3(a) depicts scatter plots of the classification accuracies in each session. Each plotted point indicates the classification accuracy obtained from one of 190 sessions. Points with the same color and mark represent the results of one single patient. The sub-figures compared the performance of FBCSP, EEGNet, ShallowConv, and DeepConv baselines against the proposed WCSN respectively. Points above the diagonal line indicate that the y-axis algorithm outperformed that on the x-axis. The results in Fig. 3(a) showed that the proposed WCSN outperformed FBCSP, EEGNet, ShallowConv, and DeepConv algorithms in 113, 177, 181, and 123 sessions out of a total of 190 sessions. For tests of significance, we employed

TABLE II

CLASSIFICATION ACCURACIES ON THE LOWER-LIMB REHABILITATION DATASET. THE HIGHEST ACCURACY IS MARKED IN BOLDFACE

Method	Avg±SD %	L008	L009	L011	L012	L013	L014	L015	L018	L020	L021	L022	L023	L024	L025
WCSN	**66.0±7.6**	**59.1**	**57.0**	**66.3**	84.2	68.9	**74.5**	**71.6**	**70.6**	**65.3**	61.0	59.5	55.3	**69.7**	61.4
FBCSP	63.0±9.7	53.0	55.2	58.0	**88.0**	**74.3**	68.3	67.8	62.4	53.4	51.8	61.0	55.6	63.8	**69.4**
EEGNet	55.8±3.2	52.0	55.3	58.0	57.5	56.3	62.7	54.9	53.3	54.9	57.4	56.7	56.4	58.2	48.1
ShallowConv	52.6±3.0	52.8	50.4	56.9	50.4	53.9	55.7	48.3	54.5	52.0	51.0	49.1	**57.0**	55.5	48.3
DeepConv	61.9±7.6	57.0	55.5	59.1	81.5	57.8	**74.5**	60.5	56.1	58.3	**64.5**	**67.4**	53.0	60.9	60.8

TABLE III

TESTS OF SIGNIFICANCE ON THE TWO NEUROREHABILITATION DATASETS. "**U**" IS FOR UPPER LIMB REHABILITATION DATASET, "**L**" IS FOR LOWER LIMB REHABILITATION DATASET, "**BCI**" IS FOR BCI COMPETITION IV DATASET 2A, AND "**HGD**" IS FOR HIGH GAMMA DATASET. "***" DENOTES $p < 0.001$, "**" DENOTES $p < 0.01$ AND "*" DENOTES $p < 0.05$

	U	L	BCI	HGD
p-value (WCSN vs. FBCSP)	**	**	0.50	*
p-value (WCSN vs. EEGNet)	***	***	0.05	*
p-value (WCSN vs. ShallowConv)	***	***	1.0	0.21
p-value (WCSN vs. DeepConv)	***	***	*	*

the Wilcoxon test to confirm if our proposed method shows significant superiority over the other computational methods. Results in Table III revealed that the WCSN algorithm performed significantly better than FBCSP ($p = 0.003$), EEGNet, ShallowConv and DeepConv ($p < 0.001$) based on results of all sessions at the 5% level.

B. Lower-Limb Rehabilitation Dataset

Table II shows the classification results of the proposed WCSN as well as the baseline algorithms on the lower-limb rehabilitation dataset. It shows that the proposed WCSN with an average classification accuracy of 66.0%, which is higher than that for FBCSP (63.0%), EEGNet (55.8%), ShallowConv (52.6%) and DeepConv (61.9%). Furthermore, the proposed WCSN yielded the highest classification accuracies in most patients (8 out of 14).

Fig. 3 (b) depicts scatter plots of the classification accuracies in each session. The results showed that the proposed WCSN outperformed FBCSP, EEGNet, ShallowConv, and DeepConv algorithms in 96, 143, 142, and 107 sessions out of a total of 168 sessions respectively. The Wilcoxon test in Table III shows that our proposed method is significantly inferior than FBCSP ($p = 0.004$), EEGNet, ShallowConv and DeepConv ($p < 0.001$).

C. Offline Adaptive Training Strategy

Non-stationarity in EEG can be attributed to fluctuations in the mental states of the subject (fatigue, disengagement, etc.) or technical factors (placement or impedance of the EEG electrodes). These lead to differences in signal quality between the calibration and evaluation sessions, which in turn lead to failure of the classifier [30]. The offline adaptation training strategy aims to address such session-to-session non-stationarity by successively including data from previous

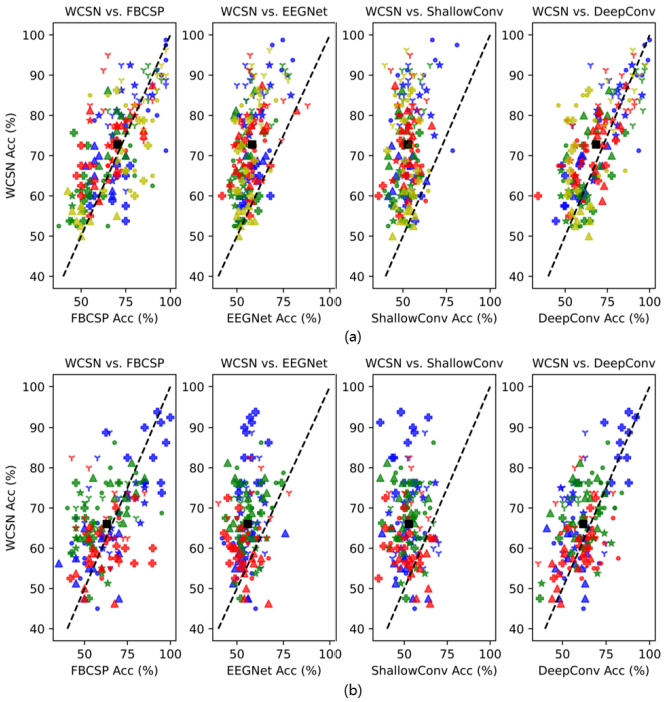

Fig. 3. Comparison of the classification accuracies using scatter plots for (a) the upper-limb dataset and (b) the lower-limb dataset. Each plotted point on the sub-figures indicates the classification accuracy obtained from one evaluation session. The accuracies of the sessions belonging to the same patient were plotted with the same color and mark. The square in black denotes the mean accuracy across all sessions.

sessions into model training. We conducted session-to-session transfer experiments to compare classification results with and without offline adaptive training. In experiments without adaptive training, the data from the calibration session is used to train an initial model before using it to continuously decode testing data in all the evaluation sessions. Fig. 4 shows the comparison results. The left sub-figure in Fig. 4(a) shows results of session-to-session transfer without adaptation on the upper limb dataset, where a trend of deterioration over sessions can be observed. It indicates that session-to-session non-stationarity exists within the neurorehabilitation process. The right sub-figure in Fig. 4(a) shows results using offline adaptive training. It is demonstrated that all methods showed no significant decline in classification accuracies. The same results can be observed in the lower limb dataset in Fig. 4 (b). Interestingly, the proposed WCSN showed higher decoding accuracies in most of sessions for both session-to-session transfer experiments. Fig. 4 (c) shows the algorithm

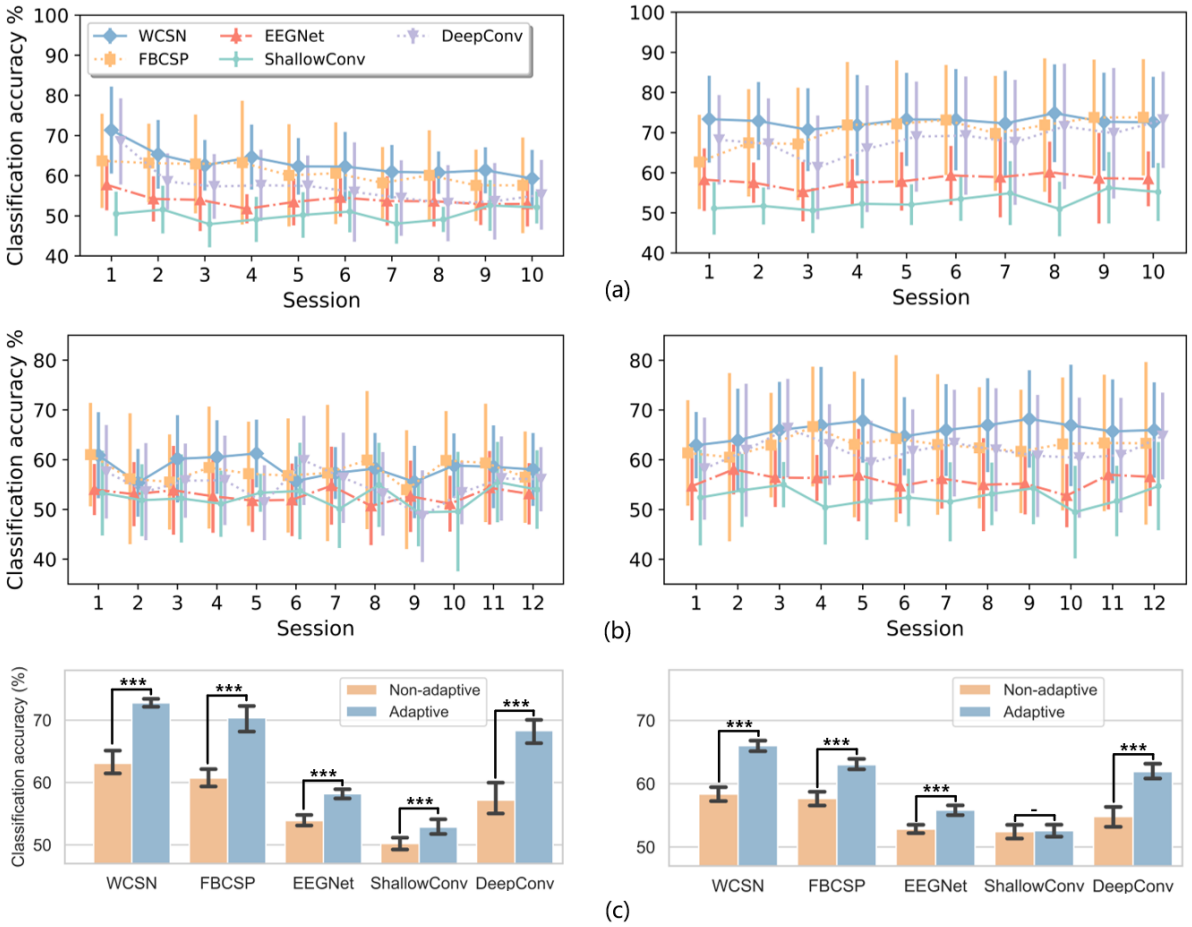

Fig. 4. Comparison between session-to-session transfer with and without offline adaptive training on (a) upper limb rehabilitation dataset and (b) lower limb rehabilitation dataset. The left sub-figures in (a) and (b) are results using the session-to-session transfer strategy without adaptation, and the right sub-figures are results using offline adaptive training. The horizontal axis represents the rehabilitation sessions that patients underwent. Vertical bar plots SDs across patients in each dataset. (c) Comparison between the algorithms with offline adaptive or not on upper limb rehabilitation dataset (left) and lower limb dataset (right). The error bars show the variability across the average sessions. "$***$" denotes $p < 0.001$, "$-$" is for no significance.

performance with or without offline adaptive training on the upper limb rehabilitation dataset (left) and lower limb dataset (right). It can be observed all algorithms show higher decoding accuracy using offline adaptive training strategy, except ShallowConv for the lower limb dataset. This result proves that offline adaptive learning can significantly boost the decoding accuracy of algorithms. The proposed WCSN has lower SDs after using this strategy on both rehabilitation datasets, which further proves the session-to-session non-stationarity is successfully resolved by our offline adaptive training strategy.

V. Discussion

In this article, we have developed a framework based on WCSN to learn representations from EEG. To the best of our knowledge, this is the first time that a deep metric learning-based method is used in the analysis of EEG data acquired from BCI-based post-stroke rehabilitation. Previously, Shahtalebi [31] explored a Siamese Neural Networks for EEG decoding. In this paper, we proved that a common problem associated with the Siamese network is its limited training efficiency as the distance values of most sample pairs center around $\sqrt{2}$. However, our WCSN does not suffer from this problem due to its temporal-spectral distance weighted sampling strategy. This sampling method is proposed to ensure that every sample has an equal chance of selection and to select informative discriminative samples for model training. To better understand the weighted sampling method, we visualized and compared the distribution of temporal distance using the weighted or unweighted sampling method in Fig. 5(a). This figure shows that the distribution of examples in terms of temporal distance using unweighted sampling follows a normal distribution that centers around $\sqrt{2}$, which is consistent with our analysis in Section 2.3. Such a sampling method will inevitably induce no loss and thereby results in a very similar distribution of learned embeddings. Of greater importance is the use of weighted sampling, wherein the distribution of examples falls in the range where all of which would possess the equal chance for selection by the algorithm in both cases, and thus can speed up convergence and stabilize the training procedure. Fig. 5(b) provides more analyses to better demonstrate the impact of our proposed temporal-spectral

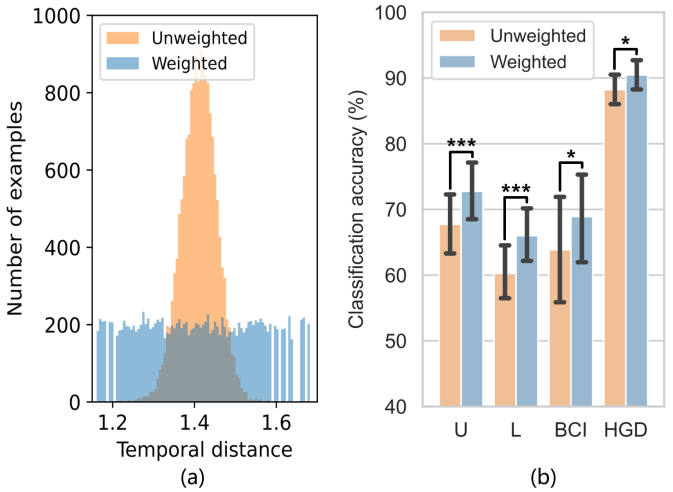

Fig. 5. Comparison between weighted sampling and unweighted sampling. (a) Histogram of the distribution of temporal distance from patient N001 in the upper limb dataset. (b) Comparison on classification accuracy of four datasets. "U": upper limb dataset, "L": lower limb dataset, "BCI": BCI Competition IV Dataset 2a, "HGD": High gamma dataset. "$***$" for $p < 0.001$ in Wilcoxon test, and "$*$" for $p < 0.05$.

distance weighted sampling method on algorithm decoding performance. For two neurorehabilitation datasets, we used classification accuracies averaged across sessions and patients, with error bars indicating SDs across sessions in each dataset. For the BCI Competition IV Dataset 2a and high gamma dataset, we used intra-subject classification accuracies averaged across subjects, with error bars indicating SDs across subjects in each dataset. The observation is that the proposed weighted sampling always achieved significantly higher classification accuracies than that of unweighted sampling in all four benchmark datasets.

Collecting EEG data is a difficult task, especially for stroke patients. Take our lower limb neurorehabilitation dataset for instance, it only contains 160 trials of EEG data for the calibration session and 20 for the evaluation sessions. Therefore, an algorithm that is robust to data size is essential in the BCI-based rehabilitation system. As the Siamese network is suitable in few-shot applications and has been successfully adopted in computer vision [32], industrial systems [33], etc., it could be the rationale behind the better performance demonstrated by our proposed WCSN as compared to existing methods in multi-session stroke patient data.

FBCSP, EEGNet, ShallowConv, and DeepConv, four state-of-arts in MI decoding, were reproduced as the benchmarking baselines in our work. All algorithms were compared against the same ones reproduced in other published works in literature before serving as our baselines. Our results using FBCSP showed slightly lower accuracies than [11] when decoding the BCI competition III dataset IVa (mean accuracy: 90.0% $<$ 90.2%). For EEGNet, our results showed slightly higher accuracies than [34] when decoding the high gamma dataset (median accuracy: 88.3% $>$ 84.7%). Therefore, they can guarantee a fair comparison against the proposed WCSN. For ShallowConv and DeepConv, Schirrmeister [12] proposed them with different training strategies such as cropped training, and early stopping. In our work, only their neural networks were adopted to ensure fair comparisons with other algorithms.

Despite the fact that EEGNet, ShallowConv, and DeepConv are all deep learning-based methods, our method still obtained higher accuracies than them in most cases. One possible reason is that our WCSN has the capability to exclude outliers in EEG by learning distance-based representations, thus it learns more discriminative representations than other algorithms. On the other hand, EEGNet, ShallowConv, and DeepConv face limitations when the number of classes becomes large. Since the number of outputs from the softmax layer needs to match the number of classes, training EEGNet may be more inefficient. Our WCSN addresses this problem by building a twin CNN through which the EEG trials to be compared are passed.

In the two neurorehabilitation datasets, our proposed WCSN attained significantly higher performance accuracies as compared to the four baseline algorithms. However, its mean accuracies were not significantly higher than the recommended accuracy of 70% for BCI control. Such results are still acceptable because BCI for neurorehabilitation is mainly utilized to provide feedback but not for device control that requires high degrees of accuracy. One possible reason for this degraded performance is the inevitable presence of noisy labels in the calibration session. EEG data, especially for those collected from stroke patients, are highly susceptible to interference. In the calibration session, patients are required to perform MI tasks, but they may not be fully engaged during the entire experiment. Yet, the system still labels each EEG trial as programmed. Such unreliable labels may mislead the training of the model and undermine the performance of the algorithm. The noisy label problem is one issue to be fixed in the future.

One future application of our proposed algorithm is zero-shot learning, where the method is expected to predict the label of unseen-class samples via knowledge learnt from other classes [35]. In the case of our WCSN, it learns a metric to determine whether two EEG trials are from the same class or not. If an EEG trial is from another unseen class, our method will be able to determine if it is from a seen class or otherwise. For example, given that our WCSN is trained only from left and right-hand MI data, we can simply use the same model to decode an additional motor imagery task such as feet without the need to train another metric separately.

VI. Conclusion

Recovery of motor function after stroke is crucial for survivors to perform day-to-day activities. In this paper, we proposed a Weighted Convolution Siamese Network to learn EEG representations from multi-session EEG for BCI-assisted post-stroke neurorehabilitation. Owing to its end-to-end structure, this network can learn representations from EEG automatically. The proposed method is evaluated on two large post-stroke clinical datasets with a total of 358 sessions and two publicly available datasets acquired from healthy subjects. Results proved that the proposed method reaches significantly better classification accuracies in most cases. Therefore, our proposed method is a promising tool to learn representations from EEG signal. Future works can be focused on using the channel selection method to increase training efficiency [36] and applying the proposed algorithm to other real-world BCI systems, such as EEG-based real-time robotic control.

ACKNOWLEDGMENT

The two stroke rehabilitation datasets used in this work are acknowledged as: 1. DSRB Ref A/06/286—"Brain Computer Interface (BCI)-Based Robotic Rehabilitation for Stroke," and 2. DSRB Ref D/10/072—"Combined Transcranial Direct Current Stimulation And Motor Imagery-Based Robotic Arm Training for Stroke Rehabilitation—a feasibility study."

REFERENCES

[1] E. S. Donkor, "Stroke in the century: A snapshot of the burden, epidemiology, and quality of life," *Stroke Res. Treatment*, vol. 2018, pp. 1–10, Nov. 2018.

[2] T. Ingwersen *et al.*, "Long-term recovery of upper limb motor function and self-reported health: Results from a multicenter observational study 1 year after discharge from rehabilitation," *Neurol. Res. Pract.*, vol. 3, no. 1, pp. 1–10, Dec. 2021.

[3] G. Bavikatte, G. Subramanian, S. Ashford, R. Allison, and D. Hicklin, "Early identification, intervention and management of post-stroke spasticity: Expert consensus recommendations," *J. Central Nervous Syst. Disease*, vol. 13, Sep. 2021, Art. no. 11795735211036576.

[4] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of EEG-based brain–computer interface paradigms," *J. Neural Eng.*, vol. 16, no. 1, Feb. 2019, Art. no. 011001.

[5] N. Tang, C. Guan, K. K. Ang, K. S. Phua, and E. Chew, "Motor imagery-assisted brain-computer interface for gait retraining in neurorehabilitation in chronic stroke," *Ann. Phys. Rehabil. Med.*, vol. 61, p. e188, Jul. 2018.

[6] V. K. Benzy, A. P. Vinod, R. Subasree, S. Alladi, and K. Raghavendra, "Motor imagery hand movement direction decoding using brain computer interface to aid stroke recovery and rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 3051–3062, Dec. 2020.

[7] P. D. E. Baniqued *et al.*, "Brain–computer interface robotics for hand rehabilitation after stroke: A systematic review," *J. Neuroeng. Rehabil.*, vol. 18, no. 1, pp. 1–25, 2021.

[8] M. A. Khan, R. Das, H. K. Iversen, and S. Puthusserypady, "Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: From designing to application," *Comput. Biol. Med.*, vol. 123, Aug. 2020, Art. no. 103843.

[9] R. Suri *et al.*, "Post-stroke movement disorders: The clinical, neuroanatomic, and demographic portrait of 284 published cases," *J. Stroke Cerebrovascular Diseases*, vol. 27, no. 9, pp. 2388–2397, Sep. 2018.

[10] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2007.

[11] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 2390–2397.

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

[13] Y. Park and W. Chung, "Frequency-optimized local region common spatial pattern approach for motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1378–1388, Jul. 2019.

[14] P. Gaur, H. Gupta, A. Chowdhury, K. McCreadie, R. B. Pachori, and H. Wang, "A sliding window common spatial pattern for enhancing motor imagery classification in EEG-BCI," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, 2021.

[15] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Netw.*, vol. 118, pp. 262–270, Oct. 2019.

[16] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.

[17] C. Zuo *et al.*, "Cluster decomposing and multi-objective optimization based-ensemble learning framework for motor imagery-based brain–computer interfaces," *J. Neural Eng.*, vol. 18, no. 2, Apr. 2021, Art. no. 026018.

[18] C.-Y. Wu, R. Manmatha, A. J. Smola, and P. Krahenbuhl, "Sampling matters in deep embedding learning," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2840–2848.

[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[20] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jul. 2006, pp. 1735–1742.

[21] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 031005.

[22] K. K. Ang and C. Guan, "EEG-based strategies to detect motor imagery for control and rehabilitation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 4, pp. 392–401, Apr. 2016.

[23] K. K. Ang *et al.*, "Facilitating effects of transcranial direct current stimulation on motor imagery brain-computer interface with robotic feedback for stroke rehabilitation," *Arch. Phys. Med. Rehabil.*, vol. 96, no. 3, pp. S79–S87, Mar. 2015.

[24] G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.

[25] R. Na *et al.*, "A wearable low-power collaborative sensing system for high-quality SSVEP-BCI signal acquisition," *IEEE Internet Things J.*, vol. 9, no. 10, pp. 7273–7285, May 2021.

[26] M. Tangermann *et al.*, "Review of the BCI competition IV," *Frontiers Neurosci.*, vol. 6, no. 1, p. 55, 2012.

[27] J. Cui, Z. Lan, O. Sourina, and W. Müller-Wittig, "EEG-based cross-subject driver drowsiness recognition with an interpretable convolutional neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Feb. 16, 2022, doi: 10.1109/TNNLS.2022.3147208.

[28] V. Lawhern, A. Solon, N. Waytowich, S. M. Gordon, C. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Jul. 2018, Art. no. 056013.

[29] Y. Shahriari, T. M. Vaughan, L. M. McCane, B. Z. Allison, J. R. Wolpaw, and D. J. Krusienski, "An exploration of BCI performance variations in people with amyotrophic lateral sclerosis using longitudinal EEG data," *J. Neural Eng.*, vol. 16, no. 5, Sep. 2019, Art. no. 056031.

[30] J. Giles, K. K. Ang, K. S. Phua, and M. Arvaneh, "A transfer learning algorithm to reduce brain-computer interface calibration time for long-term users," *Frontiers Neuroergonomics*, vol. 3, pp. 1–14, Apr. 2022.

[31] S. Shahtalebi, A. Asif, and A. Mohammadi, "Siamese neural networks for EEG-based brain-computer interfaces," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 442–446.

[32] G. Koch *et al.*, "Siamese neural networks for one-shot image recognition," in *Proc. ICML Deep Learn. Workshop*, Lille, France, vol. 2, 2015, pp. 1–30.

[33] X. Zhou, W. Liang, S. Shimizu, J. Ma, and Q. Jin, "Siamese neural network based few-shot learning for anomaly detection in industrial cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5790–5798, Aug. 2020.

[34] F. A. Heilmeyer, R. T. Schirrmeister, L. D. Fiederer, M. Volker, J. Behncke, and T. and Ball, "A large-scale evaluation framework for eeg deep learning architectures," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 1039–1045.

[35] B. Liu, L. Hu, Q. Dong, and Z. Hu, "An iterative co-training transductive framework for zero shot learning," *IEEE Trans. Image Process.*, vol. 30, pp. 6943–6956, 2021.

[36] J. Jin, C. Liu, I. M. Y. Daly, S. Li, X. Wang, and A. Cichocki, "Bispectrum-based channel selection for motor imagery based brain-computer interfacing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2153–2163, Oct. 2020.