Few-Shot Relation Learning with Attention for EEG-based Motor Imagery Classification

Sion An[1], Soopil Kim[1], Philip Chikontwe[1] and Sang Hyun Park[1]

Abstract— Brain-Computer Interfaces (BCI) based on Electroencephalography (EEG) signals, in particular motor imagery (MI) data have received a lot of attention and show the potential towards the design of key technologies both in healthcare and other industries. MI data is generated when a subject imagines movement of limbs and can be used to aid rehabilitation as well as in autonomous driving scenarios. Thus, classification of MI signals is vital for EEG-based BCI systems. Recently, MI EEG classification techniques using deep learning have shown improved performance over conventional techniques. However, due to inter-subject variability, the scarcity of unseen subject data, and low signal-to-noise ratio, extracting robust features and improving accuracy is still challenging. In this context, we propose a novel two-way few shot network that is able to efficiently learn how to learn representative features of unseen subject categories and how to classify them with limited MI EEG data. The pipeline includes an embedding module that learns feature representations from a set of samples, an attention mechanism for key signal feature discovery, and a relation module for final classification based on relation scores between a support set and a query signal. In addition to the unified learning of feature similarity and a few shot classifier, our method leads to emphasize informative features in support data relevant to the query data, which generalizes better on unseen subjects. For evaluation, we used the BCI competition IV 2b dataset and achieved an 9.3% accuracy improvement in the 20-shot classification task with state-of-the-art performance. Experimental results demonstrate the effectiveness of employing attention and the overall generality of our method.

I. INTRODUCTION

Electroencephalography (EEG) is an easy and cheap technique for recording brain activity commonly used for brain-computer interfaces (BCI), enabling communication between the brain and external devices. The high temporal resolution of EEG signals is a key characteristic that makes them useful for research and diagnosis related to brain disorders. Many studies such as robotic arm control [1], [2], wheelchair control [3], [4], autonomous driving [5] have been conducted by capturing human brain signals through the EEG. Among several types of EEG, motor imagery (MI) signals have attracted attention and are flexible for discriminating brain activity. MI is produced as a response to thinking tasks *i.e.*, intention to perform hand or leg movements. Accordingly, automated MI classification has been addressed using machine learning [6], [7], [8], [9], [10], [11], [12], [13] as well as deep learning techniques [14], [15], [16], [17], [18], [19], [20]. However, the low signal-to-noise ratio (SNR), limited

spatial resolution and complex dynamics of MI present challenges for the accurate classification and analysis of EEG data. Moreover, robust deep learning models require large amounts of labeled data to be effective given large patient-to-patient signal variations for different behaviors.

To address these limitations, we propose to model EEG-based MI classification as a few shot classification task. Few-shot classification addresses the scenario where a classifier must adapt to accommodate classes which are not seen during training given only a few labeled samples per class (or subject). Thus, we propose a model that learns how to learn representative features of unseen subject categories and be able to accurately classify them using paired samples from support and query data obtained from other subjects. The proposed method consists of feature embedding, attention and relation modules that are all connected in an end-to-end framework. In our framework, relevant features from the support and query data are obtained via the embedding module. Then, class representative features are generated via an attention module that discovers key features related to the query signal among the support data in an episode. Lastly, a relation module predicts the label of a query sample in the embedding space. Our method is able to model the relationship between several paired samples from the support and query sets among different subjects during training; further enabling generalization to classify the query signal of an unseen subject in testing even with few labeled samples.

The main contributions of this work are as follow: (1) In this study, we empirically show the benefit of few-shot learning applied to EEG-based MI classification with improvements in performance. To the best of our knowledge, this work is the first to show the application of few-shot learning in EEG-based MI classification. (2) We introduce a novel few-shot attention technique that can emphasize important signals among the support set to predict the label of a query signal. (3) We integrate an appropriate 1D convolutional neural network (CNN) in the few-shot learning embedding module, which divides a signal into three signals with different frequency ranges and adopts varying kernel sizes. The proposed method can be applied to not only EEG but also various bio-signal N-way K-shot classification problems.

II. RELATED WORKS

Conventional methods: Several methods consisting of feature extraction and classification steps have been proposed for EEG classification. For example, [6] extracted features using short time fourier transform (STFT). [7], [8]

extracted features representing time-frequency information using wavelet packet decomposition (WPD). [9], [10], [11], [12], [13] decomposed the EEG signals into spatial patterns and extracted features using common spatial pattern (CSP). Given the extracted features, classification was performed using machine learning models such as KNN classifier, support vector machine (SVM), XGBoost, or random forest. These methods required heuristic parameter setting such as defining frequency bands, and thus showed limited performances since the optimal parameters were different for each task or subject.

Deep learning based methods: Deep learning based MI EEG classification methods have shown high performance in literature. [16] applied 1D CNNs to extract multi-level features and predict the label. In contrast, [14] transformed a time domain signal into frequency domain using STFT, and then applied CNN on the frequency intensity. [15] proposed a network which contains both CNN and long-term short-term memory model (LSTM) to handle sequential time domain data. In a follow-up study, [19] proposed a CNN with hybrid convolution scale (HS-CNN) which separates a signal into three signals using bandpass filters with 4~7Hz, 8~13Hz, 13~32Hz frequency bands, and feed them into convolution layers with different filter sizes. The features including different semantic information were concatenated and then MI classification was performed. Further, as an auxiliary, data augmentation [15], [18], [19] or attention mechanism [21], [22] were also proposed. [21] applied the attention module to LSTM to utilize long range information for EEG-based hand movement classification. [22] used attention modules to focus on important part of continuous signal as well as to find an important trial among the whole trials for EEG-based emotion classification.

Despite the improved performance over the conventional methods, the deep learning methods often fail when training samples per subject are limited. Thus, a huge number of training samples need to be obtained from each target subject to train the robust model. To overcome these drawbacks, we propose a relation based few-shot learning method under a single framework that is able to accurately predict the label of query signal using only a few labeled samples.

Few shot learning: To the best our knowledge, literature using few-shot learning methods for EEG-based MI classification is limited. Although not applied for EEG-based MI classification, the work of [23] extracted features from intracranial electroencephalography (iEEG) signals using local binary patterns and learned prototype vectors representing a class in a hyper-dimensional space. Then, they performed classification based on the distance between prototype vectors and a vector from query signal. Although they considered the one-shot learning problem, this method did not provide an end-to end learning framework. [24] recently proposed a deep neural network with triplet loss to address the classification of various time series data including ECG recordings. This method demonstrates the possibility that few-shot learning can work effectively for signal analysis problems; however, they only learn feature embedding networks so that embedded features can be categorized by a fixed nearest neighborhood classifier.

In this paper, we propose a novel relation based few-shot learning method that can learn a non-linear comparator by learning end-to-end relationships between the support data distribution and query signal without calculating the distance between embedding vectors. The structure of the proposed network closely resembles the recent network proposed for image classification [25]. However, unlike the method of [25] that uses a conventional CNN for feature extraction and simply averages the embedding vectors for relation computation, we integrate a modified HS-CNN that is suitable for 1D signal analysis in the embedding module as well as incorporate an attention module that enables the emphasis of important signals in the supporting set.

III. METHOD

Our proposed 2-way K-shot learning framework is shown in Fig. 1. The framework consists of three modules: an embedding module, an attention module, and a relation module. The embedding module extracts semantic features for classification from input EEG signals. Given the extracted features, attention module predicts an attention score for each support sample using both support and query features. A class-representative vector is calculated for each class using a weighted average of k support features based on the predicted attention scores. Finally, the relation module predicts relation scores given the class-representative vectors and query features. The class with the largest relation score is predicted as the label of the query. During training, the proposed network is trained end-to-end using pairs of a support set and a query signal from subjects in training data. For testing, we predict the label of a query signal using k labeled support signals from an unseen subject. Though the figure shows an example of 2-way K-shot learning, the proposed method can be extended to N-way K-shot learning.

A. Embedding Module

The embedding module $F(\cdot)$ takes a pair of k support and a query data as input and extracts semantic features for classification using convolutional layers. To extract meaningful features from a noisy one-dimensional signal, we modified HS-CNN [19] which separates the data into 3 different frequency bands and extract features per band instead of using convolution layers directly on the data. Specifically, the data x obtained from C3, CZ, and C4 electrodes are divided into 4~7Hz, 8~13Hz, 13~32Hz frequency band signals x_b (b = 4~7, 8~13, 13~32) using a 4th order butterworth bandpass filter. For each x_b, features were extracted using three convolution layers with different kernel sizes (*i.e.*, 45×1, 65×1, and 85×1). The three features obtained are concatenated and passed through another convolution layer with kernel size of 1×3 to extract higher-level features between the C3, CZ and C4 electrodes. Then, the features are passed through a max pooling layer with a kernel size of 6×1 with a stride 6×1. Finally, the embedding feature
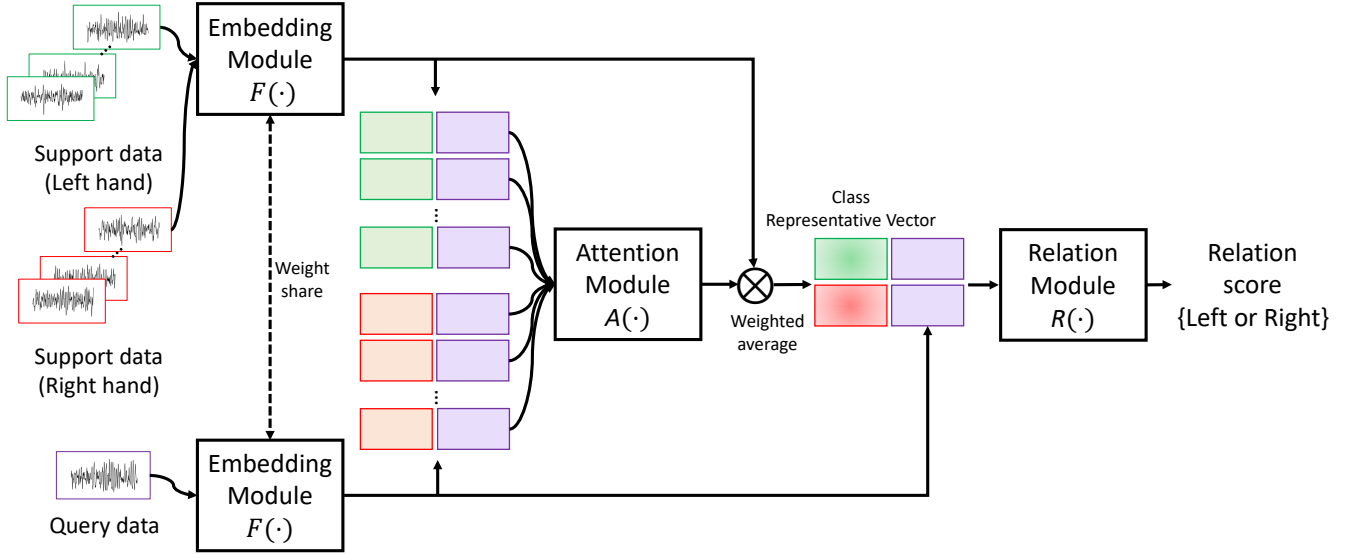
Fig. 1. Overview of the proposed few-shot relation network. Our framework has three key modules: embedding, attention and a relation module. Given support and query data, we obtain feature vectors concatenated at several stages in our pipeline via an embedding module $F(\cdot)$ which are feed to an attention module $A(\cdot)$ that focuses key features related to the query data based on attention scores. Attended features are further feed to a relation module $R(\cdot)$ that predicts relation scores from representative vectors and query features to obtain the label of the given query sample. Based on the predictions, we update the weights of the entire framework in a single training episode.
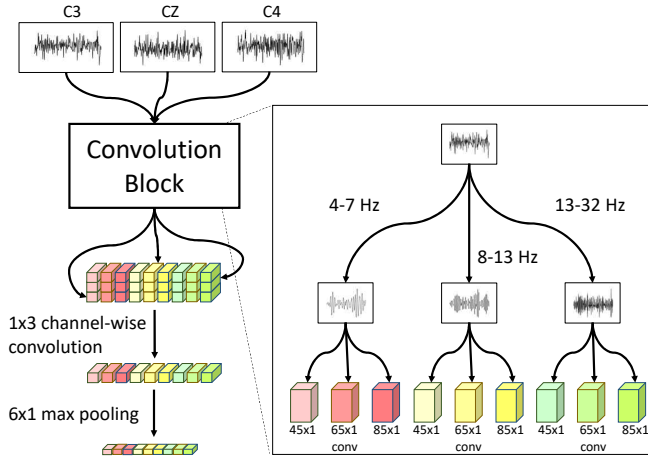

Fig. 2. Illustration of the proposed embedding module $F(\cdot)$.

$z = F(x)$ is obtained by concatenating all features. Fig. 2 shows the embedding network.

B. Attention Module

The support set of N-way k-shot learning contains $N \times k$ data samples. Given extracted features for each data sample from the embedding module, the attention module $A(\cdot)$ computes attention scores to get N class-representative features most related to the query prediction. In the original relation network [25], class representative features are defined as a summation of extracted features from support data. However, this approach does not effectively utilize the data for query prediction since all support data samples are reflected equally in the class-representative feature. Though the approach is an easy example that may explain why a given query set belongs to a certain class, it may lead to inaccurate prediction if noisy

samples exist in the support set.

To address this, we hypothesize that an attention-based approach may better help our model to focus on key support samples. Thus, we concatenate the features of each support feature $z_{i,j}^S (i = 1, 2, , N; j = 1, 2, , k)$ and query feature z^Q in the channel direction and use the feature $z_{i,j}^{SQ} = z_{i,j}^S \oplus z^Q$ as input for the attention module to predict the attention score $a_{i,j}^S = A(z_{i,j}^{SQ})$, where $\oplus$ is the concatenation operation. The attention module consists of convolutional layers using 16×1 and 4×1 kernels, a global average pooling layer, and 64 and 1-dimensional fully connected layers to consider both global and local features.

For each class, the attention-weighted feature $\overline{z_i^S}$ is computed as:

$$\overline{z_i^S} = \frac{\sum_{j=1}^{k} a_{i,j}^S * z_{i,j}^S}{\sum_{j=1}^{k} a_{i,j}^S}. \tag{1}$$

Then, $\overline{z_i^S}$ and z^Q are concatenated to obtain the final class-representative feature vector $z_i^{SQ} = \overline{z_i^S} \oplus z^Q$ for each class. Finally, the class-representative features are fed to the relation module to predict the label.

C. Relation Module

While several few-shot learning methods predict the label based on a distance metric between class-representative features and a query feature, we estimate the label by learning a relation network to distinguish them. The relation module $R(\cdot)$ predicts a relation score $r_i^S = R(z_i^{SQ})$ from z_i^{SQ}. Finally, the class with largest predicted relation score is taken as a label. To estimate the label from the 1D feature vector, the relation module uses two convolutional layers with 30×1 and 15×1 kernels, a global average pooing layer, and two fully connected layers with dimensions 256 and 100.

D. Implementation details

For training, we optimized the model using cross-entropy loss function as follows:

$$Loss = -\frac{1}{N} * \sum_{i=1}^{N} y_i * log(r_i^S), \quad (2)$$

where y_i is 1 if the class of supporting data and that of query data are the same, otherwise 0. Adam optimizer [26] was used to optimize the parameters. The batch size was set to 100 and the initial learning rate was 10^{-4}. On every iteration, learning rate decayed with an exponential decay rate of 0.033%. The model was saved when validation loss was minimum.

IV. EXPERIMENTAL RESULTS

A. Dataset

The BCI competition IV 2b dataset [27] was used to evaluate the performance of proposed network. The dataset contains raw EEG data of 9 subjects for MI classification with approximately 120 trials per experiment, where each subject imagines left- or right-hand movement according to instruction. Five experiments were conducted on each subject with a total of 45 experiments collected from 9 subjects. The collected samples were measured at a sampling frequency of 250 Hz on three electrodes C3, CZ, and C4 in accordance with the protocol of International 10-20 system. Among the 5400 trials (*i.e.*, 120×45), we ignored some rejected trials and then used signals from 3.5s to 7s of the remaining trials in this study. From 3 electrodes, we obtained three 875 values and stacked them to form a 875×3 matrix which was used as input to the model.

B. Experimental Setting

For our experiments, 9-fold cross validation was used with the model being trained with 8 subjects data samples and tested with the remainder per validation episode. Each training set consisted of the first four experiments per subject, with the fifth experiment used as the validation set. The test set consisted of the full set of the remaining subjects. During training, support and query data samples were randomly selected from both the training and validation set at each iteration. For testing, we split whole samples into two groups in each class, 20 samples as support and the rest as query data.

To assess the performance of the proposed model (**RelationNet-attention**), we evaluated the classification accuracy of $K = \{1,5,10,20\}$ shot classification models. For $K = \{1,5,10\}$ shot experiments, K support samples were randomly selected among the 20 support samples. To assess the impact of the attention, we also evaluated the performance of 5, 10 and 20-shot models without attention (**RelationNet**). In addition, we compare our method with following two supervised learning models: (1) HS-CNN [19] trained with only 40 samples (20 left and 20 right samples) in the support set of testing subjects (**HS-CNN-Few**) and (2) HS-CNN

	Method	Average ± Std
-	HS-CNN-Few	56.0 ± 3.7%
-	HS-CNN-All	64.5 ± 8.2%
1 shot	RelationNet	66.2 ± 9.5%
5 shot	RelationNet	69.4 ± 11.3%
	RelationNet-attention	71.0 ± 10.5%
10 shot	RelationNet	70.9 ± 10.9%
	RelationNet-attention	72.6 ± 11.7%
20 shot	RelationNet	73.1 ± 10.5%
	RelationNet-attention	74.6 ± 10.2%

trained with all training samples from 8 subjects (**HS-CNN-All**). Since the results may vary as the support set changes, the same experiment was repeated 10 times and the average was used as the final accuracy for all cases. The accuracy was measured based on the number of true positive and true negative over the entire dataset.

C. Results and Analysis

Table 1 shows the overall accuracy with standard deviations of the proposed methods and other methods. Fig. 3 shows the accuracy of the methods for 9 folds as a graph. In all cases, our models achieve better performance than the prior approaches. Specifically, supervised learning-based method reports that classification performance reaches 87.6% when sufficient training data acquired from the target subject is used [19], but the performance significantly drops as 65.3% when the model was trained in a cross-subject setting. When the model was trained with a small amount of data (20 samples per class) acquired from the target subject, the accuracy was also limited. On the other hand, the proposed few shot learning technique greatly improved the performance. In the one-shot setting with only one labeled sample per target subject used for training, a 1.7% improvement in performance was observed compared to HS-CNN-All, 4.9% in 5 shots, 6.4% in 10 shots, and 8.6% in 20 shots. Finally, our approach achieved 74.6% (+9.3%) mean accuracy in 20-shot with attention setting.

The proposed attention module improved the performance consistently. In particular, an average and consistent improvement of 1.6% can be observed whenever attention is employed in all cases (*i.e.*, the average percent improvement over all the proposed methods when $K = \{5,10,20\}$). Moreover, by using the attention module, we can infer that the model concentrates on the support set feature representations that are important for query prediction, while at the same time suppress noisy signal features to achieve higher performance.

We further observed that in the 2nd and 3rd folds, accuracy scores of both HS-CNN-All and the proposed model were relatively poor (See Fig. 3). These results could be an indication of the nature of EEG data; due to large SNR, it was hard to classify the query data with only a few support sets. We believe significant performance improvements may
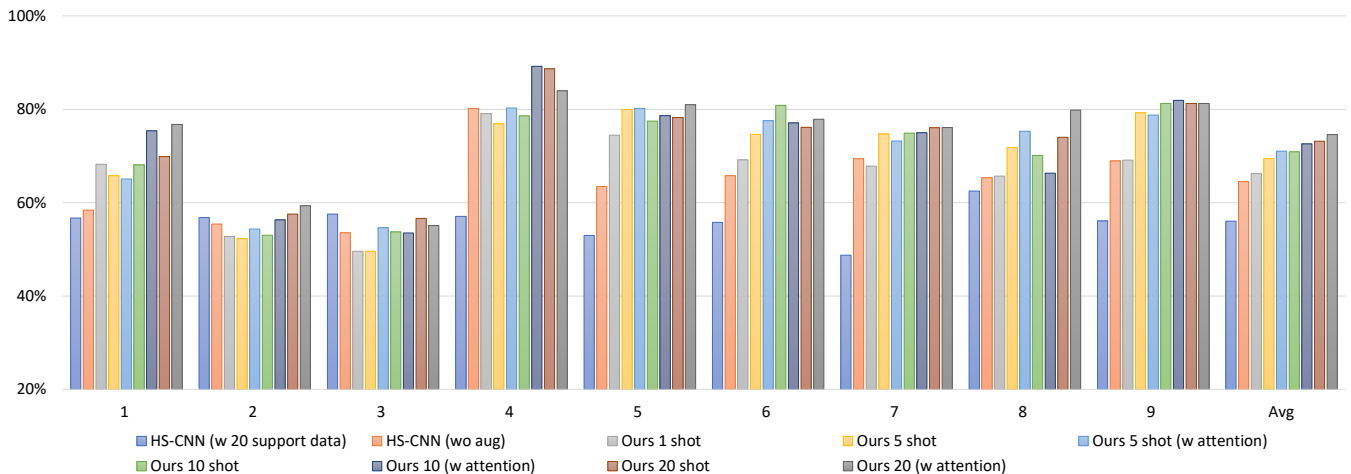
Fig. 3. Comparison of our few shot learning methods with related methods across 9 validation folds.

be observed if noise in the data is sufficiently reduced. For a more effective comparison, the average prediction accuracy, except for 2nd and 3rd folds, was calculated as 67.3% for HS-CNN-All and 79.5% for 20-shot RelationNet-attention. In particular, our few shot learning method showed comparable results in all folds except 2nd and 3rd as the std decreased, and also showed a 12.2% performance improvement over HS-CNN-All.

D. Effect of data augmentation in the few shot setting

Dai *et al.* [19] report that HS-CNN performance improved when data augmentation was used for training and testing within same subject data. To show the effect of data augmentation, we tested the HS-CNN trained by all training samples with data augmentation. Furthermore, we performed the data augmentation on our few shot learning models (both with and without attetion module) and then confirmed the results. Data augmentation was applied to the training data samples in two stages: (1) time-domain recombination: crop and recombine trials to form new trails, and (2) frequency-domain swapping: band-pass filter trials and generate new trials. The later were repeated multiple times to create increased number of samples (we refer the reader to the [19] for more details).

Table 2 shows the mean accuracy and standard deviation. In our study, the improvements gained by applying data augmentation to base method HS-CNN-All were not noteworthy. Furthermore, augmentation was equally ineffective when used in the proposed few shot framework since the quality of augmented samples generated by K images hardly reflected the EEG signals from unseen subjects. As a result, accuracy decreased in most cases when the model was trained with augmented data. On the other hand, interestingly, accuracy increased when the attention module was used even though the accuracy was slightly less than the accuracy of model without data augmentation. Overall, where the base method fails in generality, the proposed method shows consistent improvements across K-shot settings and subjects. This result further highlights utility of the proposed attention module

TABLE II
COMPARISON OF ACCURACY SCORES OF THE PROPOSED AND RELATED
METHODS TRAINED WITH DATA AUGMENTATION.

	Method with augmentation	Average $\pm$ Std
-	HS-CNN-All	65.3 $\pm$ 9.2%
5 shot	RelationNet	68.6 $\pm$ 9.0%
	RelationNet-attention	69.3 $\pm$ 10.9%
10 shot	RelationNet	71.7 $\pm$ 10.6%
	RelationNet-attention	72.8 $\pm$ 12.5%
20 shot	RelationNet	72.2 $\pm$ 10.2%
	RelationNet-attention	73.6 $\pm$ 12.3%

in emphasizing important support set features important for query prediction.

E. Validation on external dataset

To assess the generalization performance of our model, we evaluated performance on an independent external dataset. In this experiment, we tested our models on the BCI competition IV 2a dataset [28]. Note that the models trained with BCI competition IV 2b dataset were directly used for testing without re-training our models. The dataset contains a total of 288 trials with four classes (left hand, right hand, both feet, and tongue movement) for a single experiment. EEG data was measured at a sampling frequency of 250 Hz across a total of 22 electrodes, depending on the protocol of the International 10-20 system. In this study, we used the left- and right-hand trial data of C3, CZ and C4 electrodes in the same location and used 875 values as inputs from 2.5s to 6s of 1163 trials except the rejected trial in the training set.

The accuracy of the HS-CNN-All was almost 53%, while the proposed method achieved around 60%, showing a 6.0% improvement. Although the results are lower than those reported for subjects of the same dataset (*i.e.*, BCI competition IV 2b dataset [27]), we could confirm that the proposed few shot methods show comparable generalization even for tasks with slightly different settings without re-training. In addition, the model with attention module also consistently showed better results than the relation network

TABLE III

COMPARISON OF ACCURACY SCORES OF THE PROPOSED AND RELATED

METHODS ON EXTERNAL DATASET (BCI IV 2A [28]).

	Method	Average ± Std
-	HS-CNN-All	53.1 ± 2.0%
1 shot	RelationNet-attention	52.9 ± 6.6%
5 shot	RelationNet	56.3 ± 9.8%
5 shot	RelationNet-attention	55.9 ± 8.3%
10 shot	RelationNet	57.5 ± 9.3%
10 shot	RelationNet-attention	58.7 ± 11.4%
20 shot	RelationNet	58.4 ± 10.4%
20 shot	RelationNet-attention	59.1 ± 11.1%

in this experiment.

V. CONCLUSIONS

In this study, we have proposed a two way few shot classification network which performs well on unseen subjects by applying the concept of few shot learning to the EEG-based MI classification problem. Our model achieves improved performance over baseline methods and gives evidence that few shot learning is able to classify the EEG data of unseen subject even with limited samples. Moreover, we show that using attention can further improves the performance by emphasizing the information of support data that is highly correlated to query data. Empirical results show the general applicability of the proposed method across K-shot settings with a novel formulation of relation score as a similarity measure between support and query sets. The proposed model can be easily applied other types of time series data which suffer from large signal variations between subjects. Overall, the simplicity and effectiveness of this approach makes it a promising approach for EEG based MI classification.

REFERENCES

[1] J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, and B. He, "Noninvasive electroencephalogram based control of a robotic arm for reach and grasp tasks," *Scientific Reports*, vol. 6, p. 38565, 2016.

[2] A. Ramos-Murguialday, M. Schürholz, V. Caggiano, M. Wildgruber, A. Caria, E. M. Hammer, S. Halder, and N. Birbaumer, "Proprioceptive feedback and brain computer interface (BCI) based neuroprostheses," *PloS one*, vol. 7, no. 10, 2012.

[3] D. W.-K. Ng, Y.-W. Soh, and S.-Y. Goh, "Development of an autonomous BCI wheelchair," in *2014 IEEE Symposium on Computational Intelligence in Brain Computer Interfaces (CIBCI)*, 2014, pp. 1–4.

[4] F. Galán, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, and J. d. R. Millán, "A brain-actuated wheelchair: asynchronous and non-invasive brain–computer interfaces for continuous control of robots," *Clinical neurophysiology*, vol. 119, no. 9, pp. 2159–2169, 2008.

[5] L. Yang, R. Ma, H. M. Zhang, W. Guan, and S. Jiang, "Driving behavior recognition using EEG data from a simulated car-following experiment," *Accident Analysis & Prevention*, vol. 116, pp. 30–40, 2018.

[6] S. K. Bashar and M. I. H. Bhuiyan, "Classification of motor imagery movements using multivariate empirical mode decomposition and short time fourier transform based hybrid method," *Engineering science and technology, an international journal*, vol. 19, no. 3, pp. 1457–1464, 2016.

[7] D. Hu, W. Li, and X. Chen, "Feature extraction of motor imagery EEG signals based on wavelet packet decomposition," in *The 2011 IEEE/ICME International Conference on Complex Medical Engineering*, 2011, pp. 694–697.

[8] M.-a. Li, W. Zhu, H.-n. Liu, and J.-f. Yang, "Adaptive feature extraction of motor imagery EEG with optimal wavelet packets and SE-isomap," *Applied Sciences*, vol. 7, no. 4, p. 390, 2017.

[9] B. H. Yang, T. Wu, Q. Wang, and Z. J. Han, "Motor imagery eeg recognition based on WPD-CSP and KF-SVM in brain computer interfaces," in *Applied Mechanics and Materials*, vol. 556. Trans Tech Publ, 2014, pp. 2829–2833.

[10] J. F. D. Saa and M. Çetin, "A latent discriminative model-based approach for classification of imaginary motor tasks from EEG data," *Journal of neural engineering*, vol. 9, no. 2, p. 026020, 2012.

[11] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "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.

[12] M. Bentlemsan, E.-T. Zemouri, D. Bouchaffra, B. Yahya-Zoubir, and K. Ferroudji, "Random forest and filter bank common spatial patterns for EEG-based motor imagery classification," in *2014 5th International conference on intelligent systems, modelling and simulation*, 2014, pp. 235–238.

[13] A. Tiwari and A. Chaturvedi, "A multiclass EEG signal classification model using spatial feature extraction and XGBoost algorithm," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 4169–4175.

[14] M. Dai, D. Zheng, R. Na, S. Wang, and S. Zhang, "EEG classification of motor imagery using a novel deep learning framework," *Sensors*, vol. 19, no. 3, p. 551, 2019.

[15] R. Zhang, Q. Zong, L. Dou, and X. Zhao, "A novel hybrid deep learning scheme for four-class motor imagery classification," *Journal of neural engineering*, vol. 16, no. 6, p. 066004, 2019.

[16] S. U. Amin, M. Alsulaiman, G. Muhammad, M. A. Bencherif, and M. S. Hossain, "Multilevel weighted feature fusion using convolutional neural networks for EEG motor imagery classification," *IEEE Access*, vol. 7, pp. 18 940–18 950, 2019.

[17] Z. Zhang, F. Duan, J. Solé-Casals, J. Dinarès-Ferran, A. Cichocki, Z. Yang, and Z. Sun, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15 945–15 954, 2019.

[18] F. Wang, S.-h. Zhong, J. Peng, J. Jiang, and Y. Liu, "Data augmentation for eeg-based emotion recognition with deep convolutional neural networks," in *International Conference on Multimedia Modeling*. Springer, 2018, pp. 82–93.

[19] G. Dai, J. Zhou, J. Huang, and N. Wang, "HS-CNN: a CNN with hybrid convolution scale for EEG motor imagery classification," *Journal of neural engineering*, vol. 17, no. 1, p. 016025, 2020.

[20] K.-W. Ha and J.-W. Jeong, "Motor imagery EEG classification using capsule networks," *Sensors*, vol. 19, no. 13, p. 2854, 2019.

[21] G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang, and A. Etemad, "Classification of hand movements from EEG using a deep attention-based LSTM network," *IEEE Sensors Journal*, 2019.

[22] J. Chen, D. Jiang, and Y. Zhang, "A hierarchical bidirectional GRU model with attention for EEG-based emotion classification," *IEEE Access*, vol. 7, pp. 118 530–118 540, 2019.

[23] A. Burrello, K. A. Schindler, L. Benini, and A. Rahimi, "Hyperdimensional computing with local binary patterns: One-shot learning for seizure onset detection and identification of ictogenic brain regions from short-time iEEG recordings," *IEEE Transactions on Biomedical Engineering*, 2019.

[24] J. Narwariya, P. Malhotra, L. Vig, G. Shroff, and T. Vishnu, "Meta-learning for few-shot time series classification," in *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, 2020, pp. 28–36.

[25] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1199–1208.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[27] R. Leeb, C. Brunner, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008–Graz data set B," *Graz University of Technology, Austria*, pp. 1–6, 2008.

[28] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008–Graz data set A," *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, vol. 16, 2008.