EFMLNet: Fusion Model Based on End-to-End Mutual Information Learning for Hybrid EEG-fNIRS Brain-Computer Interface Applications

Lina Qiu (lina.qiu@scnu.edu.cn)

Weisen Feng (fws0104@163.com)

Zuorui Ying (854535913@qq.com)

Jiahui Pan (panjiahui@m.scnu.edu.cn)
School of Software, South China Normal University, Guangzhou, 510630, Guangdong Province, China

Abstract

Electroencephalography (EEG) and functional near infrared spectroscopy (fNIRS), both portable and non-invasive, enhance brain-computer interface (BCI) performance by integrating their spatial and temporal benefits when combined together. However, the fusion of these two signals still faces challenges. To fully unitize the complementarity of EEG and fNIRS for improved performance in EEG-fNIRS BCI, we propose an EEG-fNIRS fusion network based on end-to-end mutual information learning, named EFMLNet. In the model, EEG and fNIRS data are fed into their respective feature extractors for the extraction of temporal and spatial information. Furthermore, their complementary information is fused by two parallel mutual learning modules. We conducted classification experiments on a publicly available BCI dataset based on motor imagery (MI) task and achieved a cross-subject classification accuracy of 71.52%. This result surpasses the performance of most existing fusion methods and demonstrates the potential for real-time hybrid BCI systems.

Keywords: EEG-fNIRS; Multimodal fusion; End-to-End; Hybrid BCI; Mutual Learning

Introduction

Brain-computer interfaces (BCIs) has emerged as valuable tools for neurorehabilitation, aiding disabled individuals and detecting neurological conditions such as disorders of consciousness (Pan et al., 2023). Non-invasive BCIs utilize technologies like electroencephalography (EEG), functional near-infrared spectroscopy (fNIRS), and functional magnetic resonance imaging (fMRI). Among them, EEG and fNIRS are favored for real-world applications due to their portability and cost-effectiveness, with extensive research currently focusing on these two modalities (Gao et al., 2023).

EEG is a widely used technique for examining brain activity, recording neuronal voltage fluctuations via scalp electrodes (Buzsáki, Anastassiou, & Koch, 2012). Renowned for its excellent temporal resolution and responsiveness, EEG is favored in cognitive field (Ray & Cole, 1985). However, it faces limitations in spatial precision and susceptibility to motion artifacts and noise, potentially causing misinterpretation of resting-state signals in BCI systems (Eldele et al., 2021).

fNIRS measures cerebral cortex blood flow and metabolism by monitoring oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) using near-infrared light (Quaresima & Ferrari, 2019). It offers better spatial resolution and less noise interference than EEG (Rahman et al., 2020). Nevertheless, its temporal resolution is poor, and delayed hemodynamic response make it challenging to construct a real-time BCI alone.

Optimal BCI systems are characterized by portability, non-invasive, and superior accuracy and efficiency. Integrating various brain signal modalities has been proven to enhance BCI performance (Ferdinando et al., 2023; D. Wang et al., 2023; Park, Ha, & Kim, 2023). However, the challenge lies in utilizing the distinct and complementary data from diverse modalities to surmount the constraints of single-mode systems and enhance overall functionality in multimodal BCI applications. In the study of (Yin et al., 2015), they merged EEG and fNIRS into a single vector and optimized using the joint mutual information criterion to enhance classification performance. Shin et al. (2016) extracted prediction scores from EEG and fNIRS signals respectively, and then used LDA-based meta-classifiers to obtain final prediction results. These methods, while promising, do not effectively harness the synergistic qualities of EEG and fNIRS, leading to suboptimal predictive performance in EEG-fNIRS BCI.

Several studies have explored using fNIRS signals as an auxiliary tool for EEG-based BCI systems, that is, using fNIRS spatial prior information to optimize the processing of EEG signals in the BCI study. In an EEG-fNIRS BCI study, R. Li et al. (2017) determined the two fNIRS channels with the strongest task-induced response by general linear modeling (GLM), and then selected two EEG channels near them for performance evaluation of the hybrid fNIRS-EEG BCI system. In another study, Kwak, Song, and Kim (2022) designed a fNIRS-guided attention network for the EEG-fNIRS BCI , where fNIRS guides the important region for brain decoding and applies spatial attention to EEG features. These EEG-fNIRS BCI methods based on fNIRS- or EEG-informed can improve system performance to a certain extent by utilizing the complementary characteristics of both. However, this method naturally biases towards one modality, inevitably leading to the loss of information from the other modality, which can easily result in information bias.

In summary, while current integration techniques moderately enhance the synergistic information from EEG and fNIRS modalities, they typically fall short in thoroughly mining and capitalizing on the complementary attributes between EEG and fNIRS signals. This limitation tends to constrain the efficacy of hybrid EEG-fNIRS BCI systems. Moreover, most methods are hindered by cumbersome preprocessing

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

and manual feature extraction, impeding their efficiency and online applicability. To address this, we introduce the EEG-fNIRS Mutual Learning network (EFMLNet), a deep learning model that streamlines fusion through end-to-end mutual information learning for hybrid EEG-fNIRS BCI systems. This approach effectively streamlines the workflow for researchers and leverages the complementary characteristics of the two modalities, achieving commendable classification results in cross-subject validation. Our contributions are as follows:

- We designed personalized feature extractors for EEG and fNIRS, respectively, to extract their crucial features in temporal and spatial dimensions.
- We proposed an EEG-fNIRS deep fusion method based on end-to-end mutual information learning, named EFMLNet, which utilizes the multi-head cross attention mechanism of Transformer to achieve complementary information learning between two modalities.
- We conducted end-to-end cross-subject experiments on a publicly available EEG-fNIRS BCI dataset and achieved superior classification results.

Method

Model Framework

This study proposes a novel EEG-fNIRS fusion model based on end-to-end mutual information learning to enhance the performance of the hybrid EEG-fNIRS BCI. The model consists of a feature extractor module and a mutual learning module, as depicted in Figure 1. Two automatic feature extractors are designed for EEG and fNIRS data. The EEG extractor includes two convolutional neural networks (CNNs) and a long short-term memory network (LSTM) to capture temporal dynamics, while the fNIRS extractor uses an Embedding layer and a Transformer encoder for spatial features. The extracted modality-specific features are input to a cross-modal mutual learning module with symmetrical Transformers. Predictive losses from extractors are combined with the fusion module's loss to form a joint loss, ensuring balanced representation learning across modalities for robust outcomes.

Feature Extractor Module In order to effectively distill crucial information from both the EEG and fNIRS modalities, we developed specialized feature extractors for each.

The EEG feature extractor integrates two one-dimensional CNNs followed by a LSTM network. Research has shown that one-dimensional CNNs are adept at capturing localized temporal patterns within signals, which renders them particularly effective for analyzing time-series data (Chua & Roska, 1993). The CNN we designed consists of two one-layer convolution with a kernel size of 1×3, a step size of 1, and a padding value of 0. Each convolution layer is followed by a batch normalization layer and an activation function. As a special recurrent neural network (RNN) (Cho et al., 2014) architecture, LSTM solves the gradient disappearance problem by introducing several gates (input gates, forget gates, and output gates) and cell states (Greff et al., 2016). The key to LSTM is its cellular state, which traverses the entire chain with only a few small linear interactions. There are few barriers to the flow of information, allowing it to flow unharmed throughout the sequence. Therefore, the extractor designed based on CNN and LSTM can effectively capture and extract the key temporal features of EEG signals, which denoted as $E_{feature}$. Moreover, such a simplified design is advantageous in mitigating potential overfitting issues that may arise from complex deep learning models.

The fNIRS feature extractor is structured with an input embedding layer that incorporates positional encoding, followed by a Transformer encoder layer. The encoder layer further consists of two sub-layers: a Multi-head Self-Attention (MHSA) mechanism and a Feedforward Neural Network (FNN), each followed by residual connections and layer normalization. The embedding layer, combined with positional encoding, equips elements with positional information, allowing the encoder to discern the relative position of input elements for subsequent learning. The core of spatial information extraction process is the MHSA mechanism. In the regions of brain activity, the signals recorded by each channel play distinct roles, so the feature extraction should be more targeted. MHSA excels in identifying and linking global relationships across various positions within the input, assigning significance to channels through the calculation of attention weights for each position. In this manner, we further enhance the spatial information of the data, thereby enriching the model's comprehension of the entire sequence architecture.

The data, once filtered, undergoes normalization via a linear embedding layer that projects each element into a high-dimensional space. The outcome of this process is denoted as F_1. Unlike RNNs, Transformer is lack of sequential processing capabilities, therefore, to grasp the relative positioning of elements within the sequence, it is essential to incorporate position embeddings (PE):

$$F_2 = F_1 + F_{pos} \tag{1}$$

F_{pos} represents the position embeddings, and the specific information is frequency band and channel.

Subsequently, the data are fed into a standard Transformer encoder. The encoder utilizes the MHSA layer as one of its key components to focus on various segments of the input sequence. The integration of the residual connection (RC) surrounding this layer, along with subsequent layer normalization (LN), facilitates more effective gradient propagation and allows for easier training. After the data passes through the self-attention layer, the representation of each position undergoes a nonlinear transformation via a feedforward neural network. Similar to the processing after the attention layer, the feedforward neural network layer also uses residual connections and layer normalization to enhance the network's learning capability. Finally, the input sequence passed through the Transformer encoder will be mapped to the new representation, denoted as $F_{feature}$.

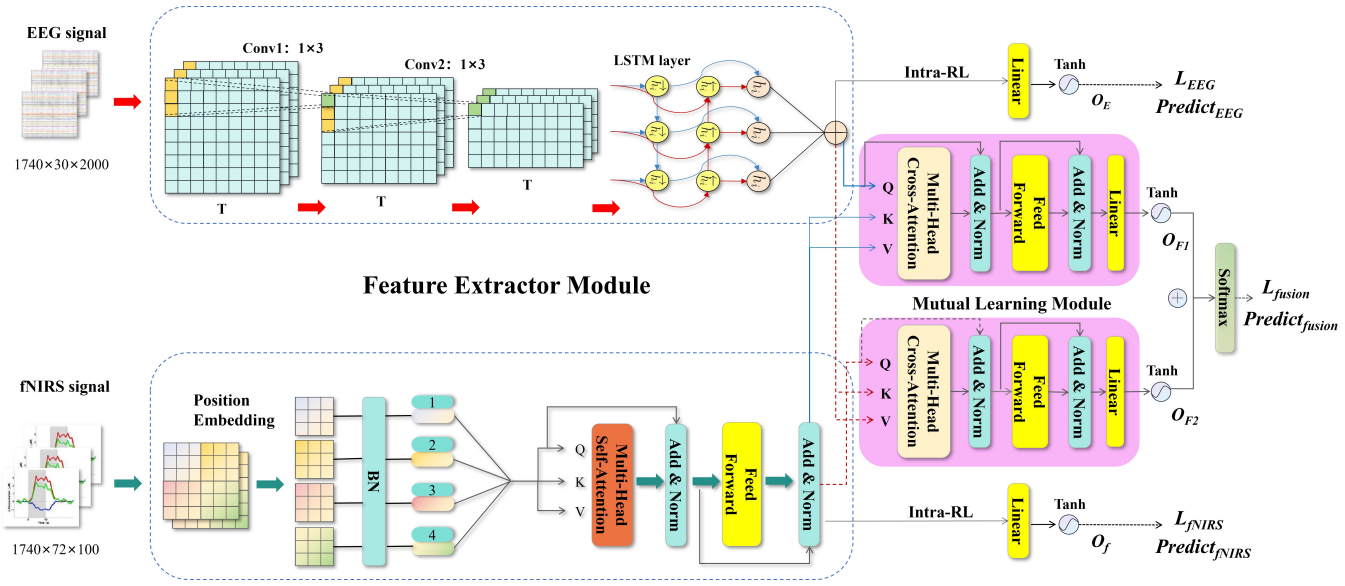

Figure 1: The Structure of EFMLNet.

Mutual-learning module Following the extraction of modality-specific extractor, the EEG and fNIRS features are separately fed into the cross-modal mutual learning module. Within this module, a pair of structurally identical cross-modal Transformers have been implemented to facilitate the mutual learning. To effectively learn the information between different modalities and achieve a complementary effect, we have also utilized the multi-head attention mechanism of the Transformer. The cross-modal transformers are similarly constituted of attention layers, Feed-Forward Networks, layer normalization, and residual connections. However, unlike the feature extractors for fNIRS, the attention layers in these transformers are specifically employed to help the model discern inter-feature relationships and to learn the mapping of information from one modality to another. The parallel processing capability allows the model to treat the input features as several blocks, enabling it to focus on different parts of the input simultaneously. With multi-head attention, the model can effectively transmit and integrate information from different modes in the process of coding.

The principle involved in implementing information mapping between different modalities are mainly related to the Q (Query), K (Key), and V (Value) components within the Transformer. They are three linear transformations in the attention mechanism that are used to map different aspects of the input sequence. Q is the representation of the input sequence after a linear transformation, expressed as:

$$Q = I \cdot W_q \tag{2}$$

where I is the representation of the input sequence and W_q is the learned weight matrix. Q is used in the attention mechanism to generate an attention score, that is, to determine the attention weight of other locations to the current location. It indicates the importance of the current location and influences the attention of other locations.

K is the representation of the input sequence after another linear transformation, expressed as:

$$K = I \cdot W_k \tag{3}$$

where I is the representation of the input sequence and W_k is the learned weight matrix. The K provides the information used to calculate the Q's attention score for other locations. By comparing Q and K, the model can determine the data relevance between different locations.

V is the representation of the input sequence obtained through different linear transformations, expressed as:

$$V = I \cdot W_v \tag{4}$$

where I is the representation of the input sequence and W_v is the learned weight matrix. The V contains information about the sequence of inputs, weighted summing the inputs based on the calculated attention score to produce the final output representation. It represents information about where the attention mechanism is focused.

In the mutual learning module, we treat the $F_{feature}$ as the Q and the $E_{feature}$ as the K and V, and vice versa—treating the $E_{feature}$ as the Q and the $F_{feature}$ as the K and V, thereby achieving mutual learning of the data through query matching. The cross attention of the two cross-modal Transformers are calculated as follows:

$$T1_i(F_{feature}, E_{feature}) = Softmax(\frac{M_{Qi}N_{Ki}^T}{\sqrt{dim}})N_{Vi} \tag{5}$$

$$T2_i(E_{feature}, F_{feature}) = Softmax(\frac{N_{Qi}M_{Ki}^T}{\sqrt{dim}})M_{Vi} \tag{6}$$

$T1_i$ and $T2_i$ respectively represent the cross-attention of the i-th head in the two cross-modal Transformers, while M_{Qi},

N_{Ki}, and N_{Vi} correspond to the linear projections of $E_{feature}$ and $F_{feature}$. Combining the attention of all the heads in a single module gives the input sequence I, which performs the corresponding operations of Q, K, and V.

$$I_E = \left[T1_1(F_{feature}, E_{feature}), \cdots, T1_n(F_{feature}, E_{feature})\right] \quad (7)$$

$$I_F = \left[T2_1(E_{feature}, F_{feature}), \cdots, T2_n(E_{feature}, F_{feature})\right] \quad (8)$$

After multiple cross-attention layers is the superposition of residual connection and feedforward layer, the processed data passes through a linear layer and is activated by the *tanh* (Fan, 2000) function, as in the case of single-modality independent prediction. Since we have two identical cross-modality Transformer modules, there will be two output results, denoted as O_{F1} and O_{F2}. By summing these two results and passing the sum through a softmax function, we obtain the final prediction outcome. This also represents the output of the entire model, as shown in Equation 9. Meanwhile, the loss function of fusion prediction L_{fusion} is used for model training.

$$Predict = Softmax(O_{F1} \oplus O_{F2}) \quad (9)$$

During the training process, we employed cross-entropy (L. Li, Doroslovački, & Loew, 2020) as the loss function for model training, assessing the training effectiveness by comparing the discrepancy between predicted labels and actual labels.

Experiment

In our study, we utilized a motor imagery (MI) public dataset (Shin et al., 2016) containing EEG and fNIRS signals to conduct experiments, which represent motor imagery (MI) tasks.

Dataset

The dataset was collected from 29 subjects (14 men and 15 women) with a mean age of 28.5±3.7 years (mean±SD). EEG signals were recorded using 30 channels at a sampling frequency of 1000 Hz, while fNIRS signals were recorded using 36 channels at a sampling frequency of 12.5 Hz. During the experiments, subjects performed MI tasks according to the given instructions.

Data Processing

Prior to conducting data fusion analysis, we performed filtering and downsampling procedures on the initial dataset. Specifically, for EEG signals, we implemented a third-order Butterworth filter to perform bandpass filtering within the 1-45Hz range (Kocsis, Herman, & Eke, 2006), and subsequently reduced the sampling rate to 200Hz. Regarding fNIRS signals, we utilized a third-order Butterworth filter to conduct bandpass filtering between 0.01-0.09Hz for both HbO and HbR, and then downscaled the sampling frequency to 10Hz.

Experiment Settings

In our study, we adopted the Leave One Subject Out (LOSO) method for training and testing our model. The dataset contains a total of 1740 samples. In each iteration, we removed all the data from one subject (60 samples) to use as the test set, while the remaining subjects' data were used for training. The input dimensions of the model were $1740\times30\times2000$ for EEG data and $1740\times72\times100$ for fNIRS data. Regarding the choice of optimizer, we utilized the Adam optimizer (Kingma & Ba, 2014) and set the weight decay to 0.0008.

Evaluation Metrics

To validate the effectiveness of the model, the experiment employed commonly used metrics in classification tasks to evaluate the model. These included Accuracy, Precision, Recall, and the F1-Score.

Results

To validate the effectiveness of our model, we conducted comprehensive classification experiments on a MI dataset, including comparative experiments and ablation experiments.

The Results of Comparative Experiments

We first compared the multimodal fusion results based on EFMLNet (i.e., EEG-fNIRS fusion) with results obtained from unimodal predictions (i.e., EEG-only or fNIRS-only).

The comparative results of EEG-fNIRS fused by EFMLNet with EEG-only and fNIRS-only in MI tasks are presented in Table 1. It can be seen that the performance of each subject in the hybrid EEG-fNIRS, fused by EFMLNet, is consistently superior to that of EEG-only and fNIRS-only.

Table 1: Comparative classification performance of EEG-fNIRS fusion based on EFMLNet with EEG-only and fNIRS-only in MI tasks

Modality	EEG	fNIRS	**EEG-fNIRS**
Accuracy (%)	63.24%	53.46%	**71.52%**
Precision (%)	62.51%	53.08%	**71.02%**
Recall	66.72%	56.74%	**73.33%**
F1-score	63.24%	53.22%	**71.08%**

To further validate that the favorable results in Table 1 are not solely attributed to the use of multimodal signals but also to our fusion model, we conducted comparisons by contrasting the fusion results of EFMLNet with those of a simple concatenation fusion of multimodal signals, as shown in Table 2. In comparison to the results obtained by employing simple concatenation fusion and classification through LDA (Meng et al., 2021), KNN, CNN (Nour, Öztürk, & Polat, 2021), and ResNet, our model consistently exhibits superior performance in MI tasks. The accuracy increased by 10.35-20.31%. Precision improved by 8.92-19.70%, while recall increased by 13.33-23.26%. F1-score also demonstrated improvements

Table 2: Comparison of EFMLNet with traditional models under EEG-fNIRS modality fusion for MI tasks.

Method	Modal	Accuracy	Precision	Recall	F1-score
Concatenate Fusion	LDA	51.21%	51.32%	50.92%	50.41%
	KNN	53.34%	53.91%	50.33%	51.57%
	CNN	61.17%	62.10%	60.00%	61.04%
	ResNet	54.74%	54.83%	56.72%	54.71%
Ours	**EFMLNet**	**71.52%**	**71.02%**	**73.33%**	**71.08%**

of 10.04-20.67%. These results suggest that our model can further enhance the performance of the EEG-fNIRS system.

Furthermore, we compared our model with some state-of-the-art methods conducted on the same dataset, including the following studies:

- **(Jiang et al., 2019)**: An independent decision path fusion (IDPF), leveraging both EEG and fNIRS technologies to distinguish various mental states.
- **(Esfahani & Sadati, 2021)**: Ensemble learning, which enhances accuracy and reduces standard deviation, leading to improved outcomes from classification models when the variance of predictions is diminished.
- **(Zhang et al., 2021)**: An 3D convolutional neural network, which can preserve the temporal information of the EEG data while maintaining its spatial topological features.
- **(He et al., 2022)**: A novel end-to-end multimodal multitask neural network (M2NN), which integrates the spatial–temporal feature extraction module, multimodal feature fusion module, and multitask learning (MTL) module.
- **(Z. Wang, Fang, & Zhang, 2023)**: fNIRSNet, incorporating the delayed hemodynamic response as domain knowledge into fNIRS classification.

The results are shown in Table 3. Although Jiang et al. (2019) achieve a classification accuracy of 70.32%, their method will result in the computational burden escalates with an increasing number of decision paths. Esfahani and Sadati (2021) focused on integrated learning, but their method relies on a voting mechanism which has poor portability. In contrast, our model can be directly transferred and used without the need for such mechanisms. The approach of Zhang et al. (2021) is capable of extracting features from signals at multiple scales, yet their focus on EEG signals results in the loss of fNIRS information. Conversely, Z. Wang et al. (2023) prioritizes fNIRS signals, overlooking EEG. Our method, however, is able to fully integrate information from both modalities. He et al. (2022) also employed an end-to-end approach; however, their classification accuracy on MI was 62.26%, which is lower than our 71.52%.

In comparison to these fully cross-subject experiments, our model demonstrates superior performance, indicating the efficacy of our model.

Table 3: The comparative results between the EFMLNet model and state-of-the-art methods

Studies	Method	Classifier	Modality	Accuracy
Jiang et al.	IDPF, LOOCV	LDA,SVM, HMM	EEG-fNIRS	70.32%
Esfahani et al.	Ensemble Learning, LOOCV	FCM-ANFIS	EEG-fNIRS	70.40%
Zhang et al.	End to end, LOSO	3D-CNN	EEG-fNIRS	70.15%
He et al.	MTL, LOSO-CV	M2NN	EEG-fNIRS	62.26%
Wang et al.	LOSO-CV	fNIRSNet	EEG-fNIRS	65.26%
Ours	**End to end, LOSO**	**EFMLNet**	**EEG-fNIRS**	**71.52%**

The Results of Ablation Experiments

Feature Extractors Experimental results have confirm the model's effectiveness, nevertheless, it's uncertain if feature extractors act as precursors to the mutual learning process or if the mutual learning modules alone are inherently performant. To ascertain this, we opted to bypass the feature extraction stage, directly feeding raw data into the mutual learning modules for examination.

As shown in the MI results of Figure 2, after eliminating the feature extractors, our model's performance decreases significantly compared to the complete model (EFMLNet). Specifically, the classification accuracy plummets by 20.48%, highlighting the vital importance of feature extractors in EFMLNet. This revelation underscores the indispensability of feature extraction for multimodal signals prior to fusion, especially focusing on personalized feature extraction tailored to the unique characteristics of each modality.

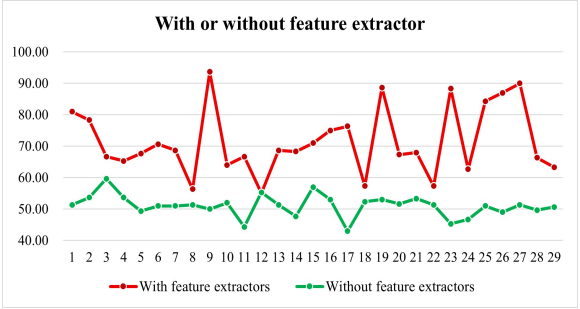

Figure 2: Comparison of MI classification accuracy for each subject of the model with and without feature extractors.

Multimodal Mutual Learning To ascertain the effectiveness and necessity of multimodal mutual learning within EFMLNet, we performed an ablation experiment, specifically by eliminating the cross-modal Transformer on one side of the mutual learning module to achieve one-way guidance. This ablation encompassed two scenarios of unilateral learning: one where the EEG signal served as the primary source and the fNIRS signal as the learning objective, and another where the fNIRS signal served as the main focus and the EEG signal as the target of learning.

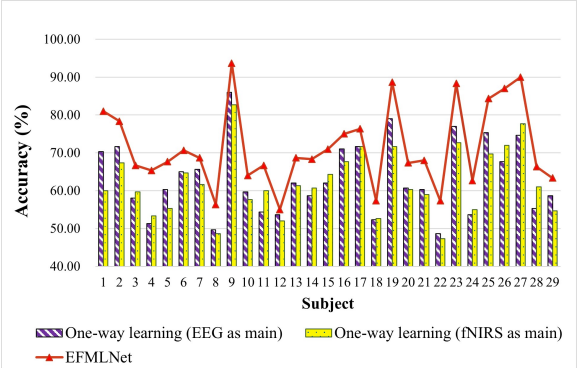

Figure 3: Comparison of classification accuracy between mutual learning and one-way learning for each subject in the MI task.

As illustrated in Figure 3, both one-way learning guided by fNIRS signals and guided by EEG signals exert comparable influences on the ultimate classification results, indicating that the extracted features from EEG and fNIRS contain a nearly equivalent amount of valid information. Through mutual learning, the pertinent information from both modalities is optimized and fully exploited. In comparison to one-way guided learning, mutual learning enhanced the average classification accuracy of MI tasks by 8.26% when EEG serves as the primary modality and by 9.37% when fNIRS is the main modality.

Conclusion

In this paper, we propose an end-to-end deep neural network for mutual learning between EEG and fNIRS data. The model automatically extracts data features internally and utilizes parallel multi-head attention to achieve complementary learning between different modal data. Experimental results indicate that this model achieves outstanding performance compared to existing models. Moreover, the model can also be generalized to the study of other physiological signals with complementary characteristics.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under grant 82302339, and the Special Innovation Projects of Colleges and Universities in Guangdong Province under grant 2022KTSCX035.

References

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents—eeg, ecog, lfp and spikes. *Nature reviews neuroscience*, *13*(6), 407–420.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Chua, L. O., & Roska, T. (1993). The cnn paradigm. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, *40*(3), 147–156.

Eldele, E., Chen, Z., Liu, C., Wu, M., Kwoh, C.-K., Li, X., & Guan, C. (2021). An attention-based deep learning approach for sleep stage classification with single-channel eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 809–818.

Esfahani, M. M., & Sadati, H. (2021). fnirs signals classification with ensemble learning and adaptive neuro-fuzzy inference system. In *2021 7th international conference on signal processing and intelligent systems (icspis)* (pp. 1–5).

Fan, E. (2000). Extended tanh-function method and its applications to nonlinear equations. *Physics Letters A*, *277*(4-5), 212–218.

Ferdinando, H., Moradi, S., Korhonen, V., Kiviniemi, V., & Myllylä, T. (2023). Altered cerebrovascular-csf coupling in alzheimer's disease measured by functional near-infrared spectroscopy. *Scientific Reports*, *13*(1), 22364.

Gao, Y., Jia, B., Houston, M., & Zhang, Y. (2023). Hybrid eeg-fnirs brain computer interface based on common spatial pattern by using eeg-informed general linear model. *IEEE Transactions on Instrumentation and Measurement*.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, *28*(10), 2222–2232.

He, Q., Feng, L., Jiang, G., & Xie, P. (2022). Multimodal multitask neural network for motor imagery classification with eeg and fnirs signals. *IEEE Sensors Journal*, *22*(21), 20695–20706.

Jiang, X., Gu, X., Xu, K., Ren, H., & Chen, W. (2019). Independent decision path fusion for bimodal asynchronous brain–computer interface to discriminate multiclass mental states. *IEEE Access*, *7*, 165303–165317.

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

Kocsis, L., Herman, P., & Eke, A. (2006). The modified beer–lambert law revisited. *Physics in Medicine & Biology*, *51*(5), N91.

Kwak, Y., Song, W.-J., & Kim, S.-E. (2022). Fganet: fnirs-guided attention network for hybrid eeg-fnirs brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 329–339.

Li, L., Doroslovački, M., & Loew, M. H. (2020). Approximating the gradient of cross-entropy loss function. *IEEE*

access, *8*, 111626–111635.

Li, R., Potter, T., Huang, W., & Zhang, Y. (2017). Enhancing performance of a hybrid eeg-fnirs system using channel selection and early temporal features. *Frontiers in human neuroscience*, *11*, 462.

Meng, M., Dai, L., She, Q., Ma, Y., & Kong, W. (2021). Crossing time windows optimization based on mutual information for hybrid bci. *Mathe. Biosci. Eng*, *18*, 7919–7935.

Nour, M., Öztürk, Ş., & Polat, K. (2021). A novel classification framework using multiple bandwidth method with optimized cnn for brain–computer interfaces with eeg-fnirs signals. *Neural Computing and Applications*, *33*, 15815–15829.

Pan, J., Cai, H., Huang, H., He, Y., & Li, Y. (2023). Multiple scale convolutional few shot learning networks for online p300-based brain-computer interface and its application to patients with disorder of consciousness. *IEEE Transactions on Instrumentation and Measurement*.

Park, S., Ha, J., & Kim, L. (2023). Improving performance of motor imagery-based brain–computer interface in poorly performing subjects using a hybrid-imagery method utilizing combined motor and somatosensory activity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 1064–1074.

Quaresima, V., & Ferrari, M. (2019). Functional near-infrared spectroscopy (fnirs) for assessing cerebral cortex function during human behavior in natural/social situations: a concise review. *Organizational Research Methods*, *22*(1), 46–68.

Rahman, M. A., Siddik, A. B., Ghosh, T. K., Khanam, F., & Ahmad, M. (2020). A narrative review on clinical applications of fnirs. *Journal of Digital Imaging*, *33*, 1167–1184.

Ray, W. J., & Cole, H. W. (1985). Eeg alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science*, *228*(4700), 750–752.

Shin, J., von Lühmann, A., Blankertz, B., Kim, D.-W., Jeong, J., Hwang, H.-J., & Müller, K.-R. (2016). Open access dataset for eeg+ nirs single-trial classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(10), 1735–1745.

Wang, D., Liu, A., Xue, B., Wu, L., & Chen, X. (2023). Improving the performance of ssvep-bci contaminated by physiological noise via adversarial training. *Medicine in Novel Technology and Devices*, *18*, 100213.

Wang, Z., Fang, J., & Zhang, J. (2023). Rethinking delayed hemodynamic responses for fnirs classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.

Yin, X., Xu, B., Jiang, C., Fu, Y., Wang, Z., Li, H., & Shi, G. (2015). A hybrid bci based on eeg and fnirs signals improves the performance of decoding motor imagery of both force and speed of hand clenching. *Journal of neural engineering*, *12*(3), 036004.

Zhang, Y., Cai, H., Nie, L., Xu, P., Zhao, S., & Guan, C. (2021). An end-to-end 3d convolutional neural network for decoding attentive mental state. *Neural Networks*, *144*, 129–137.