



A Small-Sample Method with EEG Signals Based on Abductive Learning for Motor Imagery Decoding

Tianyang Zhong¹, Xiaozheng Wei¹, Enze Shi², Jiaxing Gao¹, Chong Ma¹,
Yaonai Wei¹, Songyao Zhang¹, Lei Guo¹, Junwei Han¹, Tianming Liu³,
and Tuo Zhang¹(✉)

¹ School of Automation, Northwestern Polytechnical University, Xi'an 710072, China
tuozhang@nwpu.edu.cn

² School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

³ School of Computing, University of Georgia, Athens 30602, USA

Abstract. Motor imagery (MI) electroencephalogram (EEG) decoding, as a core component widely used in noninvasive brain-computer interface (BCI) system, is critical to realize the interaction purpose of physical world and brain activity. However, the conventional methods are challenging to obtain desirable results for two main reasons: there is a small amount of labeled data making it difficult to fully exploit the features of EEG signals, and lack of unified expert knowledge among different individuals. To handle these dilemmas, a novel small-sample EEG decoding method based on abductive learning (SSE-ABL) is proposed in this paper, which integrates perceiving module that can extract multiscale features of multi-channel EEG in semantic level and knowledge base module of brain science. The former module is trained via pseudo-labels of unlabeled EEG signals generated by abductive learning, and the latter is refined via the label distribution predicted by semi-supervised learning. Experimental results demonstrate that SSE-ABL has a superior performance compared with state-of-the-art methods and is also convenient for visualizing the underlying information flow of EEG decoding.

Keywords: MI Decoding · Small-Sample · Abductive Learning · Visualization

1 Introduction

Brain computer interface (BCI) technology plays an increasingly crucial role in the rehabilitation process of patients with nerve damage. However, the lack of large-scale labeled data makes the identification results more vulnerable to be affected. BCI technology with motor imagery (MI), which can decode neural activities through EEG signals to identify the movement intention of the human being, is widely applied in the field of EEG decoding, where the substantial problem of BCI decoding is to extract as much effective information as possible from the multi-channel and non-linear EEG signals for understanding the oscillating activities in the brain [1].

According to the development of EEG signal processing technology, the traditional EEG signal feature extraction methods [2, 3] rely on rich human experience or expert knowledge. Furthermore, the existing deep learning models for EEG motion decoding [4, 5] rely on a large amount of labeled data. More recently, the active learning framework [6, 7] based on semi-supervised learning is used to solve a feature extraction problem with a large amount of unlabeled data, but it is sensitive to the initial model accuracy and sampling strategy. Therefore, there is still an unsolved urgent challenge that needs to be considered further, which is how to make biologically reasonable use of the information contained in a large number of unlabeled EEG data on the above basis.

The starting point of this paper is that the following biological findings can be observed in the MI decoding [8–10]: 1) the EEG rhythm energy in the contralateral motor sensory area (MSA) of the cerebral cortex is significantly reduced, while that in the ipsilateral MSA is increased, 2) different functional brain regions activities can be reflected by different sub-bands of EEG signals, and 3) the distribution density of different rhythms are clearly distinguished. And the rules are similar to the knowledge reasoning used in human decision-making. Then, the purpose of this paper is to employ these knowledge constraints to guide the model to extract MI information from unlabeled EEG data in a mutually beneficial way.

Therefore, based on abductive learning (ABL) [11], this article proposes a small-sample EEG decoding method, which can adaptively extract abstract features from complex and dynamic EEG signals, and an efficient knowledge base is designed to constrain the training process of the model. The main contributions are listed as follows:

- 1) A novel EEG decoding method is proposed to tackle the MIR (motion intention recognition) problem with small-sample EEG signals. This method does not rely on strict mathematical assumptions for datasets and its accuracy and robustness are well-maintained under strong interference.
- 2) A multi-scale feature fusion network is designed to enhance abstract features, which can capture temporal and frequency information across multi-channel EEG signals and spatial relationships among different electrodes.
- 3) An effective knowledge base module of motor imagery is constructed and symbolized, which can upgrade the model space under this constraint by mining the potential facts of large-scale unlabeled EEG signals.

2 Problem Definition and Method

2.1 Problem Definition

In the SSE-ABL framework for EEG decoding, the given input is defined as **Input** = $\{X_l, X_u, KB_\theta\}$, where tensor X_l denotes labeled EEG data, tensor X_u denotes unlabeled EEG data, the data size of X_u is much larger than that of X_l , and KB_θ denotes knowledge base on brain science in MI task. Concretely, $X_l = \{(x_{l1}, y_{l1}), (x_{l2}, y_{l2}), \dots, (x_{li}, y_{li})\}$, where variable x_{li} represents multi-channel EEG signals, and y_{li} implies the corresponding labels. And the assignment of X_l is to learn a mapping from x to y . $X_u = \{x_{u1}, x_{u2}, \dots, x_{uj} | i \ll j\}$, where x_{uj} denotes unlabeled multi-channel EEG signals, which are utilized to boost representative capability of the above mapping. KB_θ consists of a series of first-order logic sentences with learnable parameters θ , which

integrate labels EEG data X_l and unlabeled data X_u to optimize a perceptual model and train parameters θ with the constraint of knowledge base. The final result is regarded as **Output** = $\{f, W, \theta'\}$, where f is a mapping from EEG signals to motion intention, $W \in \mathbf{R}^{m \times n}$ indicates the proportion of the m^{th} sub-band to the n^{th} channel in EEG data, $\theta' \in \mathbf{R}^k$ represents the contribution rate of the k^{th} channel to the whole in MIR. The SSE-ABL algorithm yields the corresponding pseudo-labels to the unlabeled data by the classifier optimized by a small amount of labeled EEG signals, and the produced labels may be incorrect due to the small number of training samples, which is difficult to guarantee good performance. Therefore, the SSE-ABL modifies the pseudo-labels and optimizes the internal parameters of the knowledge base at the same time, so that the consistency of them is maximized under the constraint of the knowledge base. Formally, the problem definition of the SSE-ABL can be summarized as an optimization problem of searching **Output** under a given **Input**:

$$\min_{f, W, \theta} \mathcal{Loss}_{label}(y_{li}, f_{li}) + \mathcal{Loss}_{unlabel}(\delta(y'_{uj}), f_{uj}) \quad (1)$$

$$\text{s.t. } \underset{\delta}{\operatorname{argmax}} \operatorname{constraint}(\delta(y'_{uj}), f_{li}, KB_{\theta}) \quad (2)$$

where y'_{uj} is the pseudo-label corresponding to the j^{th} unlabeled instance, which is generated by the perceptual module. $\delta(\cdot)$ indicates a heuristic function obtained by optimization, which aims to revise pseudo-labels by logical abduction process. In addition to correcting inconsistent pseudo-labels, this goal also helps the knowledge base to learn accurate parameter θ . It can be seen from Eq. 1 and Eq. 2 that the major challenge is how to mine the effective information of massive unlabeled EEG data under the KB_{θ} constraints and react to the iterative update of itself.

2.2 Architecture of the SSE-ABL Framework

The proposed method, as shown in Fig. 1, consists of four phases. A workflow outline of the proposed method in this paper is displayed in Algorithm 1.

- 1) Phase 1 (Sample Representation): For the purpose of promoting the ability to describe the local and global details of the brain activities, a time-frequency-space data representation method based on brain region division is proposed, which makes the receptive field of view cover the whole brain region in a fine-grained way.
- 2) Phase 2 (Multiscale Feature Fusion): Inspired by the neuroscience findings shown in Sect. 1, a multi-scale feature fusion model is proposed to adaptively integrate time-frequency-space information of EEG signals aiming at reducing the potential difference of feature distribution and selectively focusing on the MI-related materials.
- 3) Phase 3 (Motion Intention Estimation): Based on the common laws of EEG signals distribution, a motor intention evaluation model is constructed, in which the rules are used as the supervisory information to judge the authenticity of motor imagery recognition, so as to ensure that the overall process conforms to the actual criteria.
- 4) Phase 4 (Abductive Reasoning Optimization): When the classifier is insufficient-trained, the pseudo-labels could be wrong, the SSE-ABL method in this paper needs to correct the wrong pseudo-labels to achieve consistent abductions by using gradient free optimization method under the principle of minimal inconsistency [12, 13].

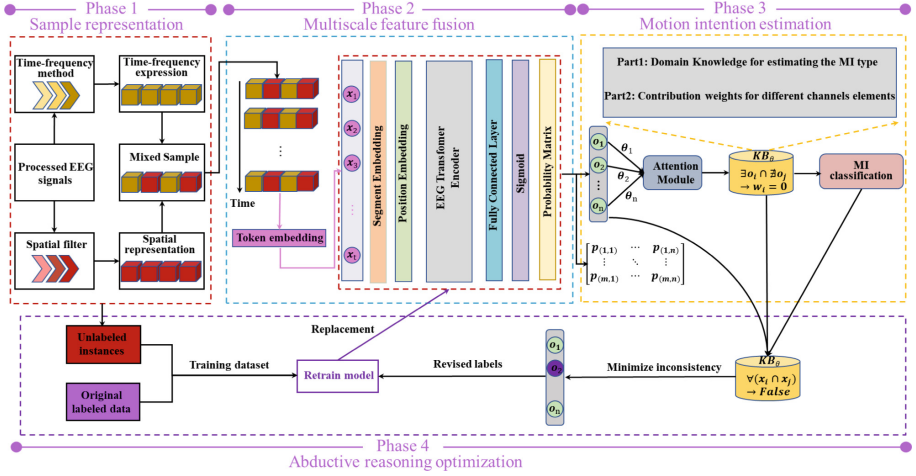


Fig. 1. Architecture of the proposed SSE-ABL method.

Algorithm 1 Small-Sample EEG Decoding Method Based on Abductive Learning

Input: Labeled EEG signals $\{(\mathbf{x}_{l1}, \mathbf{y}_{l1}), (\mathbf{x}_{l2}, \mathbf{y}_{l2}), \dots, (\mathbf{x}_{li}, \mathbf{y}_{li})\}$; Unlabeled EEG signals $\{\mathbf{x}_{u1}, \mathbf{x}_{u2}, \dots, \mathbf{x}_{uj}\}$; Knowledge base KB_{θ}

Parameters: Training parameters(batch size, learning rate, epoch ,etc.)

Output: Perception model f ; Parameters \mathbf{W}, θ'

```

1:  for VarEpoch = 1 → epoch do    # Training epochs
2:     $f = \text{TrainModel}(\{(\mathbf{x}_{l1}, \mathbf{y}_{l1}), (\mathbf{x}_{l2}, \mathbf{y}_{l2}), \dots, (\mathbf{x}_{li}, \mathbf{y}_{li})\})$   # Pre-train model
3:     $\theta = \text{UpdateParameter}(\{\mathbf{y}_{l1}, \mathbf{y}_{l2}, \dots, \mathbf{y}_{li}\})$   # Pre-train learnable parameters
4:  end for
5:  for  $i = 1 \rightarrow E$  do
6:     $\mathbf{Y}' = f(\mathbf{X}_u)$   # Generate pseudo-labels  $\mathbf{Y}'$ 
7:     $\delta(\mathbf{Y}') = \text{Abduce}(KB_{\theta}, \mathbf{Y}')$   # Revise pseudo-labels
8:     $\theta' = \text{UpdateParameter}(\mathbf{Y}_l, \delta(\mathbf{Y}'))$   # Update parameters  $\theta'$ 
9:     $f = \text{RetrainModel}(f, \{(\mathbf{x}_{l1}, \mathbf{y}_{l1}), (\mathbf{x}_{l2}, \mathbf{y}_{l2}), \dots, (\mathbf{x}_{li}, \mathbf{y}_{li})\}, \mathbf{X}_u, \mathbf{Y}')$   # Reupdate  $f$ 
10:    $i = i + 1$ 
11: end for
12: return Output

```

2.3 Sample Representation

The preprocessed EEG signals are first divided into five bands containing delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–31 Hz) and gamma (31–50 Hz) wavebands by using digital band-pass filters whose corresponding cut-off frequency is performed

according to the division standard in [1], and the segmentation process is completed by utilizing the local time window method (0.5s). Then, the comprehensive information of these rhythm signals about brain activities is obtained by continuous wavelet transform method (CWT) [14], and the time-frequency expression ($\mathbf{tf}_t \in \mathbb{R}^{N \times C \times H \times W}$) of the EEG data is obtained by convolution layer, normalization layer, nonlinear activation function and max pooling layer operations (defined as $\text{Conv}(\cdot)$), as described in Eq. 3.

$$\mathbf{tf}_t = \text{Conv} \left(\text{Cat}^{(t,i)} \left[\text{Cat}^{(ch,j)} \left(\frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} \text{div}^{(i,j)}(f^{(i)}(t)) \bar{\psi}_{s,a} dt \right) \right] \right) \quad (3)$$

where $f^{(i)}(t) \in \mathbf{R}^N$ denotes the $i^{\text{th}} \in \{1, 2, \dots, M\}$ segment signal of EEG, $\text{div}^{(i,j)}(\cdot)$ represents the decomposition of the signal $f^{(i)}(t)$ into j sub-bands. The inner product process is the principle of CWT, $\text{Cat}^{(x,y)}$ implies that the y^{th} tensor data is concatenated according to the x^{th} dimension. The procedure of fusing time-frequency information directly in multiple frequency bands may ignore the spatial distribution of the electrode. Therefore, multi-channel EEG signals are executed across the channel direction through the channel-by-channel convolutional operations, which aims at capturing spatial dynamic correlation characteristics ($\mathbf{S}_t \in \mathbb{R}^{N \times C \times H \times W}$) among brain regions, as shown in Eq. 4.

$$\mathbf{S}_t = \text{Cat} \left(\text{Append} \left(g \left(\mathbf{W} f^{(i)}(t) + \mathbf{b} \right) \right) \right) \quad (4)$$

where $g(\cdot)$ implies 1-D convolutional operation, tensor \mathbf{W} and \mathbf{b} denotes convolution weights and bias, $\text{Append}(\cdot)$ means adding each element in turn. With the above configurations, the mixed sample ($\mathbf{MS}_t \in \mathbb{R}^{N \times 2C \times H \times W}$), which is composed of \mathbf{tf}_t and \mathbf{S}_t , are represented as 4-D tensor, where (C, H, W) is the number of channels and the resolution of the feature map respectively.

2.4 Multiscale Feature Fusion

Algorithm 2 Multiscale Feature Fusion method

Input: the mixed sample \mathbf{MS}_t

Parameters: \mathbf{W}

Output: Multi-channel EEG sub-bands features (\mathbf{SSF}) about multiple brain regions

- 1: **Input** = $\text{Tok}(\mathbf{MS}_t) + \text{Seg}(\mathbf{MS}_t) + \text{Pos}(\mathbf{MS}_t)$ # Generate embedding
 - 2: **for** $n = 1 \rightarrow \text{NUM}$ **do**
 - 3: $\mathbf{Q}_n = \mathbf{W}_{Q_n} \times \text{Norm}(\mathbf{MS}_{ti})$
 - 4: $\mathbf{K}_n = \mathbf{W}_{K_n} \times \text{Norm}(\mathbf{MS}_{ti})$
 - 5: $\mathbf{V}_n = \mathbf{W}_{V_n} \times \text{Norm}(\mathbf{MS}_{ti})$ # Calculate query (\mathbf{Q}), key (\mathbf{K}), value (\mathbf{V})
 - 6: $\mathbf{A}_n = \mathbf{XSA}(\mathbf{Q}_n, \mathbf{K}_n, \mathbf{V}_n)$ # Forward temporal/spatial self-attention block with eeg data
 - 7: **end for**
 - 8: $\mathbf{SSF} = \mathbf{A}_{\text{NUM}} + \text{MLP}(\text{Norm}(\mathbf{A}_{\text{NUM}}))$ # Conduct residual operation
 - 9: **return** \mathbf{SSF}
-

Given the mixed samples, the multiscale feature fusion model embeds them into feature vectors and extracts the time-frequency-space features of multi-channel EEG signals. We introduce the embedding layers in phase 2 to better describe the semantic and location information of EEG and improve the EEG transformer encoder to pay attention to seizing the global and local information of long-term EEG signals. Concretely, the normalized operation, the position mark of multiple MS_t and the order inside it are added to attain the input of the network by conducting token, segment and position embedding steps, respectively, as showed in Eq. 5.

$$\text{Input} = \text{Tok}(MS_t) + \text{Seg}(MS_t) + \text{Pos}(MS_t) \quad (5)$$

Note that the addition here is bitwise addition. Tok, Seg and Pos correspond to the above three operations. Then tensor-patches in MS_t are vectorized and carried to the network to calculate temporal and spatial self-attention block described in [15], as showed in Algorithm 2.

2.5 Motion Intention Estimation

We construct MI classification module based on attention mechanism and knowledge rules, which consists of two parts: domain knowledge expressed by first-order logic formula and learnable parameters weights, as shown in Fig. 2.

Part1: Domain Knowledge for estimating the MI type

$false \leftarrow \exists(right_electrodes_energy_inc(X, Z_1) \wedge left_electrodes_energy_inc(Y, Z_2))$
 $false \leftarrow \exists(right_electrodes_energy_dec(\bar{X}, Z_1) \wedge left_electrodes_energy_dec(\bar{Y}, Z_2))$
 $left_move(X, Y) \leftarrow right_electrodes_energy_dec(\bar{X}, Z_1) \vee left_electrodes_energy_inc(Y, Z_2)$
 $right_move(X, Y) \leftarrow right_electrodes_energy_inc(X, Z_1) \vee left_electrodes_energy_dec(\bar{Y}, Z_2)$

Part2: Contribution weights for different channels elements

$element_weight(channels, delta, score)$	$element_weight(channels, theta, score)$
$element_weight(channels, alpha, score)$	$element_weight(channels, beta, score)$
$element_weight(channels, gamma, score)$	

Notice: $left_move(X, Y)$ denotes **SSF** X corresponds to MI classification Y, other analogies can be inferred "Λ", "V" represents conjunction and disjunction. "∃" is the meaning of existing, and "←" implies implication.

Fig. 2. MIR rules of the knowledge base *KB*. Part 1): the domain knowledge for MI type is to revise pseudo-labels and inference. 2) the contribution weights for different channels.

3 Results

We adopt the 2008 BCI competition IV-2a EEG dataset (<https://www.bbc.de/competition/iv/>) including 9 subjects with a 250 Hz sampling rate and band-pass filtered from 0.5 to 100 Hz, which consists of four MI tasks: imagine left hand (class 1), right hand (class

2), foot (class 3) and tongue (class 4) movements. And the preprocessing operations in this article refers to eliminating the wrong experiments and using digital band-pass filters of 1–50 Hz. In the experiments, all methods with self-teaching plan are set to the same number of modules using the default hyperparameters in the source code.

The following experimental tests of the proposed method (SSE-ABL) are designed to conduct: 1) Compared with the current mainstream algorithms containing transformer (TF) [16], BERT [17] and MEET [15] models to verify the performance of the SSE-ABL. 2) By changing proportion of training dataset size to verify the robustness of the SSE-ABL under the interference of different small-sample data, that is, the extrapolation ability. 3) The visualization of the underlying information flow of EEG decoding by the proposed method is displayed.

Table 1. Performance comparison of SSE-ABL method to state-of-the-art methods as the proportion change of training dataset in term of accuracy metrics (%) for BCI IV-2a dataset.

Method	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	AVE
TF-10	74.10	73.27	73.41	74.69	70.63	76.07	89.27	81.61	75.55	76.51
BERT-10	72.30	69.72	76.01	71.27	65.12	72.91	82.97	82.42	75.31	74.23
MEET-10	78.63	66.80	73.62	74.70	67.91	75.92	86.73	82.51	73.29	75.57
SSE-ABL-10	79.22	76.01	78.90	78.43	72.44	78.66	89.19	88.05	79.40	80.03
TF-50	79.03	76.59	79.24	77.48	73.53	79.36	89.38	88.26	80.54	80.38
BERT-50	79.85	75.93	78.83	77.56	73.05	78.79	89.83	88.73	79.76	80.26
MEET-50	80.20	76.26	78.93	78.59	73.19	79.56	90.51	89.19	79.83	80.70
SSE-ABL-50	80.78	76.68	79.34	78.91	73.62	79.95	90.56	92.82	81.07	81.53
TF-100	80.91	75.36	79.12	78.68	74.13	77.38	91.21	91.32	80.09	80.91
BERT-100	81.65	77.13	79.92	79.16	73.59	80.34	89.72	92.68	81.39	81.73
MEET-100	81.42	77.98	79.53	78.51	73.06	79.98	90.44	92.46	81.38	81.64

As shown in Table 1 are the accuracy results of experiments 1 and 2, which consists of three parts in the case of 10%, 50% and 100% labeled data as the training set. And there are 9 subjects and average accuracy in the horizontal direction. It can be seen from the first and second part that SSE-ABL has obvious advantages in the individual accuracy and overall average accuracy of different subjects, up to 89.19%, 80.03% and 92.82%, 81.53% respectively. By comparing the first two parts, can be found that when the labeling rate is low, the benefits of using both unlabeled data and MI knowledge are much higher during EEG decoding. Further comparison with the third part shows that the performance can even be competitive the methods with the 100% training set.

Then, the visualization results of experiment 3 is exhibited in Fig. 3, which is the analysis of FC_1 , FC_3 , CP_1 , CP_3 and FC_2 , FC_4 , CP_2 , CP_4 channels (corresponding to the left and right brain regions respectively) when imagining a left-handed task. And there are also three parts: the time domain EEG signal (P1), the time-frequency analysis (P2) and visualization analysis of the SSE-ABL (P3). From the time domain perspective

shown in the P1, it seems that there is not much discrimination that can separate specific motor imagery tasks. However, it can be found in the P2 that the energy distribution on the left and right sides of the brain can be clearly found by the data representation method proposed in this paper, which can be distinguished by the brightness of the color. At the same time, there is also a phenomenon that the energy of the right brain region is lower than that of the left brain region, which is consistent with biological cognition. Moreover, it can be observed in the P3 that the high energy component of EEG signals is surrounded by a large amount of dark red, which is the information that the SS-ABL algorithm focuses on. And the contribution rates of the above channels to this recognition task are marked under the channels, respectively, which gives us an intuitive understanding of the EEG signal decoding process.

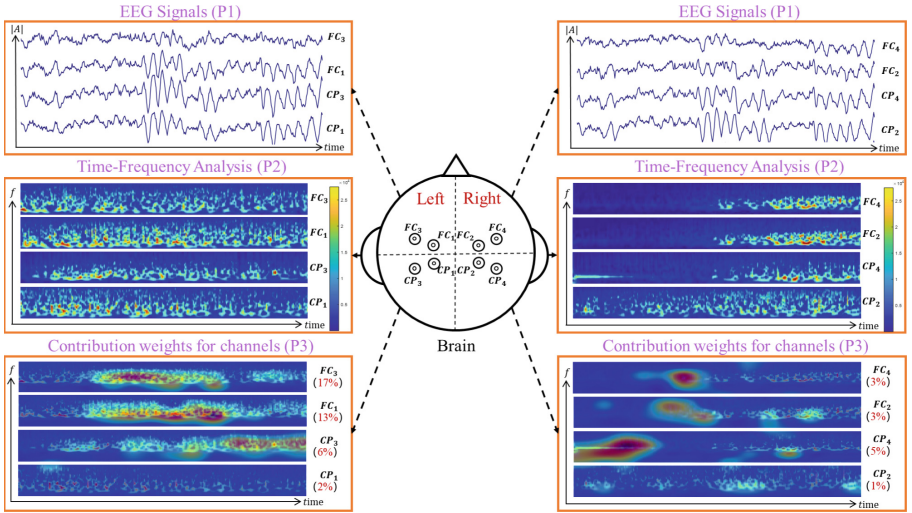


Fig. 3. Visual analysis of the typical channels (FC_{1-4} and CP_{1-4}) when imaging left hand movements through the SSE-ABL method.

4 Conclusion

A novel small-sample method with EEG signals based on abductive reasoning is studied for MI recognition. This method solves the problem of low precision and poor robustness ability of EEG decoding under a small amount of labeled data. The multiscale feature fusion module based on self-attention mechanism improves the ability of adaptive feature mining by capturing time-frequency-space information cover the whole brain region, which realizes the enhancement of abstract features. An effective knowledge base module of motor imagery is constructed and symbolized, which can upgrade the model space under this constraint by mining the potential facts of large-scale unlabeled EEG signals. Through the comparison experiments with other mainstream inversion methods, it can be founded that our method with 10% and 50% labeled EEG data can reach the accuracy

of 80.03% and 81.53%, achieving the high precision standard of EEG decoding. In the future, we will consider a fine-grained visualization method of EEG signals in the case of partial data damage, such as signal coupling.

References

1. Li, D.L., Xu, J.C., Wang, J.H., Fang, X.K., Ji, Y.: A multi-scale fusion convolutional neural network based on attention mechanism for the visualization analysis of EEG signals decoding. *IEEE Trans. Neur. Sys. Rehabil. Eng.* **28**(12), 2615–2626 (2020)
2. Zhang, X., Yao, L.N., Wang, X.Z., et al.: A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers. *J. Neur. Eng.* **18**(3), 031002 (2021)
3. Valizadeh, S.A., Riener, R., Elmer, S., et al.: Decrypting the electrophysiological individuality of the human brain: identification of individuals based on resting-state EEG activity. *NeuroImage* **197**, 470–481 (2019)
4. Li, D.D., Xie, L., Chai, B., Wang, Z., Yang, H.: Spatial-frequency convolutional self-attention network for EEG emotion recognition. *Appl. Soft Comput.* **122**, 108740 (2022)
5. Jana, G.C., Sabath, A., Agrawal, A.: Capsule neural networks on spatio-temporal EEG frames for cross-subject emotion recognition. *Biomed. Signal Process. Control* **72**, 103361 (2022)
6. Ye, Z., Sun, T., Shi, S., et al.: Local-global active learning based on a graph convolutional network for semi-supervised classification of hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **20**, 1–5 (2023)
7. Zhou, C.H., Zou, L.Y.: Semi-supervised Gaussian processes active learning model for imbalanced small data based on tri-training with data enhancement. *IEEE Access* **11**, 17510–1724 (2023)
8. Teng, X.B., Tian, X., Doelling, K., Poeppel, D.: Theta band oscillations reflect more than entrainment: behavioral and neural evidence demonstrates an active chunking process. *Eur. J. Neurosci.* **48**(8), 2770–2782 (2018)
9. Lechat, B., Hansen, K.L., Melaku, Y.A., et al.: A novel electroencephalogram-derived measure of disrupted delta wave activity during sleep predicts all-cause mortality risk. *Ann. Am. Thorac. Soc.* **19**(4), 649–658 (2022)
10. Musaeus, C.S., Engedal, K., Høgh, P., et al.: EEG theta power is an early marker of cognitive decline in dementia due to Alzheimer's disease. *J. Alzheimers Dis.* **64**(4), 1359–1371 (2018)
11. Huang, Y.X., Dai, W.Z., Cai, L.W., et al.: Fast abductive learning by similarity-based consistency optimization. *Adv. Neural. Inf. Process. Syst.* **34**, 26574–26584 (2021)
12. Dai, W.Z., Xu, Q., Yu, Y., et al.: Bridging machine learning and logical reasoning by abductive learning. *Adv. Neural Inf. Process. Syst.* **32** (2019)
13. Zhou, Z.-H.: Abductive learning: towards bridging machine learning and logical reasoning. *Sci. China Inf. Sci.* **62**(7), 1–3 (2019). <https://doi.org/10.1007/s11432-018-9801-4>
14. Shi, X., Qin, P., Zhu, J., et al.: Feature extraction and classification of lower limb motion based on sEMG signal. *IEEE access* **8**, 132882–132892 (2020)
15. Shi, E.Z., Yu, S.G., Kang, Y.Q., et al.: MEET: Multi-band EEG Transformer. *arXiv preprint* (2023)
16. Song, Y., Jia, X., Yang, L., et al.: Transformer-based spatial-temporal feature learning for EEG decoding. *arXiv preprint arXiv:2106.11170* (2021)
17. Devlin, J., Chang, M.W., Lee, K., et al.: Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018)