# An Adaptive Channel Selection and Graph ResNet Based Algorithm for Motor Imagery Classification

Yongquan Xia, Jianhua Dong, Duan Li*, Keyun Li, Jiaofen Nan, Ruyun Xu
School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, Henan, China

***Abstract*—In Brain-Computer interface (BCI) applications, achieving accurate control relies heavily on the classification accuracy and efficiency of motor imagery electroencephalogram (EEG) signals. However, factors such as mutual interference between multi-channel signals, inter-individual variability, and noise interference in the channels pose challenges to motor imagery EEG signal classification. To address these problems, this paper proposes an Adaptive Channel Selection algorithm aimed at optimizing classification accuracy and Information Translate Rate (ITR). First, C3, C4, and Cz are selected as key channels based on neurophysiological evidence and extensive experimental studies. Next, the channel selection is fine-tuned using spatial location and absolute Pearson correlation coefficients. By analyzing the relationship between EEG channels and key channels, the most relevant channel combination is determined for each subject, reducing confounding information and improving classification accuracy. To validate the method, the SHU Dataset and the PhysioNet Dataset are used in experiments. The Graph ResNet classification model is employed to extract features from the selected channel combinations using deep learning techniques. Experimental results show that the average classification accuracy is improved by 5.36% and 9.19%, and the Information Translate Rate is improved by 29.24% and 26.75%, respectively, compared to a single channel combination.**

***Keywords*—Brain-Computer Interface; motor imagery; channel selection; deep learning; graph convolutional neural network**

## I. Introduction

Brain-Computer Interface (BCI) systems allow for direct communication with the outside world without relying on the brain's typical output pathways [1,2]. Within the field of BCI, motor imagery is one of the most commonly utilized paradigms. Motor imagery (MI) involves mentally simulating the movement of a limb without actually moving it [3]. BCI systems can capture EEG signals from the brain during these imagined movements through electrodes, and thereby enable control over external devices [4]. With their potential applications in fields such as rehabilitation and medical care, communication security, and environmental protection, BCI systems have a wide range of possible uses [5].

Deep learning [6] is a widely used approach in the field of BCI. Compared to shallow learning models like traditional machine learning, deep learning uses neural network architecture with multiple complex network layers that perform varying functions. This leads to a significant improvement in the quantity and quality of feature extraction and recognition. Many studies have applied deep learning methods such as convolutional neural networks (CNN), long short-term memory (LSTM), and deep Boltzmann machine (DBM) for motor imagery classification, achieving good recognition results. However, these methods overlook the rich topological relationships between electrodes by treating EEG data as simple two-dimensional data. In contrast, the utilization of graph convolutional neural networks for EEG signal classification addresses this issue.

Different regions of the human brain have distinct functions. For instance, the motor cortex is responsible for motor functions, while the parietal lobe processes various sensory information like touch, smell, and taste. However, most current studies have used EEG data from all channels to extract features, which inevitably introduces redundancy. Additionally, the brain's electrical activity may vary among individuals. Therefore, selecting task-specific signal recording sites can reduce preparation time and improve user comfort in nonclinical BCI applications [7]. By choosing the optimal channel, the impact of noise interference can be minimized, and the computational costs of processing high-dimensional data can be reduced.

The objective of this study is to address the issue of individual variability and to decrease the computational complexity involved in processing high-dimensional EEG data. The main contributions of the paper are as follows:

- An Adaptive Channel Selection algorithm is proposed. For different individuals, the method selects all or partial channels as input to the BCI system by itself, and the partial channels are selected based on the correlation between channels and spatial location.
- A residual-based graph neural network is used to decode the MI signal.
- This method achieves good results on the SHU dataset and the PhysioNet dataset.

The main part of this paper consists of five sections. The first section is the introduction, which provides the background, significance, and methodology of the motor imagery study, as well as the reasons for conducting channel selection. The second section is the related work, where research related to feature extraction and classification algorithms of EEG signals is discussed. Section III presents the core part of the paper, introducing the channel selection algorithm and the Graph ResNet model. Section IV presents the experimental testing of the algorithms proposed in Section III, along with the corresponding results and analysis. Finally,

Section V concludes the research approach and provides future outlook.

## II. Related Works

### A. Feature Extraction

The method of feature extraction method has a significant impact on the performance of BCI systems. Popular methods include CNN and RNN [8-13]. In [11], a combination of convolutional and recurrent neural networks achieved an accuracy of 98.3% on the PhysioNet dataset. However, these methods only considered regular data in Euclidean space and did not take into account the topological relationships between electrodes. To address the issue of non-Euclidean space data being unable to be convolved, the first graph convolutional neural network (GCN) was proposed in [14] and applied to non-Euclidean space structured data. In [15] and [16], GCN was applied to MI decoding, achieving the highest accuracy of 93.056% and 98.08% on the PhysioNet dataset, respectively, demonstrating the superiority of GCN in processing EEG signals. However, in order to fully utilize the topological relationship between electrode channels, all channel data are typically used as network inputs, resulting in a large amount of redundant data [17].

### B. Channel Selection

EEG signals are acquired using multi-channel electrodes, but using all channels as network input can consume more computational resources and degrade system performance due to noise in some channels. Therefore, channel selection is necessary [17]. One simple method is to select data from the three channels (C3, C4, and Cz) where potential changes during motor imagery are mainly concentrated [18]. Additionally, the StEEGCS algorithm, proposed in [19], utilizes EEG shapelet-transformed for EEG channel selection. It selects the top-k EEG channels by solving a logistic loss-embedded minimization problem while simultaneously learning EEG shapelets, hyperplanes, and channel weights. and in [20], a Spatiotemporal-Filtering-Based channel selection (STECS) method was introduced to extract discriminative information from EEG signals. The STECS method was able to achieve the same classification performance as the full channel by using only half of the number of channels. However, these methods lack a neurophysiological basis. In [21], it was assumed that the channels associated with MI should contain public information, and channels were selected based on inter-channel correlation. An iterative multi-objective optimization channel selection (IMOCS) algorithm was proposed in [7] that selects optimal channels using anatomical and functional correlations of EEG channels, but it does not consider the spatial distribution of electrodes.

Subject-independence can reduce the time of data preprocessing in BCI systems. However, the investigation results showed [22] that subject-specific classification accuracy is higher than that of subject-independent BCI systems. This suggests that the optimal channel varies across subjects.

In summary, the paper addresses the problems of data redundancy and individual variability faced by using deep neural networks for EEG signal classification. An adaptive Channel Selection is proposed, which considers both the neurophysiological basis and the correlation between channels to select the optimal channel for each subject individually.

## III. Materials and Methods

### A. Overview

The framework of this paper is shown in Fig. 1.

*1)* Computing the channel correlation and obtaining the graph Laplacian.

*2)* Selecting electrodes using an Adaptive Channel Selection algorithm.

*3)* Generating a graph representation of the channel correlation from the graph Laplacian.

*4)* Applying the residual graph neural network to decode EEG signals.

*5)* Determining the optimal channel selection scheme based on the test results.

The Adaptive Channel Selection algorithm automatically selects the combination of channels with a higher Information Translate Rate (ITR) based on the ITR obtained after training data from both schemes. Fig. 2 illustrates the algorithm framework, which includes one scheme that uses all channels in the original data and another that uses proposed channel selection algorithm. Channel selection is based on the following two criteria.

### B. Adaptive Channel Selection Algorithm

To minimize computational effort during graph pooling, the number of selected channels is chosen as integer powers of two since the number of channels is reduced by half each time.

*1) Neurophysiological basics and spatial location:* Motor imagery EEG refers to the spontaneous electrical activity of brain tissue that reflects the functional state of the brain. Different bands of EEG activity typically appear in different functional areas of the brain, and changes in their activity can reflect various brain states. Previous studies [23] have shown that during motor imagery, the phenomena of ERD and ERS in the mu (8-13 Hz) and beta (13-30 Hz) bands are concentrated in the C3 and C4 electrodes of the cerebral motor cortex. in addition to Cz, which also receives the influence of hand movements. Therefore, first selected the C3, C4, and Cz channels that are most relevant to motor imagery EEG activity. Next, selected another set of channels that are spatially closest to the selected channels, with the already selected channels being referred to as “Fixed Channels”.

*2) Absolute pearson coefficient:* For the remaining channels, they needed to be closely related to EEG activity in the motor cortex of the brain, and considering the variability among individuals, this work selected a unique set of channels for each subject based on the absolute Pearson correlation coefficient. These channels were referred to as “Free Channels”, and the three sets of channels were combined to form the input to the model. The composition of each part of the channel is shown in Fig. 3.

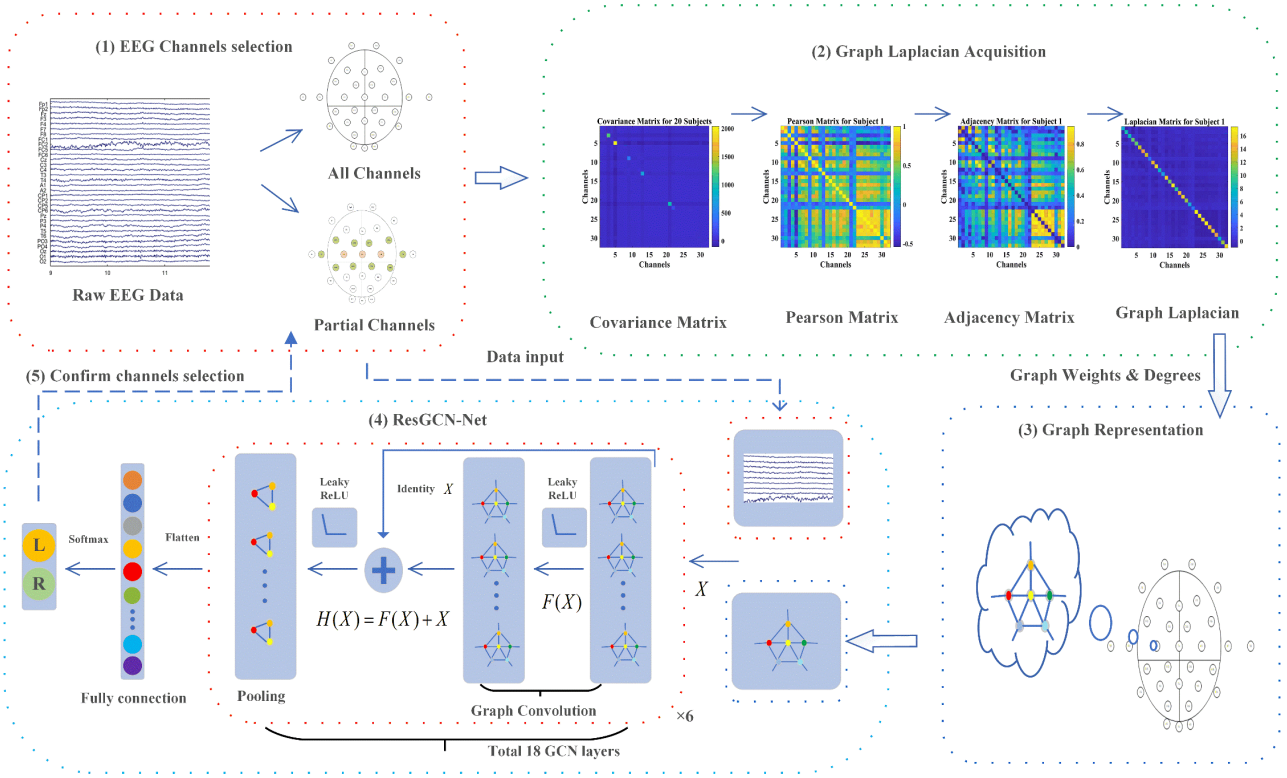



Fig. 1. Illustration of the proposed framework.

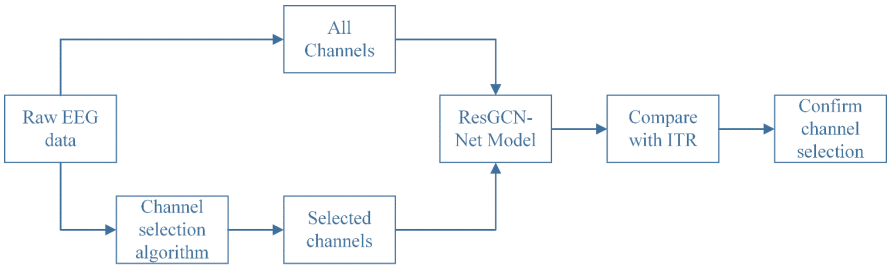



Fig. 2. Adaptive channel selection algorithm flowchart.

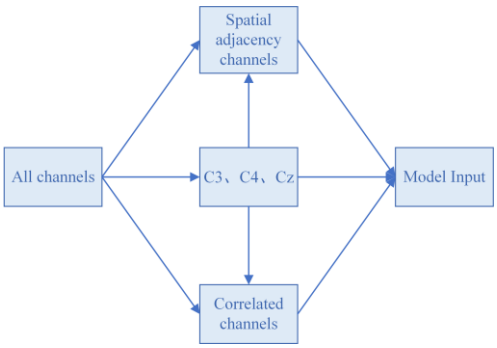



Fig. 3. Diagram of channel composition.

The absolute Pearson coefficient is a commonly used method for feature selection and extraction, as it measures the degree of correlation between two variables. In channel selection, this work aims to identify channels that are highly correlated with changes in motor cortex EEG signals. To achieve this, this work uses the absolute Pearson coefficient to evaluate the degree of correlation between each channel and the motor functional areas of the brain, and select the channels with high correlation [21].

The Pearson correlation coefficient between two variables is calculated as the quotient of the covariance and the standard deviation between the two variables.

$$\rho_{X,Y} = \frac{\mathrm{cov(X,Y)}}{\sigma_X \sigma_Y} \tag{1}$$

where $\rho_{X,Y}$ denotes the overall Pearson correlation coefficient of variables $X$ and $Y$, $\mathrm{cov(X,Y)}$ represents the covariance between variables $X$ and $Y$, and $\sigma_X \sigma_Y$ represents the product of the standard deviations of variables $X$ and $Y$. The estimation is based on the sample with respect to the variance and covariance.

$$\sigma_X^2 = \frac{1}{\mathrm{n}-1}\sum_{i=1}^{n}(X_\mathrm{i} - \bar{X})^2 \tag{2}$$

$$\mathrm{cov(X,Y)} = \frac{1}{\mathrm{n}-1}\sum_{i=1}^{n}(\mathrm{X_i} - \bar{X})\,(\mathrm{Y_i} - \bar{Y}) \tag{3}$$

Bring (2) and (3) into (1). The absolute Pearson coefficients of the samples are obtained as follows.

$$\mathbf{r} = \left| \frac{\sum_{i=1}^{n}(\mathrm{X_i}-\bar{X})(\mathrm{Y_i}-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \right| \quad (4)$$

With the EEG channels considered as variables and their corresponding data as observations, the Pearson correlation coefficients were calculated for all channels and the three channels most closely associated with motor imagery, C3, C4, and Cz. To ensure that negative correlations between channels and changes in EEG signals in the motor cortex of the brain were not overlooked, the absolute values of the Pearson correlation coefficients were taken. The correlation degree between all channels and the three motor imagery-related channels was calculated, and the channels with the highest correlation were selected as "Free Channels" for model input. The number of "Fixed Channels" and "Free Channels" selected for different datasets are shown in Table I.

### C. Graph ResNet

The graph convolution network can be divided into two categories: spectral graph convolution network and spatial graph convolution network [24]. The spectral method defines graph convolution in the spectral using the convolution theorem, while the spatial method aggregates the central node and its neighboring nodes by defining an aggregation function in the node domain. In this paper, the spectral graph convolution method is employed.

*1) Graph represents:* For an undirected graph, it can be represented as $G = \{V, E, A\}$, where $|V| = n$ denotes the number of nodes, $E$ denotes the set of edges, and $A$ denotes the adjacency matrix that defines the connectivity between nodes. The Laplacian matrix of the graph is denoted by $L = D - A$, where $D$ is a diagonal matrix. The normalized Laplacian matrix [24] is defined as

$$\mathrm{L} = I_n - D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \quad (5)$$

and since $L$ is a real symmetric matrix, an eigen-decomposition of $L$ yields $L = U\Lambda U^T$, where $U$ denotes the identity matrix of $L$, $\Lambda$ is the diagonal array of eigenvalues. The correlation matrix, Pearson matrix, absolute Pearson coefficient matrix, adjacency matrix, degree matrix, and Laplace matrix of subject 1 in SHU-Dataset are given in in Fig. 4.

*2) Graph convolution:* In the spectral, the graph convolution of signals $x_1, x_2$ is defined as

$$x_1 *_g x_2 = U\big((U^T x_1) \odot (U^T x_2)\big) \quad (6)$$

where $*_g$ represents the graph convolution operator and $\odot$ represents the Hadamard product. For an input signal $x$, the graph convolution operation through a convolution kernel filter $g \in \mathbb{R}^n$ is defined as

$$x *_G g = U(U^T x \odot U^T g) \quad (7)$$

Define $g$ as $g_\theta = \mathrm{diag}(U^T g)$,then the graph convolution operation for $x$ can be simplified to

$$x *_G g_\theta = U g_\theta U^T x \quad (8)$$

To parameterize the convolution kernel $g_\theta$, Chebyshev network (ChebyNet) [25] is used instead of the convolution kernel in the spectral. $g_\theta$ is defined as $g_\theta = \sum_{i=0}^{K} \theta_i T_i(\tilde{\Lambda})$, where $\tilde{\Lambda}$ can be represented as

$$\tilde{\Lambda} = \frac{2\Lambda}{\lambda_{max}} - I_n \quad (9)$$

TABLE I. DISTRIBUTION OF DIFFERENT CHANNEL TYPES

| Dataset | Number of Channels | Fixed Channels | Free Channels |
|---|---|---|---|
| **SHU Dataset** | 32 | | |
| | 16 | 13 | 3 |
| **Physionet MI Dataset** | 64 | | |
| | 32 | 21 | 11 |
| | 16 | 3 | 13 |

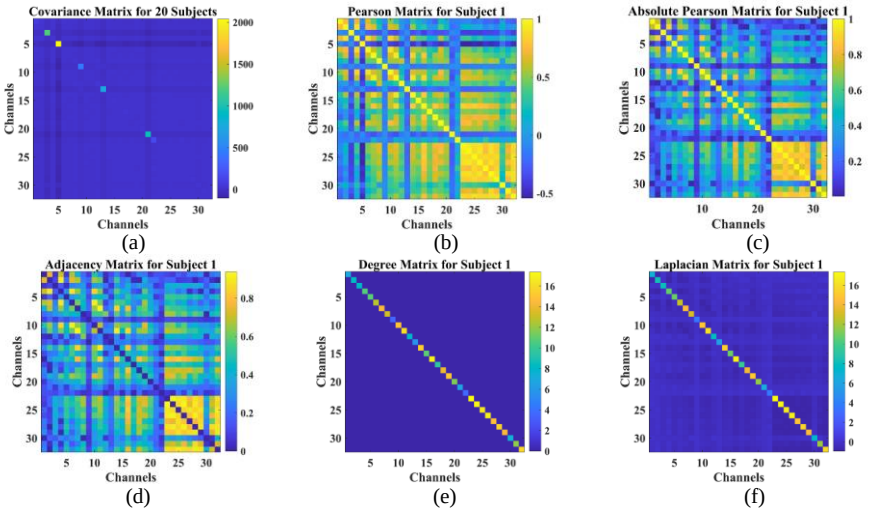



Fig. 4. The correlation matrix, Pearson matrix, absolute pearson coefficient matrix, adjacency matrix, degree matrix, and Laplace matrix of subject 1 in SHU-dataset.

and the Chebyshev polynomial is defined as $T_i(x) = 2xT_{i-1}(x) - T_{i-2}(x)$, where $T_0(x) = 1$ and $T_1(x) = x$. Therefore, the ChebyNet graph convolution operation is

$$x *_G g_\theta = \mathrm{U}\left(\sum_{i=0}^{K} \theta_i T_i(\tilde{\Lambda})\right) U^T x \quad (10)$$

Let $\tilde{L} = \frac{2L}{\lambda_{max}} - I_n$, then $T_i(\tilde{L}) = UT_i(\tilde{\Lambda})U^T$, and the ChebyNet graph convolution operation can be simplified as

$$x *_G g_\theta = \sum_{i=0}^{K} \theta_i T_i(\tilde{L})\, x \quad (11)$$

ChebyNet convolution does not require feature decomposition of the Laplacian matrix, and the convolution kernel has only K + 1 trainable parameters [25]. Therefore, the parameter complexity is significantly reduced.

*3) Graph pooling:* ChebyNet implements pooling operation using a complete binary tree. In the coarsening phase based on the Graclus multi-level clustering algorithm [26], the input feature tensor is divided into blocks of equal size. These blocks are then aligned based on the nature of complete binomial trees where each non-leaf node has 2 children. At each non-leaf node, the most matching feature block is selected for pooling using the greedy criterion, Pooling two nodes into one node. The pooling process is shown in Fig. 5.

*4) Residual learning:* Deep neural networks are known for their improved ability to fit nonlinear functions as the number of layers and neurons increases. However, simply stacking layers can lead to problems such as vanishing gradients, exploding gradients, and network degradation. To address these issues, Residual Learning proposed in [27]. This framework assumes that $H(X) = X$ represents the optimal solution mapping, and the general convolutional neural network is directly fitted with $H(X) = X$. In contrast, the residual network aims to fit the residual mapping, $F(X) = H(X) - X$. The optimal solution mapping is then given by $H(X) = F(X) + X$.

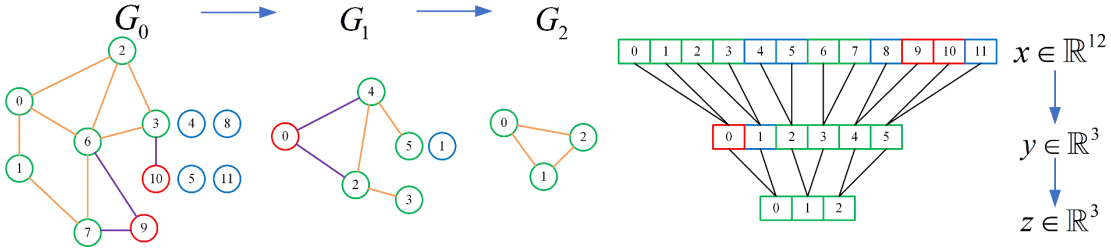



Fig. 5. ChebyNet implements pooling operators using complete binary trees[25].

## IV. Results and Discussion

### A. Dataset Description

*1) SHU dataset:* This work utilized the SHU Dataset, a publicly available motor imagery dataset from Shanghai University in 2022[28]. The dataset includes 25 subjects who underwent a total of 5 sessions every 2-3 days. Each session included 100 trials, which were automatically labeled by using EEGLAB with amplitudes greater than 100 $\mu$V. Experts then reviewed and eliminated any bad segments, resulting in a small number of missing trials in some sessions. The motor imagery tasks consisted of left-hand (L) and right-hand (R). EEG data were recorded using 32 electrodes based on the International 10-10 system with a 250 Hz sampling rate and a 4-second time window. Each subject had a total of 500 trials (5 sessions * 100 trials), and each trial included 1000 sampling points. The EEG data acquisition paradigm is shown in Fig. 6.

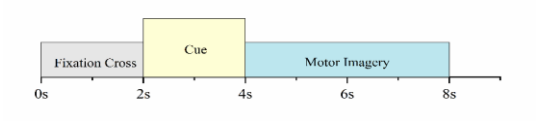



Fig. 6. Timing paradigm of SHU dataset.

*2) PhysioNet dataset:* The PhysioNet dataset included 109 subjects who performed motor imagery tasks using their left hand (L), right hand (R), both fists (B), and both feet (F). EEG data was recorded from 64 electrodes based on the International 10-10 system, with a sampling rate of 160 Hz and a time window of 4 seconds. Each subject completed 84 trials (3 runs * 7 trials * 4 tasks), with each trial consisting of 640 sampling points.

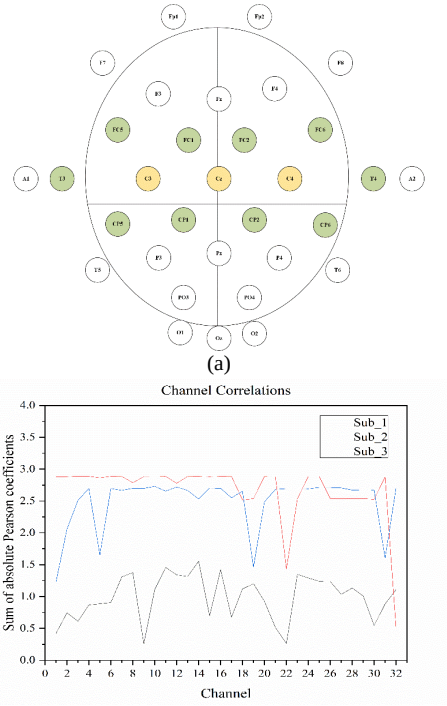
### B. Channel Selection Results

*1) SHU dataset:* The original SHU-Dataset had 32 channels. Firstly, three channels C3, C4, and Cz -were selected, and then 10 additional channels were selected based on their spatial location: FC1, FC2, FC5, FC6, T3, T4, CP1, CP2, CP5, and CP6. In total, 13 channels were selected and has been marked in Fig. 7(a). Finally, three Free Channels are selected based on the absolute Pearson coefficients, and a total of 16 channels were used as input for the model.

Taking the first three subjects in the dataset as an example, computing the sum of the absolute Pearson coefficients for all channels and C3, C4, and Cz channels. The computation results are presented in Fig. 7(b), and the channels that were selected based on this result are listed in Table II.

*2) PhysioNet dataset:* The original dataset consists of 64 channels, which are reduced to 32 and 16 channels respectively. In both cases, three channels (C3, C4, Cz) were first selected from the source channels, and the remaining channels were selected using the following method:

- For the 32 EEG channels, 18 channels were initially selected based on their spatial location, resulting in a total of 21 Fixed Channel as shown in Fig. 8. Then, 11 Free Channels were selected based on the absolute Pearson coefficients.
- For the 16-channel subset, due to the densely arranged 64 channels based on the international 10-10 system and the need for fewer channels, the remaining 13 Free Channels were all selected based on the absolute Pearson coefficients.




(a)

(b)

Fig. 7. (a) Shows schematic diagram of SHU Dataset Fixed Channels, (b) Shows the sum of the absolute Pearson coefficients between all channels and the C3, C4, and Cz channels for the first three subjects in the SHU-Dataset was calculated.

TABLE II. FREE CHANNELS SELECTED OF THE THREE SUBJECTS

| Subject | Added channels label | Added channels |
|---|---|---|
| **1** | 7、23、24 | F8、Pz、P3 |
| **2** | 3、6、17 | Fz、F7、A1 |
| **3** | 25、26、27 | P4、T5、T6 |

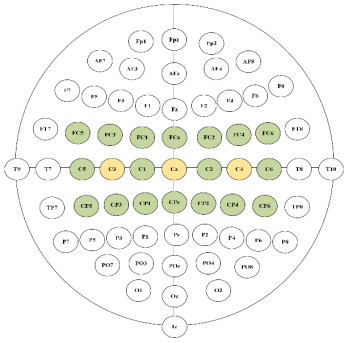


Fig. 8. Schematic diagram of physionet dataset fixed channels.

### C. Evaluation Metrics

To evaluate this algorithm, this work introduces the Information Translate Rate (ITR) [29] as an additional evaluation metric, in addition to the traditional precision metric. ITR is a standard method for measuring communication systems, expressing the amount of information transmitted per unit time. The bit rate depends on both speed and accuracy. The formula for calculating ITR is as follows

$$ITR = 60 \cdot \frac{log_2 M + P log_2 P + (1-P) log_2 \frac{1-P}{M-1}}{T_0} \quad (12)$$

Here, $T_0$ represents the prediction time of each sample in the model, $M$ represents the number of categories, and $P$ represents the classification accuracy. The unit of ITR is bits/min.

### D. Equipments

This work utilized a remote server that runs on the Ubuntu 18.04 operating system to conduct experiments. The experiments were conducted in a Python 3.6 environment using the TensorFlow 1.15 deep learning framework. The models were trained and evaluated using 1 NVIDIA Tesla V100-PCIE-32GB GPU and 1 Intel Xeon Processor (Skylake) 2.4GHz CPU. The system was equipped with 147GB RAM for system memory and had 6TB hard disk capacity.

### E. Model Parameters

The network architecture consists of 12 convolutional layers, each connected to a pooling layer every two layers. The dataset is split into 90% training data and 10% test data. A training cycle of 100 batches is utilized and the performance of the trained model is evaluated on the test set after each cycle. Evaluation metrics including ITR and accuracy are calculated, and finally the average value of the successive ITR is taken to reduce randomness. The model hyperparameters are shown in Table III.

TABLE III. MODEL HYPERPARAMETERS

| Hyperparameter | Value |
|---|---|
| **Chebyshev Order** | 3 |
| **Activation Function** | Leaky ReLu |
| **Batch size** | 1024 |
| **Epoch** | 50 |
| **Optimization Algorithm** | Adam |
| **Learning Rate** | 0.001 |
| **Dropout** | 0.5 |

### F. Experimental Results

*1) SHU dataset:* For this dataset, choosing to train the EEG dataset of the first 15 subjects separately, and the outcomes are presented in Fig. 9.

Table IV presents the results of motion imagery classification accuracy using all 32 channels, with an impressive accuracy of 83.98% reported in [16]. This outperforms the various benchmark classification methods used in [28], which achieved a maximum accuracy of 78.9%

on the cross-session task. After applying proposed method to halve the channels, the classification accuracy improved by 1.86%, the ITR improved by 14.4%, and the sample prediction time decreased by 22.7%. This result indicates that reducing the number of channels effectively reduces interference information, reduces data redundancy, and improves information transmission rate. In particular, some subjects who previously had poor classification accuracy and ITR, such as subjects 4, 5, 8, and 10, showed significant improvement after channel reduction. Using only half of the channel data, higher average classification accuracy and ITR were achieved compared to using all 32 channels. Furthermore, after applying the Adaptive Channel Selection algorithm, significant improvements were observed in classification accuracy and ITR compared to using all channels and only 16 channels. The classification accuracy improved by 5.36% and 3.5%, ITR improved by 29.24% and 12.97%, and single-sample prediction time was reduced by 12.56% and 13.13%, respectively.

*2) Physionet dataset:* For this dataset, training the EEG data of the first 20 subjects separately, and the experimental results were obtained by taking the mean values of the different subjects. These results are presented in Fig. 10.

Table V shows that with the input of all 64 channels data [16], the classification accuracy is improved by 1.26% and 8.21% compared to 32 and 16 channels, respectively, due to the availability of more extracted features. However, this leads to more data redundancy and the longest single-sample prediction time, resulting in a lower ITR compared to 32 channels. Although the ITR of 16 channels has only a 3.85% difference compared to 64 channels, the classification accuracy loss is higher due to less information available and less training time. By using the proposed Adaptive Channel Selection algorithm, the classification accuracy is improved by 0.98%, 2.24%, and 9.19% compared to 64, 32, and 16 channels, respectively, and the ITR is improved by 26.75%, 7.66%, and 22.05%, respectively.

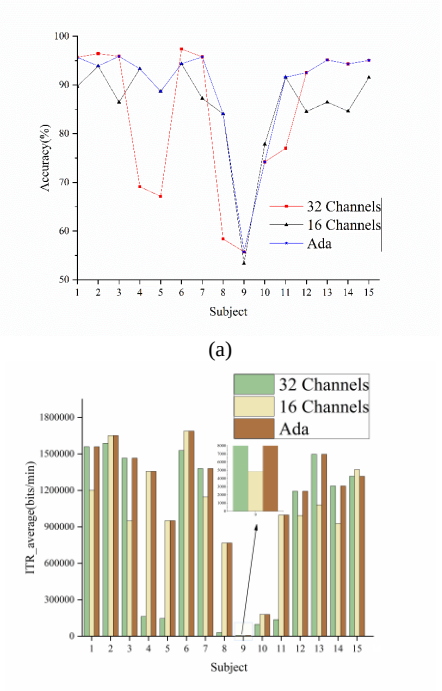
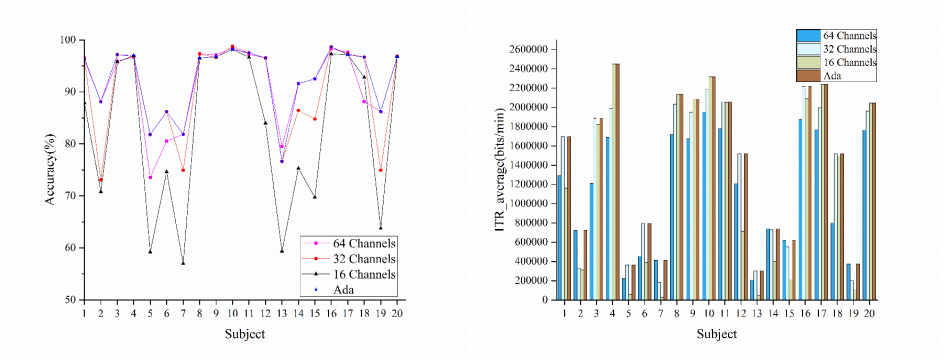



Fig. 9. (a) Shows the classification accuracy of the model under different channel strategies. (b) Shows the ITR of the BCI system under different channel strategies.

TABLE IV. SHU DATASET EXPERIMENTAL RESULT

| Subject | Channels Strategies | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | *32* | | | *16* | | | *Ada* | | |
| | *ACC (%)* | *ITR (bits/min)* | $T_0$ *(μs)* | *ACC (%)* | *ITR (bits/min)* | $T_0$ *(μs)* | *ACC (%)* | *ITR (bits/min)* | $T_0$ *(μs)* |
| **1** | 95.67 | 1558783 | 22 | 89.7 | 1201300 | 22 | 95.67 | 1558783 | 22 |
| **2** | 96.44 | 1588442 | 25 | 93.89 | 1651697 | 19 | 93.89 | 1651697 | 19 |
| **3** | 95.87 | 1466000 | 24 | 86.45 | 949636 | 18 | 95.87 | 1466000 | 24 |
| **4** | 69.1 | 162564 | 28 | 93.32 | 1356800 | 18 | 93.32 | 1356800 | 18 |
| **5** | 67.13 | 146810 | 28 | 88.67 | 951056 | 19 | 88.67 | 951056 | 19 |
| **6** | 97.38 | 1528246 | 28 | 94.34 | 1688184 | 21 | 94.34 | 1688184 | 21 |
| **7** | 95.76 | 1380068 | 25 | 87.23 | 1146226 | 19 | 95.76 | 1380068 | 25 |
| **8** | 58.39 | 30620 | 24 | 84.06 | 767902 | 19 | 84.06 | 767902 | 19 |
| **9** | 55.74 | 8438 | 24 | 53.32 | 4860 | 19 | 55.74 | 8438 | 24 |
| **10** | 74.2 | 97775 | 28 | 77.83 | 179420 | 22 | 74.2 | 179420 | 22 |
| **11** | 76.99 | 136505 | 23 | 91.58 | 997918 | 19 | 91.58 | 997918 | 19 |
| **12** | 92.54 | 1193187 | 24 | 84.53 | 993657 | 18 | 92.54 | 1193187 | 24 |
| **13** | 95.12 | 1497135 | 22 | 86.47 | 1081009 | 18 | 95.12 | 1497135 | 22 |
| **14** | 94.32 | 1236212 | 25 | 84.61 | 927602 | 19 | 94.32 | 1236212 | 25 |
| **15** | 95.03 | 1315861 | 24 | 91.55 | 1371521 | 19 | 95.03 | 1315861 | 24 |
| **Average** | 83.98 | 889776 | 24.93 | 85.84 | 1017919 | **19.27** | **89.34** | **1149910** | 21.8 |




(a) (b)

Fig. 10. (a) Shows the classification accuracy of the model under different channel strategies. (b) Shows the ITR of the BCI system under different channel strategies.

TABLE V. PHYSIONET DATASET EXPERIMENTAL RESULTS

| Channels Strategies | Acc (%) | ITR (bits/min) | $T_0$ (μs) |
|---|---|---|---|
| **64** | 91.53 | 1123454 | 33.8 |
| **32** | 90.27 | 1322665 | 30.1 |
| **16** | 83.32 | 1166775 | **27.25** |
| **Ada** | **92.51** | **1424018** | 29.75 |

TABLE VI. PERFORMANCE OF DIFFERENT METHODS ON PHYSIONET DATASET

| Work | Channels Strategies | Average Acc(%) |
|---|---|---|
| **This work** | **Ada** | **92.51** |
| **Jia et al. [16]** | All Channels | 91.53 |
| **Hou et al. [31]** | All Channels | **92.5** |
| **Ma et al. [32]** | All Channels | 68.2 |

In short, the proposed method utilizes less channel data while achieving comparable accuracy to some studies using full channel data, and higher ITR as shown in Table VI. Specifically, the method achieves an average classification accuracy of 92.51%, while in [16], using the same feature extraction method with full channel data, achieves 91.53% accuracy. The method is slightly better than the 92.5% average classification accuracy obtained in [30] using scout EEG source imaging (ESI) with a convolutional neural network (CNN) algorithm. Furthermore, it is much higher than the average classification accuracy of 68.20% obtained in [31] using Spatial and Temporal Recurrent Neural Networks for classification.

## V. CONCLUSION

This work proposes an Adaptive Channel Selection algorithm that automatically selects the optimal combination of EEG channels for each subject to maximize the information transmission rate of the BCI system. The channel combinations include all channels and some selected channels based on the spatial location of each channel in the dataset and the absolute Pearson coefficients of each channel with the key channels C3, C4, and Cz, which are commonly used in motor imagery experiments.

To extract features, this work uses a deep graph convolutional neural network with residual blocks added to prevent overfitting. Applying this method to the SHU dataset improved the classification accuracy by up to 5.36% and the information transfer rate (ITR) by up to 29.24% compared to using a single channel combination. On the Physionet dataset, the classification accuracy is improved by up to 9.19% and the ITR is improved by up to 26.75%. The proposed algorithm effectively addresses the problems of data redundancy and individual differences faced by deep neural networks in extracting features, and significantly improves the classification accuracy and ITR.

In summary, the study demonstrates the effectiveness of the proposed adaptive channel selection algorithm for decoding EEG-based motor imagery. The algorithm outperforms the single-channel combination scheme and has the potential to improve the generalization capability of deep neural networks for BCI applications. Future work will focus on exploring the generalization capability of this channel selection algorithm across different datasets and experimental conditions.

## REFERENCES

[1] Lance, B.J.; Kerick, S.E.; Ries, A.J.; Oie, K.S.; McDowell, K. Brain–Computer Interface Technologies in the Coming Decades. Proceedings of the IEEE 2012, 100, 1585–1599, doi:10.1109/JPROC.2012.2184830.

[2] Wolpaw, J.R.; Birbaumer, N.; McFarland, D.J.; Pfurtscheller, G.; Vaughan, T.M. Brain-Computer Interfaces for Communication and Control. Clin Neurophysiol 2002, 113, 767–791, doi:10.1016/s1388-2457(02)00057-3.

[3] Raza, H.; Rathee, D.; Zhou, S.; Cecotti, H.; Prasad, G. Covariate Shift Estimation Based Adaptive Ensemble Learning for Handling Non-Stationarity in Motor Imagery Related EEG-Based Brain-Computer Interface 2018.

[4] Lee, H.K.; Choi, Y.-S. A Convolution Neural Networks Scheme for Classification of Motor Imagery EEG Based on Wavelet Time-Frequecy Image. In Proceedings of the 2018 International Conference on Information Networking (ICOIN); January 2018; pp. 906–909.

[5] Al-Saegh, A.; Dawwd, S.A.; Abdul-Jabbar, J.M. Deep Learning for Motor Imagery EEG-Based Classification: A Review. Biomedical Signal Processing and Control 2021, 63, 102172, doi:10.1016/j.bspc.2020.102172.

[6] LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. Nature 2015, 521, 436–444, doi:10.1038/nature14539.

[7] Handiru, V.S.; Prasad, V.A. Optimized Bi-Objective EEG Channel Selection and Cross-Subject Generalization With Brain–Computer Interfaces. IEEE Transactions on Human-Machine Systems 2016, 46, 777–786, doi:10.1109/THMS.2016.2573827.

[8] Luo, J.; Shi, W.; Lu, N.; Wang, J.; Chen, H.; Wang, Y.; Lu, X.; Wang, X.; Hei, X. Improving the Performance of Multisubject Motor Imagery-Based BCIs Using Twin Cascaded Softmax CNNs. J Neural Eng 2021, 18, doi:10.1088/1741-2552/abe357.

[9] Ortiz-Echeverri, C.J.; Salazar-Colores, S.; Rodríguez-Reséndiz, J.; Gómez-Loenzo, R.A. A New Approach for Motor Imagery Classification Based on Sorted Blind Source Separation, Continuous Wavelet Transform, and Convolutional Neural Network. Sensors 2019, 19, 4541, doi:10.3390/s19204541.

[10] Densely Feature Fusion Based on Convolutional Neural Networks for Motor Imagery EEG Classification | IEEE Journals & Magazine | IEEE Xplore Available online: https://ieeexplore.ieee.org/document/8840855 (accessed on 23 March 2023).

[11] Zhang, D.; Yao, L.; Chen, K.; Wang, S.; Chang, X.; Liu, Y. Making Sense of Spatio-Temporal Preserving Representations for EEG-Based Human Intention Recognition. IEEE Trans. Cybern. 2020, 50, 3033–3044, doi:10.1109/TCYB.2019.2905157.

[12] Sakhavi, S.; Guan, C.; Yan, S. Learning Temporal Information for Brain-Computer Interface Using Convolutional Neural Networks. IEEE Trans. Neural Netw. Learning Syst. 2018, 29, 5619–5629, doi:10.1109/TNNLS.2018.2789927.

[13] Dai, G.; Zhou, J.; Huang, J.; Wang, N. HS-CNN: A CNN with Hybrid Convolution Scale for EEG Motor Imagery Classification. J. Neural Eng. 2020, 17, 016025, doi:10.1088/1741-2552/ab405f.

[14] Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y. Spectral Networks and Locally Connected Networks on Graphs 2014.

[15] Hou, Y.; Jia, S.; Lun, X.; Hao, Z.; Shi, Y.; Li, Y.; Zeng, R.; Lv, J. GCNs-Net: A Graph Convolutional Neural Network Approach for Decoding Time-Resolved EEG Motor Imagery Signals. IEEE Trans. Neural Netw. Learning Syst. 2022, 1–12, doi:10.1109/TNNLS.2022.3202569.

[16] Jia, S.; Hou, Y.; Shi, Y.; Li, Y. Attention-Based Graph ResNet for Motor Intent Detection from Raw EEG Signals 2020.

[17] Qiu, Z.; Jin, J.; Lam, H.-K.; Zhang, Y.; Wang, X.; Cichocki, A. Improved SFFS Method for Channel Selection in Motor Imagery Based BCI. Neurocomputing 2016, 207, 519–527, doi:10.1016/j.neucom.2016.05.035.

[18] Tabar, Y.R.; Halici, U. A Novel Deep Learning Approach for Classification of EEG Motor Imagery Signals. J. Neural Eng. 2017, 14, 016003, doi:10.1088/1741-2560/14/1/016003.

[19] Dai, C.; Pi, D.; Becker, S. Shapelet-Transformed Multi-Channel EEG Channel Selection. ACM Trans. Intell. Syst. Technol. 2020, 11, 58, doi:10.1145/3397850.

[20] Qi, F.; Wu, W.; Yu, Z.L.; Gu, Z.; Wen, Z.; Yu, T.; Li, Y. Spatiotemporal-Filtering-Based Channel Selection for Single-Trial EEG Classification. IEEE Transactions on Cybernetics 2021, 51, 558–567, doi:10.1109/TCYB.2019.2963709.

[21] Jin, J.; Miao, Y.; Daly, I.; Zuo, C.; Hu, D.; Cichocki, A. Correlation-Based Channel Selection and Regularized Feature Optimization for MI-Based BCI. Neural Networks 2019, 118, 262–270, doi:10.1016/j.neunet.2019.07.008.

[22] Lotte, F.; Guan, C.; Ang, K.K. Comparison of Designs towards a Subject-Independent Brain-Computer Interface Based on Motor Imagery. In Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; September 2009; pp. 4543–4546.

[23] Ha, K.-W.; Jeong, J.-W. Motor Imagery EEG Classification Using Capsule Networks. Sensors 2019, 19, 2854, doi:10.3390/s19132854.

[24] Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. IEEE Trans. Neural Netw. Learning Syst. 2021, 32, 4–24, doi:10.1109/TNNLS.2020.2978386.

[25] Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering 2017.

[26] Dhillon, I.S.; Guan, Y.; Kulis, B. Weighted Graph Cuts without Eigenvectors A Multilevel Approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 2007, 29, 1944–1957, doi:10.1109/TPAMI.2007.1115.

[27] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition 2015.

[28] Ma, J.; Yang, B.; Qiu, W.; Li, Y.; Gao, S.; Xia, X. A Large EEG Dataset for Studying Cross-Session Variability in Motor Imagery Brain-Computer Interface. Sci Data 2022, 9, 531, doi:10.1038/s41597-022-01647-1.

[29] Ming Cheng; Xiaorong Gao; Shangkai Gao; Dingfeng Xu Design and Implementation of a Brain-Computer Interface with High Transfer Rates. IEEE Trans. Biomed. Eng. 2002, 49, 1181–1186, doi:10.1109/TBME.2002.803536.

[30] Hou, Y.; Zhou, L.; Jia, S.; Lun, X. A Novel Approach of Decoding EEG Four-Class Motor Imagery Tasks via Scout ESI and CNN. J. Neural Eng. 2020, 17, 016048, doi:10.1088/1741-2552/ab4af6.

[31] Ma, X.; Qiu, S.; Du, C.; Xing, J.; He, H. Improving EEG-Based Motor Imagery Classification via Spatial and Temporal Recurrent Neural Networks. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); IEEE: Honolulu, HI, July 2018; pp. 1903–1906.