

TESI DOCTORAL

Expanding the Potential of Non-Invasive Brain-Computer Interfaces: Advancements in Upper Limb Movement Classification and Data Augmentation Strategies

Hao Jia

Director de tesi: Jordi Solé Casals, Cesar F. Caiafa, Zhe Sun

Programa de doctorat: Experimental Sciences and Technology

2023

 UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL DE CATALUNYA
Escola de Doctorat

ABSTRACT

This dissertation explores the potential of non-invasive brain-computer interfaces (BCIs) in connecting the brain with external devices. Electroencephalogram (EEG) signals are the signals acquired from the scalp of the brain with BCI system, which is used in the brain activity analysis. While previous studies on limb movement classification have primarily focused on left/right limb classification, this research highlights the significance of classifying different types of upper limb movements. These movements, including elbow extension, supination, and more, offer more active-evoked commands to BCIs. However, only a limited number of machine learning approaches can be utilized as baseline methods for the classification of limb movements.

The dissertation introduces the Filter Bank Task-Related Component Analysis (FBTRCA) method for binary classification of two motions. The FBTRCA framework consists of three main steps: spatial filtering, similarity measurement, and filter bank selection. Initially, spatial filtering, referred to as task-related component analysis, is employed to eliminate noise from EEG signals. Canonical correlation is then utilized to measure the similarity of the spatial-filtered signals and extract features. Correlation features are extracted from multiple low-frequency filter banks. The minimum-redundancy maximum-relevance technique is employed to select essential features from all correlation features. Finally, support vector machines are used to classify the selected features.

The FBTRCA method is further extended to address multi-class classification. In this case, the FBTRCA framework is similar to binary classification, but with an optimized spatial filter and similarity measure. The extended method achieves state-of-the-art performance in upper limb classification.

To enhance the temporal features in similarity measurement, the dissertation proposes a two-stage-training temporal-spectral neural network (TTSNet). By incorporating a convolutional neural network, TTSNet achieves improved classification performance compared to the multi-class FBTRCA method.

Furthermore, this dissertation tackles the challenge of acquiring an adequate amount of brain signals from impaired patients for Brain-Computer Interfaces (BCIs). In order to overcome the issue of data scarcity during neural network training, a data augmentation method based on the decomposition and recombination strategy is proposed and explored using small-size Alzheimer's datasets. This method begins by decomposing the original EEG signals within the training set into multiple intrinsic mode functions using empirical mode decomposition. Artificial signals are then generated by summing randomly selected IMF components. The enhanced training set encompasses both the original signals and the artificially generated signals, thereby enabling the training of neural network models while avoiding the problem of overfitting.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors, Jordi Solé-Casals, Zhe Sun, and Cesar F. Caiafa. Specifically, I am grateful to Professor Jordi Solé-Casals for his efforts in helping me to adapt to life in Catalonia and for his guidance throughout the process of preparing and finalizing the manuscripts of the articles. His support has been invaluable.

I would also like to extend my thanks to Professor Zhe Sun for his insightful suggestions regarding my potential future research area and for his assistance during the experiments. His input has greatly contributed to the quality of this thesis.

Additionally, I am grateful to Professor Cesar F. Caiafa for his suggestions on manuscript preparation. It was his advice, particularly his recommendation to "read more papers," that sparked my academic career when I was a master's student.

Special thanks go to Dr. Patrick Ofner for generously providing the public datasets that enabled me to focus on the development of algorithms in this thesis. His contribution has been instrumental in the success of my research.

I would also like to acknowledge the support and encouragement of my friends and family throughout this journey. Their unwavering belief in me has been a constant source of motivation.

Lastly, I extend my appreciation to the staff and faculty at the University of Vic - Central University of Catalonia for creating a conducive environment for academic growth and fostering a spirit of inquiry.

CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	ii
TABLE OF CONTENTS	iii
LIST OF FIGURES	v
LIST OF TABLES	vii
1. THESIS AS A COMPENDIUM OF PREVIOUSLY PUBLISHED WORKS . . .	1
2. INTRODUCTION.	2
2.1. Background.	2
2.2. Objectives.	5
2.3. Contributions of the doctoral student.	6
3. COPY OF PUBLISHED AND SUBMITTED WORKS.	7
3.1. Assessing the Potential of Data Augmentation in EEG Functional Connectivity for Early Detection of Alzheimer’s Disease	7
3.2. Graph Empirical Mode Decomposition-Based Data Augmentation Applied to Gifted Children MRI Analysis.	22
3.3. Improving Pre-movement Pattern Detection with Filter Bank Selection	35
3.4. Multi-class Classification of Upper Limb Movements with Filter Bank Task-related Component Analysis	71
3.5. Enabling Temporal-Spectral Decoding in Multi-class Single-side Upper Limb Classification	87
4. SUMMARY OF THE WORK CARRIED OUT WITHIN THE THESIS	116
4.1. Data augmentation on the small-size dataset	116
4.1.1. Pre-requisition	116
4.1.2. Data augmentation with a decomposition and recombination strategy	117
4.1.3. Deep-learning classifiers.	125
4.2. Single-side upper limb movement classification.	128
4.2.1. Pre-requisition	128
4.2.2. Spatial filter.	132
4.2.3. Correlation feature	134

4.2.4. Filter bank technique.	136
4.3. Conclusion	139
4.4. Future work.	140
BIBLIOGRAPHY.	142

LIST OF FIGURES

2.1	Flowchart of a brain-computer interface.	2
2.2	In the brain-computer interface, EEG signals can be analyzed to extract brain activities such as motor imagery, steady-state visual evoked potential, and movement-related cortical potential.	3
4.1	Evaluation of classification performance using the training set and testing set.	116
4.2	Data augmentation is a technique used to increase the size of the training set, allowing deep learning models to learn more effectively.	118
4.3	The IMFs decomposed by CEMD in the time domain (a) and the frequency domain (b)	120
4.4	Flowchart of SEMD. An example is shown using a two-channel time series. The blue block represents white noise or empty signals containing no data. After signal decomposition, the signals in the blue block are removed from the IMFs.	121
4.5	Data decomposition with MEMD. MEMD decomposes multi-channel signals into IMFs. The IMFs are located in different frequency bands, and within each decomposed channel, the k -th IMF corresponds to the same frequency band. In this figure, the IMFs are sorted in descending order in the frequency domain.	122
4.6	Decomposition of multi-channel time series with SEMD-MEMD. Multi-channel EEG signals of multiple trials are first concatenated along the time axis. The concatenated EEG signals are then decomposed with MEMD.	124
4.7	Signal decomposition of MRI data. In MRI data, the signals are represented as a graph. The graph edges measure the distances of the brain regions and the graph nodes denote the brain regions. The features of the brain regions are the signals in the nodes. After converting MRI data into a graph, the MRI data are decomposed in a graph format by GEMD. . . .	125
4.8	Flowchart depicting the recombination of IMFs to generate artificial signals in multi-channel EEG signals using MEMD.	126
4.9	The E2E block in BrainNet CNN.	127
4.10	The residual module with a skip connection in the ResNet.	127

4.11 EEG signal processing: From raw signals to three-dimensional EEG tensors. The EEG signals are acquired using 11 electrodes. The raw EEG signals around the movement onset are sliced and concatenated to form EEG tensors.	129
4.12 Determination of hyperparameters of classification models using nested cross-validation.	129
4.13 Concept of grand average MRCP.	130
4.14 Comparison between signals in the low-frequency band and the high-frequency band.	131
4.15 The flowchart of FBTRCA and TTSNet in the multi-class classification task.	131
4.16 Three frequency range settings in filter bank division. Because the differences of the grand average MRCPs are reflected in the low-frequency band, as shown in Figure 4.13(a) and Figure 4.14, the frequency range M3 is used in the filter bank division.	137
4.17 Feature selection using mutual information. Relevance refers to the mutual information between a single feature and the corresponding label. Redundancy represents the mutual information between two features. . .	138

LIST OF TABLES

4.1	Model structure of BrainNet CNN	127
4.2	An overall view on the modules of the proposed two methods	132
4.3	Model structure of the adopted CNN layers	136
4.4	Model structure of fully-connected layers in TTSNet	138

1. THESIS AS A COMPENDIUM OF PREVIOUSLY PUBLISHED WORKS

This doctoral thesis, in accordance with the corresponding report, authorized by the thesis supervisors and the body responsible for the doctoral program, is presented as a compendium of three previously published papers and a manuscript under review. The complete references of the articles that belong to the body of the thesis are the following:

* Same contribution

- **Hao Jia***, Zihao Huang*, Cesar F. Caiafa, Feng Duan, Yu Zhang, Zhe Sun, Jordi Solé-Casals, "Assessing the Potential of Data Augmentation in EEG Functional Connectivity for Early Detection of Alzheimer's Disease", *Cognitive Computation*, doi: 10.1007/s12559-023-10188-7, September, 2023. (*Impact Factor: 5.4, Citescore: 7.70, JCR Q2*)
- Xuning Chen, Binghua Li, **Hao Jia**, Fan Feng, Feng Duan, Zhe Sun, Cesar F. Caiafa, Jordi Solé-Casals, "Graph Empirical Mode Decomposition-Based Data Augmentation Applied to Gifted Children MRI Analysis", *Frontiers in Neuroscience*, doi: 10.3389/fnins.2022.866735, July 2022. (*Impact Factor: 4.30, Citescore: 6.80, JCR Q2*)
- **Hao Jia**, Zhe Sun, Feng Duan, Yu Zhang, Cesar F. Caiafa, and Jordi Solé-Casals, "Improving Pre-movement Pattern Detection with Filter Bank Selection", *Journal of Neural Engineering*, doi: 10.1088/1741-2552/ac9e75, November 2022. (*Impact Factor: 4.00, Citescore: 7.50, JCR Q2*)
- **Hao Jia**, Fan Feng, Cesar F. Caiafa, Feng Duan, Zhe Sun, Yu Zhang, and Jordi Solé-Casals, "Multi-class Classification of Upper Limb Movements with Filter Bank Task-related Component Analysis", *IEEE Journal of Biomedical and Health Informatics*, doi: 10.1109/JBHI.2023.3278747, May 2023. (*Impact Factor: 7.70, Citescore: 11.90, JCR Q1*)
- **Hao Jia**, Shuning Han, Cesar F. Caiafa, Feng Duan, Yu Zhang, Zhe Sun, Jordi Solé-Casals, "Enabling Temporal-Spectral Decoding in Multi-class Single-side Upper Limb Classification", has been submitted to *Engineering Applications of Artificial Intelligence* for publication (*Impact Factor: 8.00, Citescore: 12.30, JCR Q1*). My contributions cover all of this work, including idea, innovation point, method, experiments and manuscript writing. This article is wholly included in the thesis.

2. INTRODUCTION

2.1. Background

The COVID-19 pandemic has significantly altered daily routines across various fields and has had a profound impact on society [1]. One of the most widely recognized and effective measures to combat the pandemic is the practice of social distancing [2]–[4]. Reducing human-to-human interaction is strongly advised to prevent the further spread of the virus [5], [6]. However, given the uncertain duration of the pandemic, it is crucial for people’s daily lives to continue. In this context, human-robot interaction can be viewed as a safer alternative within society due to the reduced direct interaction between individuals [7]–[9].

Human-robot interaction refers to the interaction between humans and client devices such as service robots, smartphones, wearable devices, and more [10]–[12]. In the current post-pandemic period, human-robot interaction serves three key roles. Firstly, it provides an indirect approach for interaction among healthy individuals, thereby reducing the risk of spreading the pandemic [13]. Secondly, it facilitates contactless communication between patients and the outside world [14]. Lastly, it assists in the rehabilitation of individuals with wearable devices like exoskeletons [15]. However, despite the significant roles played by human-robot interaction in the post-pandemic society, not everyone can fully benefit from its convenience, particularly stroke survivors and individuals with disabilities. The most effective solution to this problem is to bridge the gap between the brain and external devices using brain-computer interfaces [16]–[18].

The brain-computer interface (BCI) encompasses four key steps, as illustrated in Figure 2.1. Firstly, brain signals are acquired from the scalp. Secondly, the acquired signals are analyzed to determine brain activities. Thirdly, the brain activities are converted into control commands. Finally, these control commands are transmitted to external devices. Through these four steps, the brain-computer interface effectively translates human intentions into control commands, enabling communication with external devices [19]–[23].

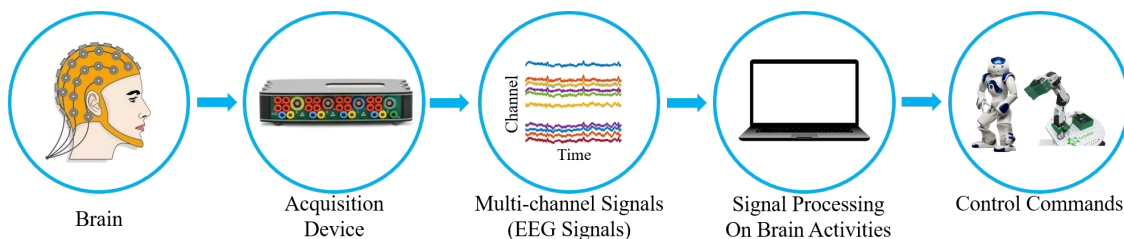


Figure 2.1: Flowchart of a brain-computer interface.

The electroencephalogram (EEG) signal is a type of non-invasive multi-channel brain signal acquired from the scalp with BCI [24]–[30]. It serves as an effective commu-

nication bridge, offering advantages such as good temporal resolution and a harmless acquisition process [31], [32]. In EEG signal processing, two popular brain activities are studied: steady-state visual evoked potential [33]–[38], and motor imagery [39]–[44]. The conversion of these brain activities into commands involves a classification task that distinguishes different brain states within each activity, as depicted in Figure 2.2. For instance, motor imagery processing typically involves binary classification between left and right limb movements, representing double-sided limb movements. However, this approach may not capture more precise movements, such as elbow flexion or hand closure of a single limb. On the other hand, steady-state visual evoked potential relies on classifying brain signal frequencies evoked by external stimuli. Nevertheless, this method has limitations, as subjects can easily become fatigued when required to sustain focus on the external visual stimulus.

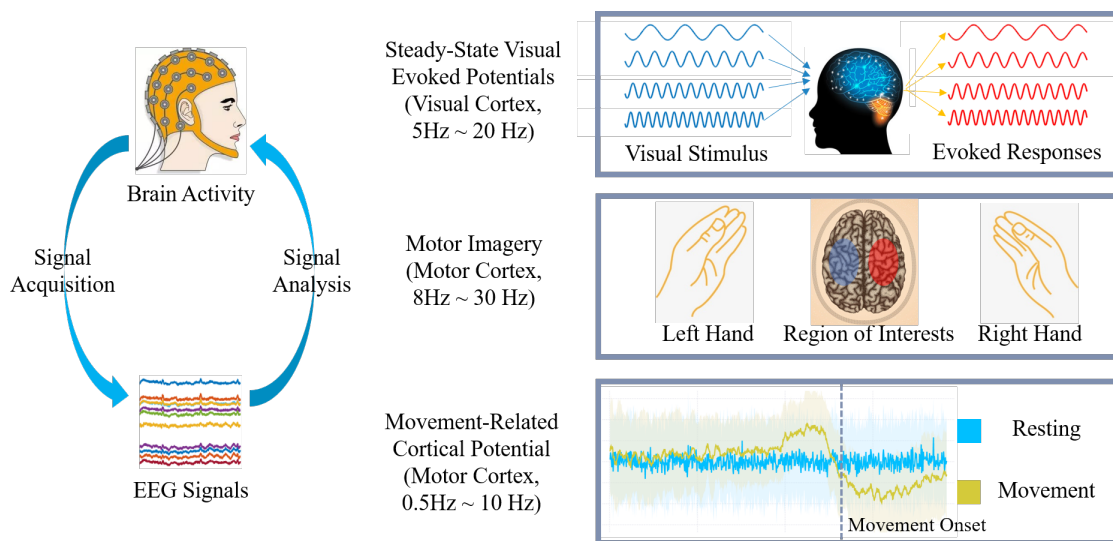


Figure 2.2: In the brain-computer interface, EEG signals can be analyzed to extract brain activities such as motor imagery, steady-state visual evoked potential, and movement-related cortical potential.

The movement-related cortical potential (MRCP) is a brain activity associated with limb movements and is characterized by low-frequency bands in EEG signals [45]–[51]. In comparison to motor imagery and steady-state visual evoked potentials, MRCP offers advantages such as (1) being evoked by the subject’s own movement intention without external stimuli and (2) being related to multiple movements on the same limb, specifically single-side limb movements. The classification of MRCP holds promise for providing more effective control commands in brain-computer interfaces. While numerous methods have been proposed for classifying double-side limb movements based on motor imagery, there is a scarcity of machine learning approaches suitable for multi-class classification of single-side limb movements.

The analysis of EEG signals not only enables the generation of control commands for external devices but also contributes to the early diagnosis of Alzheimer’s disease (AD). AD is a clinical syndrome that primarily affects the elderly, characterized by progressive

impairment of memory and cognitive function [52]. Despite ongoing research efforts and advancements, the underlying pathogenesis of AD remains incompletely understood, and the disease remains irreversible. Consequently, early diagnosis is crucial for implementing interventions and delaying symptom progression through medication [53].

Previous research has demonstrated the efficacy of deep learning methods, such as ResNet, for the early detection and diagnosis of AD using functional connectivity generated with EEG signals [54], [55]. However, training neural networks requires a substantial amount of data for parameter tuning [56]. Obtaining EEG signals for training purposes is relatively straightforward in healthy individuals with proper consent. However, acquiring EEG datasets from AD patients poses challenges due to their limited ability to perform tasks and their propensity for fatigue. As a result, EEG datasets for AD are typically small in size, making it difficult to fully leverage the potential of neural networks for classification. Therefore, it is imperative to maximize the utility of EEG signals acquired from AD patients when employing deep learning methods for classification.

2.2. Objectives

Based on the background descriptions, the overall objective of this doctoral thesis is to explore the potential of EEG signals in contemporary society. This includes the development of classification methods for EEG signals in controlling external devices and enhancing the performance of neural networks in the diagnosis of Alzheimer's disease (AD). To achieve this, the following specific objectives have been identified:

- Propose a data augmentation method to improve the performance of neural networks when working with small-size datasets in a BCI environment;
- Develop a binary single-side upper limb classification method to differentiate MRCP signals within a motion pair;
- Extend and adapt the binary classification method to handle multi-class classification tasks.

These objectives aim to address the challenges and limitations highlighted in the background, with a focus on advancing the field of EEG signal analysis and its applications in controlling external devices and diagnosing AD. By achieving these objectives, this research aims to contribute to the broader goal of utilizing EEG signals effectively in current societal contexts.

2.3. Contributions of the doctoral student

The contributions of the doctoral student in the processing of EEG signals, as evidenced by the various publications, are as follows:

- The doctoral student proposes a data augmentation method based on a decomposition and recombination strategy to address the challenges posed by small-size Alzheimer's disease (AD) datasets. The proposed method is evaluated on neural networks, including ResNet and BrainNet CNN, to assess its impact on performance.
- The doctoral student introduces a novel method called filter bank task-related component analysis (FBTRCA) to address the binary classification problem of single-side upper limb movements. This method incorporates the filter bank technique into the standard task-related component analysis method. Firstly, the MRCP signals are divided into multiple subbands. Then, spatial filter task-related component analysis is applied to each subband to reduce signal noise. Correlation features are extracted from the filtered signals, and these features from all subbands are fused. Furthermore, a mutual-information-based feature selection method is employed to reduce the feature dimension. Finally, a support vector machine classifier is trained to predict the signal class.
- Building upon the binary FBTRCA method, the doctoral student extends it to handle multi-class classification tasks. This extension overcomes the limitations of the binary FBTRCA method, which was restricted by the structure of the standard task-related component analysis. The multi-class FBTRCA method optimizes both the spatial filter and correlation features, demonstrating its superiority compared to previous machine learning and deep learning methods.
- In the context of multi-class FBTRCA, the doctoral student introduces a novel approach where features are extracted by calculating the correlations between the filtered MRCP signals and pre-trained signals, specifically the grand average MRCP. This correlation calculation is equivalent to the weighted sum of the MRCP signals, with the grand average MRCP acting as the weight. To enhance the weighting process, the doctoral student proposes replacing the weight with a shift-invariant weight and develops a two-stage training temporal-spectral neural network (TTSNet).

These contributions highlight the innovative methods and techniques introduced by the doctoral student to advance the field of EEG signal processing, specifically in addressing challenges related to small-size AD datasets, and in the areas of classification of single-side upper limb movements.

3. COPY OF PUBLISHED AND SUBMITTED WORKS

3.1. Assessing the Potential of Data Augmentation in EEG Functional Connectivity for Early Detection of Alzheimer's Disease



Assessing the Potential of Data Augmentation in EEG Functional Connectivity for Early Detection of Alzheimer's Disease

Hao Jia^{1,2} · Zihao Huang² · Cesar F. Caiafa³ · Feng Duan² · Yu Zhang⁴ · Zhe Sun⁵ · Jordi Solé-Casals^{1,6}

Received: 21 February 2023 / Accepted: 30 July 2023
© The Author(s) 2023

Abstract

Electroencephalographic (EEG) signals are acquired non-invasively from electrodes placed on the scalp. Experts in the field can use EEG signals to distinguish between patients with Alzheimer's disease (AD) and normal control (NC) subjects using classification models. However, the training of deep learning or machine learning models requires a large number of trials. Datasets related to Alzheimer's disease are typically small in size due to the lack of AD patient samples. The lack of data samples required for the training process limits the use of deep learning techniques for further development in clinical settings. We propose to increase the number of trials in the training set by means of a decomposition–recombination system consisting of three steps. Firstly, the original signals from the training set are decomposed into multiple intrinsic mode functions via multivariate empirical mode decomposition. Next, these intrinsic mode functions are randomly recombined across trials. Finally, the recombined intrinsic mode functions are added together as artificial trials, which are used for training the models. We evaluated the decomposition–recombination system on a small dataset using each subject's functional connectivity matrices as inputs. Three different neural networks, including ResNet, BrainNet CNN, and EEGNet, were used. Overall, the system helped improve ResNet training in both the mild AD dataset, with an increase of 5.24%, and in the mild cognitive impairment dataset, with an increase of 4.50%. The evaluation of the proposed data augmentation system shows that the performance of neural networks can be improved by enhancing the training set with data augmentation. This work shows the need for data augmentation on the training of neural networks in the case of small-size AD datasets.

Introduction

Alzheimer's disease (AD) is a clinical syndrome characterized by the progressive deterioration of the memory and cognitive functions, particularly in elderly people. The

disease usually appears silently, and the process is slow and irreversible. According to the 2019 Alzheimer's World Report [1], there are more than 50 million people with AD. The figure may rise to 152 million by 2050.

✉ Feng Duan
duanf@nankai.edu.cn

✉ Zhe Sun
z.sun.kc@juntendo.ac.jp

✉ Jordi Solé-Casals
jordi.sole@uvic.cat

Yu Zhang
yuzhang@lehigh.edu

¹ Data and Signal Processing Research Group, University of Vic-Central, University of Catalonia, Vic, Catalonia, Spain

² College of Artificial Intelligence, Nankai University, Tianjin, China

³ CONICET CCT La Plata/CIC-PBA/UNLP, Instituto Argentino de Radioastronomía, Villa Elisa, Argentina

⁴ Department of Bioengineering and the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem 18015, PA, USA

⁵ Faculty of Health Data Science, Juntendo University, Urayasu, Chiba, Japan

⁶ Department of Psychiatry, University of Cambridge, Cambridge, UK

In recent years, the attention paid to AD has been gradually increasing. So far, only five drugs have been approved by the Food and Drug Administration (FDA) for the treatment of AD [2], and all of them can only delay the development of AD and alleviate symptoms, but not cure or even treat AD. Consequently, early diagnosis is important to delay the symptoms through medication. Typically, AD is divided into four stages, and the best time to diagnose the disease is during the early stages of mild cognitive impairment (MCI) and mild AD [3–5].

Electroencephalography (EEG) is the non-invasive acquisition of signals corresponding to electrical activity in the brain using electrodes positioned directly on the scalp. Magnetoencephalography (MEG) is also a non-invasive technique which is used to acquire signals by recording the magnetic activity of the brain. Functional magnetic resonance imaging (fMRI) indirectly detects changes of the brain neuronal activity based on the linked alterations of cerebral blood flow as exhibited by the differentiated magnetic properties of the hemoglobin molecule between its oxygen saturated and desaturated states. The difference between AD patients and normal control subjects can be detected using these brain signals, each coming with different advantages and disadvantages. Machine learning methods related to the classification between AD patients and normal control subjects using EEG, MEG, and fMRI brain signals are listed in Table 1.

With the increasing use of deep learning techniques, many deep AD detection methods have recently emerged. Sarraf and Tofighi [14] used LeNet-5, a convolutional neural network (CNN) architecture, to classify fMRI data from AD subjects and normal controls, with an accuracy on the testing dataset of 96.85%. They used 5-fold cross-validation on a dataset containing 28 AD subjects and 15 normal controls. Kim and Kim [15] proposed a classifier based on deep neural networks using the relative power of EEG to fully exploit and recombine features through its own learning structure. Their dataset contained 10 MCI subjects and 10 normal controls, and leave-one-out cross-validation was used to evaluate the model's performance. The accuracy obtained on the testing dataset was 59.4%. Duan et al. [16] used EEG functional connectivity

as the network input to train ResNet-18, achieving an accuracy of 93.42% and 98.5% on the MCI and mild AD datasets, respectively, where the former contained 22 MCI subjects and 38 normal controls, and the latter contained 17 mild AD subjects and 24 normal controls.

Among the aforementioned brain signals (EEG, MEG, and fMRI), EEG has the best temporal resolution. Nevertheless, since EEG signals are acquired from several locations on the scalp with electrodes, their spatial resolution is not as good as that of the measurements for the other two types of signals. Despite this, the spatial distribution of the signals can be optimized in the processing steps with the use of well-designed algorithms [17–21]. Given that EEG signals are easier to acquire and is less expensive than other techniques, EEG-based methods for AD detection are currently more popular.

In studies based on EEG signals, deep learning methods are trained on small datasets, as electrophysiological signals are more difficult to acquire in AD patients. The learning capability of deep learning models partially relies on their large number of hyper-parameters. A high amount of samples is required to fit these hyper-parameters and avoid the over-fitting problem [22, 23]. One way to deal with the issue is using data augmentation.

Data augmentation can be implemented by generating artificial data [24, 25]. The strategy of decomposing and recombining the original EEG signals is one possible way to create new artificial data for data augmentation [26–28]. EEG signals can be decomposed into different filter banks. In each filter bank, the frequency of the decomposed EEG signals is within a certain frequency band. All filter banks cover a wide range of frequencies. This strategy helps to achieve a better performance using deep-learning models in the enhancement of small-size datasets. Note that in studies where this particular data augmentation strategy has been implemented, the details about the models used are not entirely the same throughout, even though the same overall approach is being used. For instance, Zhao et al. [26] proposed a method of

Table 1 Summary of papers using EEG/MEG/fMRI signals to design a classification system for AD/MCI detection

Ref	Method	Signal	Disease type	Accuracy	Year
[6]	Correlation, phase synchrony, and Granger causality measures	EEG	MCI and mild AD	83% and 88%, respectively	2012
[7]	Hybrid feature selection	EEG	MCI and mild AD	95% and 100%, respectively	2015
[8]	Complex network theory and TSK fuzzy system	EEG	AD	97.3%	2019
[9]	Functional connectivity and effective connectivity analysis	MEG	AD	86%	2019
[10]	Phase locking value, imaginary part, and correlation of the envelope	MEG	MCI	75%	2019
[11]	High-order FC correlations	fMRI	MCI	88.14%	2016
[12]	Hierarchical high-order functional connectivity networks	fMRI	MCI	84.85%	2017
[13]	Strength and similarity guided GSR using LOFC and HOFC	fMRI	MCI	88.5%	2019

random recombination of EEG signals in different filter banks, which are decomposed by the discrete cosine transform. This approach enhances the classification performance of one-dimension convolutional neural networks in the epileptic seizure focus detection task. Zhang et al. [27] used the augmentation strategy to enhance the classification performance of motor imagery. Instead of decomposing signals with the discrete cosine transform, the empirical mode decomposition (EMD) technique was adopted [29]. In the decomposition–recombination strategy, EMD has the advantage that the signals can be recovered by simply adding up the decomposed intrinsic mode functions (IMFs). Besides the decomposition–recombination strategy, generative adversarial networks (GANs) also offer a solution to generate artificial signals [30]. However, GANs require a large dataset to tune the parameters and fit the model. Since the goal of data augmentation in small Alzheimer’s datasets is to solve the problem of insufficient samples, it is not possible to use GANs to generate artificial data.

In this paper, we propose a decomposition and recombination model for data augmentation in a small Alzheimer’s data set, which is used to distinguish AD patients from normal controls. The decomposition and recombination approach consists of three steps. First, empirical multivariate mode decomposition (MEMD) is used to decompose EEG signals into IMFs. These IMFs are then randomly recombined within each of the two groups. Finally, in each group, the IMFs are added up to generate a new artificial trial. These artificial trials are used to extend the AD training dataset.

This work is organized as follows. "Method" includes the description of the small Alzheimer’s datasets used, the scheme of the proposed decomposition and recombination approach, and the neural networks used for classification. "Results" presents the experimental results, including the classification performance of the neural networks during the training process and the effects of data augmentation in the datasets. Then, these results are discussed in "Discussion", together with the limitations associated with the method. Finally, the conclusions are presented in "Conclusion".

Method

Alzheimer’s Datasets

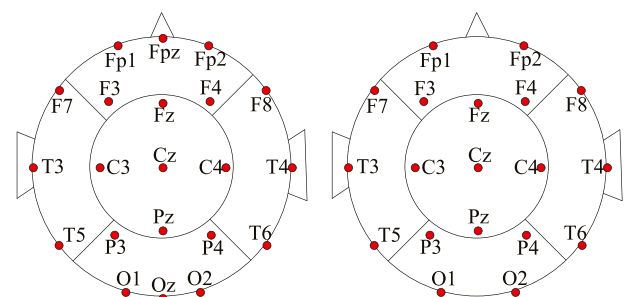
All experiments in this work use two datasets: the MCI dataset, containing 22 subjects with MCI and 38 normal controls, and the mild AD dataset, containing 17 subjects

with mild AD and 24 normal controls. Other studies have been conducted based on these datasets [5, 7, 31].

The MCI Dataset

The MCI dataset is comprised of data from subjects who complained of memory impairment and of control subjects who did not have memory impairment or other diseases. The patient group included 53 subjects who underwent a comprehensive neuropsychological test; the results showed quantitative and objective evidence of memory impairment, but their overall cognitive, behavioral or functional status was not significantly lost. The classification of mild dementia impairment requires a score of at least 24 in the Mini-Mental State Examination (MMSE) [32], a score of 0.5 on the Clinical Dementia Rating (CDR) scale [33] and a standard deviation lower than the normal memory performance reference value. All subjects met these criteria. Then, these subjects underwent an initial assessment, and their progress was monitored in the clinic during the subsequent 12–18 months. According to the criteria defined by the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA), 25 of these 53 mild AD patients might develop into AD. The average age of the 25 subjects in the MCI data set is 71.9 ± 10.2 years old, and the MMSE score is 28.5 ± 1.6 . The control group had 56 age-matched healthy subjects with an average age of 71.7 ± 8.3 years old and an MMSE score of 26 ± 1.8 .

Twenty-one electrodes from Biotop 6R12 (NEC-Sanei, Tokyo, Japan) were placed on the subject’s scalp in a 10–20 international system with a sampling frequency of 200 Hz. In addition, Fpz and Oz electrodes were added to the system, as shown in Fig. 1a. After the data was collected, analog bandpass filtering was used



(a) Placement of the EEG electrodes in the MCI dataset (b) Placement of the EEG electrodes in the mild AD dataset

Fig. 1 Schematic display of the electrode positions from above

to retain data between 0.5 and 250 Hz, and then third-order Butterworth filters (forward and reverse filtering) were used to perform digital band-pass filtering between 0.5 and 30 Hz.

The Mild AD Dataset

The mild AD dataset is comprised of data from 17 mild AD patients (age: 69.4 ± 11.5 years) and 24 healthy subjects (age: 77.6 ± 10.0 years). The patient group underwent a full set of cognitive tests (MMSE, Rey auditory verbal learning, Benton visual retention, and memory recall tests) along with psychological tests. The results were graded and interpreted by psychologists and then discussed in meetings with multidisciplinary teams. The subjects in the control group were all healthy volunteers, and their EEG was judged to be normal by the clinical neurophysiology consultants.

Nineteen electrodes were placed on the subject's scalp using the Maudsley system, which is similar to the international 10–20 system. The sampling frequency was 128 Hz, as shown in Fig. 1b. After data acquisition was carried out, a third-order Butterworth filter (forward filter and reverse filter) was used for digital band-pass filtering between 0.5 and 30 Hz.

Recording Conditions in Both Datasets

During the collection process of the two aforementioned datasets, the subjects were awake and with their eyes closed. The whole process lasted for 5 min. After that, the EEG data was checked by EEG experts, and the data containing artifacts were discarded. Finally, only clean EEG data of 20 s of length was saved for each subject, discarding the subjects whose data did not meet this condition. Based on this procedure, the MCI dataset finally comprised of 22 subjects with MCI and 38 normal controls, while the mild AD dataset comprised of 17 subjects with mild AD and 24 normal controls.

A Decomposition and Recombination System

In small data sets, neural networks often face overfitting problems. Data augmentation is used to enlarge the size of the training set, as shown in Fig. 2.

In this work, we propose a decomposition and recombination system to generate artificial trials and thus enlarge the training set. For the decomposition part, the empirical mode decomposition (EMD) method is used. EMD can divide a signal into multiple intrinsic mode functions (IMFs). These IMFs cover different frequency bands, with low overlap. The original signal can then be

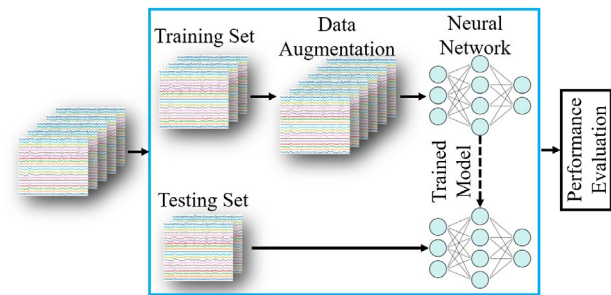


Fig. 2 The concept of data augmentation. In a small data set, the training set is small in size, since it is generated from only a portion of the (few) original data. When a neural network is used to fit the training set, there is a potential overfitting problem. Data augmentation is used to mitigate this issue by enlarging the size of the training set

recovered by adding up these IMFs [29]. The recombination part consists of adding IMFs from different trials, taking each of the IMFs from a different one.

The simplest EMD method is classical empirical mode decomposition (CEMD), which is the original version of EMD, as shown in the algorithm 1. A faster version of EMD is serial EMD (SEMD), which is used to deal with multi-channel signals. SEMD converts multi-channel signals into a single channel by concatenating them over time, ensuring the continuity of the signals by suitably adding a transient part between channels. CEMD is then used to decompose the single (long) channel. Multivariate EMD (MEMD) is also a method used for decomposing multi-channel signals, as shown in the algorithm 2. First, it places the multi-channel signals in a tangent space and then decomposes these signals into IMFs. The IMFs are finally reverted to normal space. Figure 3 shows the original multi-channel signals and the signals decomposed by MEMD. MEMD ensures that IMFs with the same index (shown in Fig. 3) cover the same frequency band.

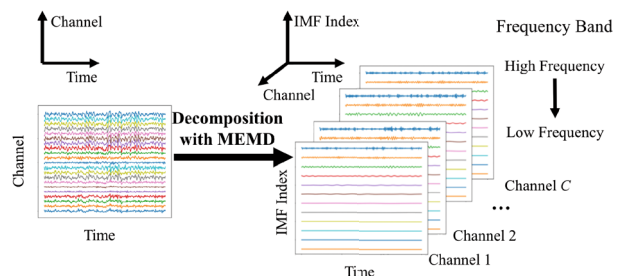


Fig. 3 Data decomposition with MEMD. MEMD can decompose multi-channel signals into IMFs. The IMFs are located in different frequency bands, but in all the decomposed channels, the k th IMF covers the same frequency band. In this figure, the IMFs are sorted in descending order in the frequency domain

Algorithm 1 CEMD

Require: $T \in \mathbb{Z}$
Ensure: $\mathbf{x} \in \mathbb{R}^{T \times 1}$
while number of extreme values in \mathbf{x} is greater than 3 **do**
interpolate the maxima extremes (minima extremes) of \mathbf{x}
obtain the envelop of extremes e_{max} (e_{min})
 $\mathbf{m} = (e_{max} + e_{min})/2$
 $\mathbf{h} = \mathbf{x} - \mathbf{m}$
if \mathbf{h} is an IMF **then**
save \mathbf{h} as an IMF
 $\mathbf{x} = \mathbf{x} - \mathbf{r}$
else
 $\mathbf{x} = \mathbf{h}$
end if
end while
save \mathbf{x} as an IMF

Algorithm 2 MEMD

Require: $T \in \mathbb{Z}$
Ensure: $\mathbf{x} \in \mathbb{R}^{T \times C}$
obtain N_{dir} uniformly-distributed sequences ($\theta = \theta_1, \theta_2, \dots, \theta_P$)
from Hammersley sequence, $N_{dir} > P$;
cast θ into tangent space and obtain direction vector \mathbf{v} .
while number of extreme values in \mathbf{x} is greater than 3 **do**
for $don = 1 : N_{dir}$
cast X into tangent space with inner product ($X \cdot \mathbf{v}$)
locate the positions of extremes in $X \cdot \mathbf{v}$, which are assumed
to be the positions of extremes in X
interpolate the maxima extremes (minima extremes) of X
obtain the envelop of extremes e_{max}^n (e_{min}^n) of the n -th
uniform-distributed sequence.
end for
 $e_{max} = \text{mean}(e_{max}^n)$ for $n = 1 : N_{dir}$
 $e_{min} = \text{mean}(e_{min}^n)$ for $n = 1 : N_{dir}$
 $\mathbf{m} = (e_{max} + e_{min})/2$
 $\mathbf{h} = \mathbf{x} - \mathbf{m}$
if \mathbf{h} is an IMF **then**
save \mathbf{h} as an IMF
 $\mathbf{x} = \mathbf{x} - \mathbf{r}$
else
 $\mathbf{x} = \mathbf{h}$
end if
end while
save \mathbf{x} as an IMF

In order not to decompose each trial separately, which would result in IMFs covering non-equal frequency bands in the same trial, and also to decrease the processing time, we combine the MEMD and SEMD methods as shown in Fig. 4. Multi-channel signals from several trials are first concatenated along the time axis as in SEMD, and then MEMD is used to decompose the concatenated signals, ensuring that each trial has the same number of IMFs. Figure 5 presents an example of generating an artificial trial with the original EEG signals.

Neural Network Classifiers

In the analysis of EEG signals, there are two traditional options used as inputs for the neural networks. In the first case, the original multi-channel signals are used as inputs. In the second, the multi-channel signals are converted into a functional connectivity (FC) matrix [34]; this is an

EEG-based connectivity matrix between brain regions obtained by calculating the inter-channel EEG similarity, e.g., by means of the coherence measure. The degree of similarity between two brain regions can be reflected in the FC matrix. In this way, the generated matrix preserves the spatial information of the multi-channel signals. To distinguish between controls and AD patients, EEG is often analyzed in four frequency bands: *delta* (0.1–4 Hz), *theta* (4–8 Hz), *alpha* (8–13 Hz), and *beta* (13–30 Hz). The signal in each band contains different information about brain connectivity and synchronization [35]. In this work, however, we adopt slightly different frequency bands, namely 4–8 Hz, 8–10 Hz, 10–13 Hz, and 13–30 Hz. These bands are derived from a previous work [16] and are optimized for the datasets used [7].

The main goal of this work is to measure the effect of the data augmentation method on the performance of the classifiers when functional connectivity matrices are used as inputs to the models. Therefore, it is not in the scope of this work to determine the best possible model. To evaluate the effects of the data augmentation method on the small AD datasets, three neural networks are used: BrainNet CNN [36], ResNet [37], and EEGNet [38]. To simplify the explanation of the networks, some symbols are defined here. In the following, B is the batch size, C is the number of input EEG signals, and T is the number of sample points of the EEG signals.

Methods such as Pearson’s correlation coefficient or coherence can be used to compute the correlation or relationship between channels. Here, we adopt the coherence to compute the FC matrices. EEG coherence measures the degree of phase synchronization of EEG spectral activity between two electrodes [39]. For two temporal signals $x(t)$ and $y(t)$, the coherence between them can be defined as follows:

$$C_{xy} = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)}, \tag{1}$$

where G_{xy} is the cross-spectral density between x and y , and G_{xx} and G_{yy} are the power-spectral density of x and y , respectively. Considering an EEG sample that has 21 channels containing data of 20 s of length, we can obtain an FC matrix with a size of $C \times C$ by calculating the coherence between each pair of EEG signals. Here, we first divide the original signals into the four aforementioned frequency bands, namely 4–8 Hz, 8–10 Hz, 10–13 Hz, and 13–30 Hz. As a consequence, the input of the neural networks is of size $4 \times C \times C$ (where C is the channel number of EEG signals). The inputs for BrainNet CNN and ResNet are the FC matrices of the four frequency bands. The input for EEGNet is the original multi-channel time series.

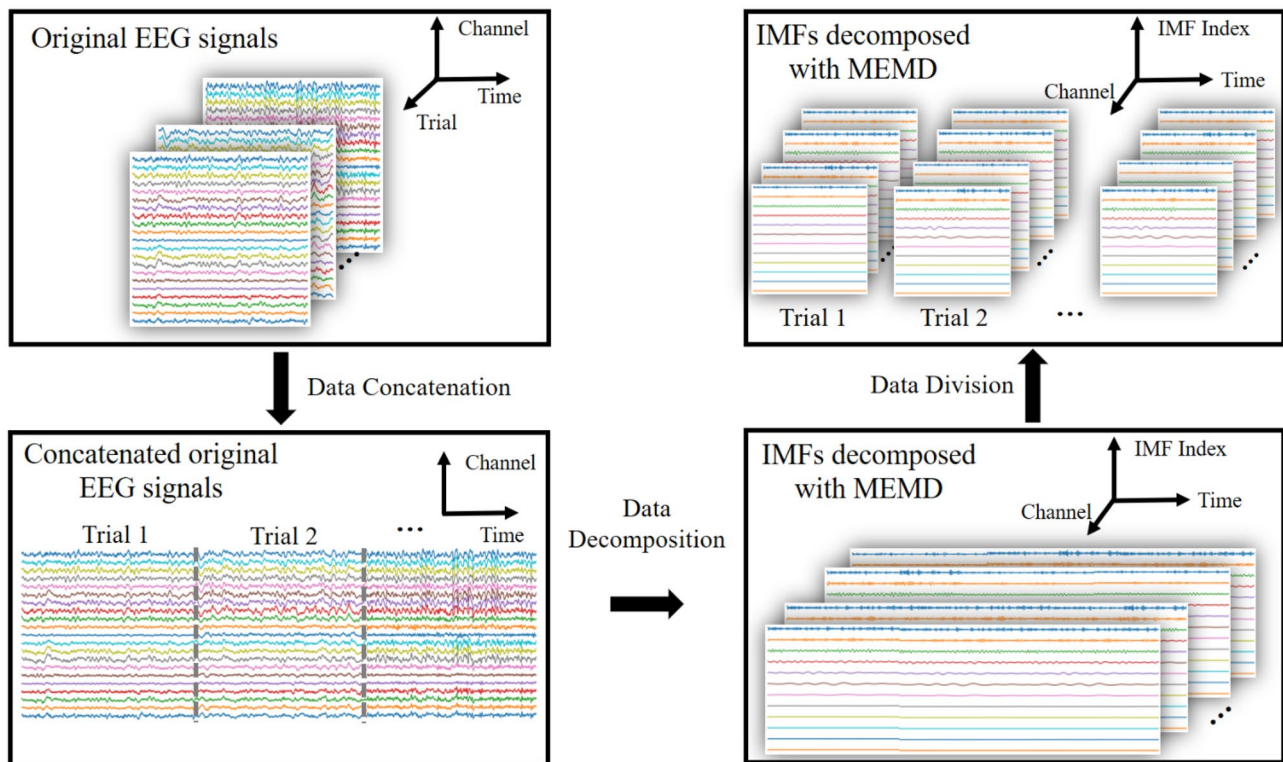


Fig. 4 The procedure of SEMD-MEMD decomposition for multiple trials of multi-channel EEG signals. Trials of EEG signals are concatenated along the time axis and then decomposed

BrainNet CNN

BrainNet CNN is a network architecture that analyzes the FCs of different frequency bands [36]. This network has three basic convolutional blocks: edge-to-edge (E2E), edge-to-node (E2N), and node-to-graph (N2G), which are specially designed for FC matrix processing. The three blocks are convolutional layers with different kernels. E2N is a convolutional layer with kernel size $(1, C)$ which converts the edges in FC matrices to nodes. N2G is a convolutional layer with kernel size $(C, 1)$ which suppresses the output nodes of the E2N layer. Finally, E2E is the added-up output of convolutions with kernel size $(1, C)$ and $(C, 1)$. An illustration of E2E is given in Fig. 6. The structure of the BrainNet CNN is given in Table 2.

ResNet

In the training process of deep learning methods, the back-propagation of multiple layers faces the problem of gradient vanishing [40]. The residual module of the deep residual network can reduce the influence of gradient vanishing by introducing a shortcut connection [37]. The deep residual network is a network that has already been validated on a

large number of classification problems. Compared with that of deep neural networks without shortcut connections, the shortcut connection of the deep residual network allows raw input information to be sent directly to a later layer. Assuming that the input of the residual block is x , the expected output is $H(x)$. The learning target of the deep residual network is then $F(x) = H(x) - x$, which is called residual, and then the input and output of this block are added together through the shortcut (Fig. 7). This approach greatly increases the training speed of the model, improves the training effect, and effectively solves the vanishing problem when the number of layers is deepened without adding extra parameters and calculations to the network. In this study, we employed the ResNet-18 deep residual network.

EEGNet

EEGNet is a universal solution to the classification of multi-channel EEG signals, which has been validated in the classification of other brain activity signals such as motor imagery and movement-related cortical potential [38]. EEGNet takes the original multi-channel EEG signals as the input instead of the FC matrices. Even though EEGNet has not been validated in the classification of early AD, in this work, we use

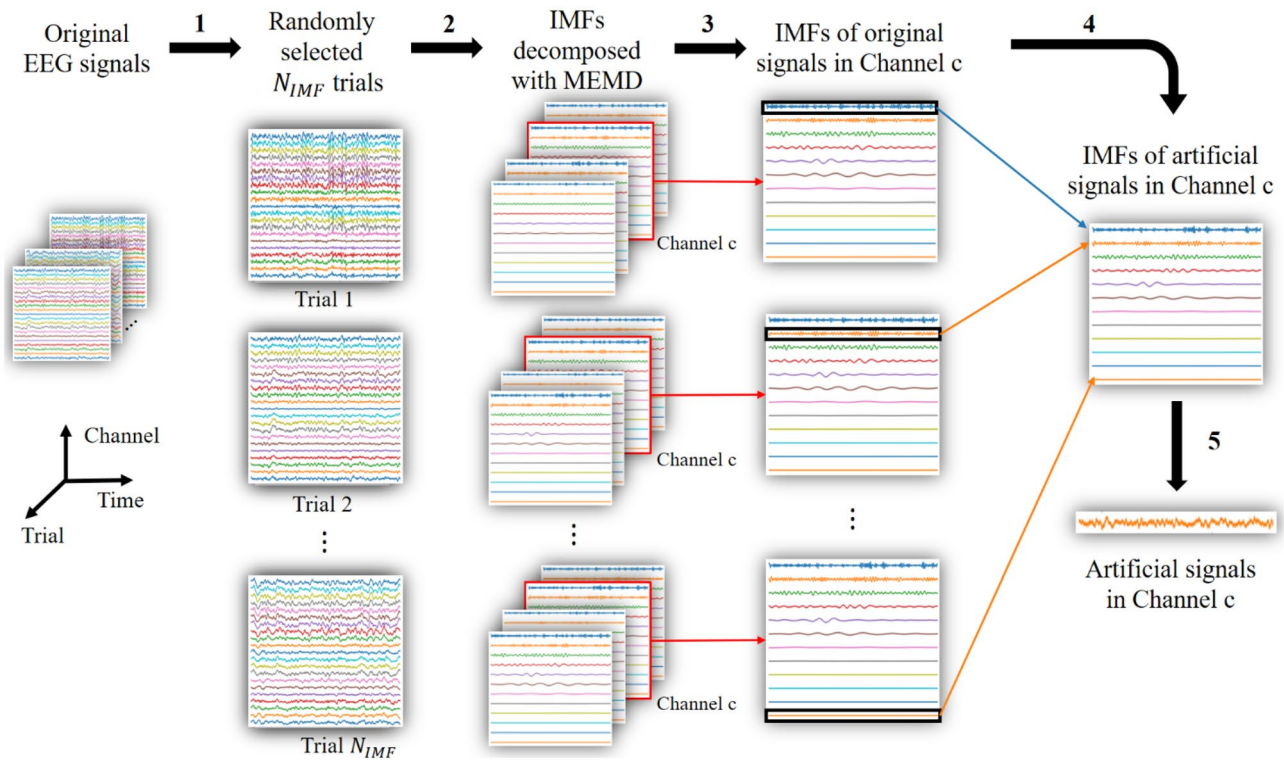


Fig. 5 Outline of the proposed decomposition and recombination system. As an example, for an artificial signal generated in channel c , the procedure consists of **i** randomly selecting N_{IMF} trials from the original EEG signals; **ii** obtaining the IMFs, which are decomposed using the method outlined in Fig. 4; **iii** collecting the decomposed IMFs

of channel c from randomly selected N_{IMF} trials; **iv** recombining the IMFs in channel c . The n_{imf} -th IMF of the artificial signal is the n_{imf} -th IMF of the n_{imf} -th randomly selected trial; and **v** adding the IMFs and obtaining the artificial signal of channel c

it to test and explore the data augmentation performance. The structure of EEGNet is given in Table 3.

Parameter Setting

In the training of these neural networks, the adaptive moment estimation (Adam) optimizer was used, with $\beta_1 = 0.9$, $\beta_2 = 0.99$ and 0.0001 for the learning rate. ResNet and BrainNet CNN were trained using 100 epochs,

and EEGNet was trained using 200 epochs. The mini-batch size was 50.

Results

The experiments aim to explore the effects of data augmentation on the small AD dataset with the decomposition and recombination strategy using FC matrices as inputs and with three different neural networks as classifiers. In Table 4, the number of trials in the training and testing sets is given. In the training set, 10 trials are randomly selected from the original EEG signals of AD patients and controls to avoid the imbalance of the training set. Five hundred artificial trials are generated from the 10 original trials for each class. The rest of the original trials are used in the testing set. The chance level is calculated with the stratified dummy classifier in Python’s scikit-learn toolbox [41]. The training set consists of both original and artificial EEG signals. Artificial EEG signals in the training set are generated exclusively from the real EEG data of this set (Fig. 5). The original EEG signals in the training set

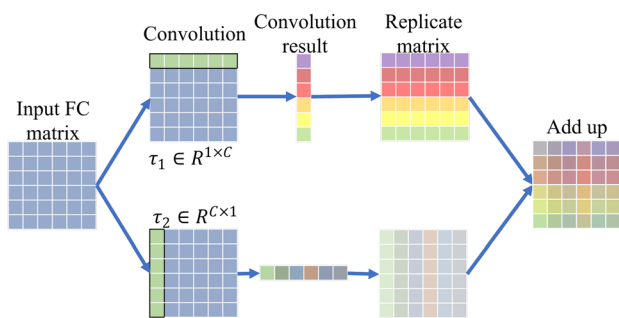


Fig. 6 A schematic depiction of the E2E block in BrainNet CNN. The output of the block is the sum of two convolution results

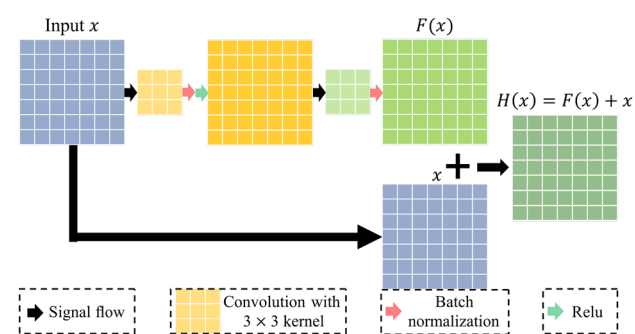
Table 2 The structure of BrainNet CNN

Layer	Output size	Parameter
Input layer	$[B, 4, C, C]$	
BatchNorm	$[B, 4, C, C]$	
ReLU	$[B, 4, C, C]$	
E2E	$[B, 16, C, C]$	$(C, 1)$
BatchNorm	$[B, 16, C, C]$	
ReLU	$[B, 16, C, C]$	
E2E	$[B, 32, C, C]$	
ReLU	$[B, 32, C, C]$	
E2N	$[B, 64, C, 1]$	$(1, C)$
N2G	$[B, 512, 1, 1]$	$(C, 1)$
Flatten	$[B, 512]$	
Linear and softmax	$[B, 2]$	

are randomly selected ten times, and the classification is repeated as a cross-validation procedure.

Feature Distribution

First, the feature distributions of the artificial data generated by data augmentation are assessed. To clearly illustrate this, the FC matrices of mild AD patients vs controls are depicted in Fig. 8 using the uniform manifold approximation and projection method (UMAP) [42, 43]. There are four FC matrices for each trial: 4–8 Hz, 8–10 Hz, 10–13 Hz, and 13–30 Hz. Since FC matrices are symmetric, their upper triangle is taken as the feature of said matrix. For each trial, we have $4 \times C \times (C - 1)/2$ features. The UMAP model is first trained with features from 10 mild AD trials and 10 control trials. Then, 100 artificial mild AD trials and 100 artificial control trials generated with SEMD-MEMD, SEMD, or CEMD are transformed with the trained UMAP model. In the UMAP setting, the size of the local neighborhood used for manifold approximation is set to 10, and the effective minimum distance between embedded points is set to 1; the training epoch number for embedding optimization is 1000. The dimension of the features is reduced and projected onto

**Fig. 7** A residual block with a shortcut in ResNet

a two-dimensional map with UMAP. Figure 8 shows that artificial data of the two classes generated with MEMD are more easily separable than those generated with SEMD or CEMD.

Performance Analysis

The evolution of the classification accuracy of the classifiers during the training process is depicted in Fig. 9. The training set is augmented with SEMD-MEMD. For the mild AD dataset, EEGNet has the worst classification performance with an average accuracy of around 53%. The data augmentation deteriorates the performance of EEGNet compared to the case of not using artificial data. On the other hand, the classification accuracy for BrainNet CNN improves with data augmentation when the number of artificial trials is greater than 20, as the accuracy converges faster than without data augmentation. The ResNet performance also improves with data augmentation.

In Fig. 10, the trend of the accuracy of the classification is given. The accuracies of ResNet and BrainNet CNN in this figure are obtained after 100-epoch training, while the number of training epochs of EEGNet is 200. We note that data augmentation does not always help to improve the training of neural networks.

Finally, Fig. 11 shows the confusion matrices, with only real data (before) or with 10 artificial trials per class (after), respectively. The number of artificial trials generates an increase of 100% of samples in the training dataset (factor of 2). These confusion matrices are calculated using MATLAB's "confusionmat" function [44]. The results were

Table 3 The structure of EEGNet

Layer	Output size	Parameter
Input layer	$[B, 1, C, T]$	
ZeroPad2d	$[B, 1, C, T+63]$	$(31, 32, 0, 0)$
Conv2d	$[B, 8, C, T]$	$(1, 64)$
BatchNorm2d	$[B, 8, C, T]$	
Conv2d	$[B, 16, 1, T]$	$(C, 1), \textit{grouped}$
BatchNorm2d	$[B, 16, 1, T]$	
ELU	$[B, 16, 1, T]$	
AvgPool2d	$[B, 16, 1, T//4]$	$(1, 4)$
Dropout	$[B, 16, 1, T//4]$	0.25
ZeroPad2d	$[B, 16, 1, T//4+15]$	$(7, 8, 0, 0)$
Conv2d	$[B, 16, 1, T//4]$	$(1, 15), \textit{grouped}$
Conv2d	$[B, 16, 1, T//4]$	$(1, 1)$
BatchNorm2d	$[B, 16, 1, T//4]$	
ELU	$[B, 16, 1, T//4]$	
AvgPool2d	$[B, 16, 1, T//32]$	$(1, 8)$
Dropout	$[B, 16, 1, T//32]$	0.25
Flatten	$[B, 16 \times T//32]$	
Linear	$[B, K]$	$\textit{bias} = \textit{False}$

Table 4 Distribution of the number of trials

Data type	Training set		Testing set	Chance level
	Artificial	Original	Original	
Data type	Artificial	Original	Original	
Mild AD	0–500	10	7	0.3333
Control	0–500	10	14	
MCI	0–500	10	12	0.3000
Control	0–500	10	28	

obtained by averaging over ten folds, and the final values were normalized by dividing by the sum of each row. The experiment was carried out using only a small number of artificial trials, as the results depicted in Fig. 10 pointed out that this was a good value in almost all the models. Table 5 contains the accuracy, sensitivity, and precision calculated using the “confusionmat” function.

Discussion

In this work, we proposed a decomposition and recombination system to enlarge the size of two AD datasets and explored the data augmentation performance on three different neural networks. This work is based on the following two assumptions:

1. The AD dataset is a small dataset.
2. Neural networks need a considerable amount of data to tune the parameters.

Most patients affected by AD are elderly people. In contrast to the EEG signal acquisition of healthy people, AD patients are easily exhausted, weak, or less willing in the process of acquiring EEG signals. Sometimes, the acquisition can even be interrupted for unexpected reasons such as the non-collaboration of the patients. Therefore, AD datasets

are very valuable and are usually small in size. To protect the health of the patients and to facilitate data acquisition in experiments, a data augmentation method is needed to process small AD datasets.

When it comes to the second assumption, note that deep neural networks can accurately find the unknown relationship between the raw data and the corresponding labels because of their intrinsic nature and huge number of parameters. At the same time, these parameters can only be learned from the available data, but the higher the number of parameters, the higher the number of signals needed to train the model. Therefore, data augmentation on small AD datasets is again of great interest.

In addition to the decomposition and recombination strategy in data augmentation, generative adversarial networks (GANs) are also a universal solution for time series data augmentation. However, in these, both the generator and discriminator parameters require a certain amount of data to be tuned. For an AD dataset of limited size, this requirement on the amount of data is not met, and hence, GANs are not suitable in this case.

In the classification of mild AD, data augmentation has a positive effect on the training of ResNet. When the number of artificial trials increases, the average accuracy of ResNet increases from 72.38 to 77.62%, with a consistent performance. In the BrainNet CNN case, a positive outcome is also obtained in the classification performance when using data augmentation in the mild AD dataset. However, this effect is only positive for a small number of artificial trials in the MCI dataset; if the number of artificial trials increases above 30, the mean accuracy decreases. Finally, the EEGNet network is the one with the poorest results for the mild AD dataset, and artificial trials only have a moderate positive effect for the MCI dataset again when the number of artificial trials is small.

In Fig. 11, the confusion matrices before and after data augmentation are given. Both ResNet and BrainNet CNN obtain a consistent accuracy, sensitivity, and precision

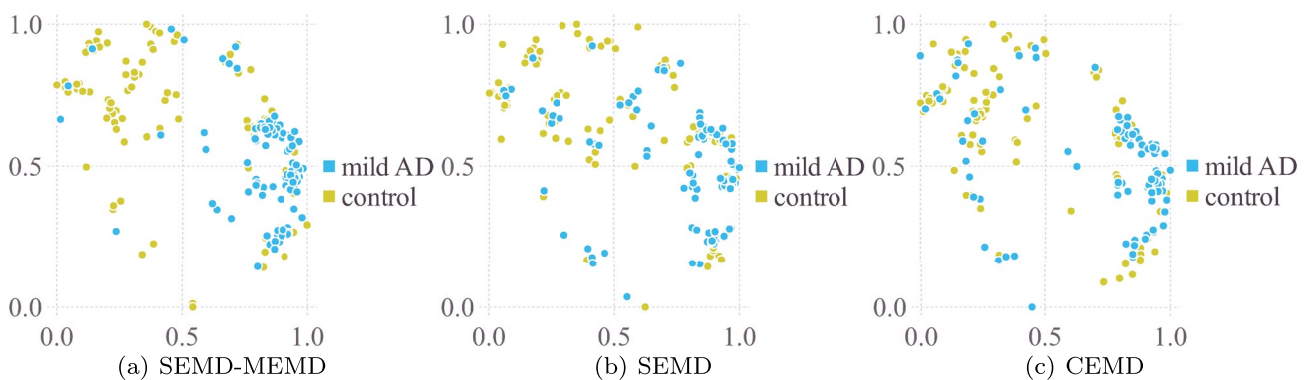


Fig. 8 Feature map of artificial mild AD patients vs controls, plotted with UMAP. For each class, 100 artificial samples are generated using MEMD **a**, SEMD **b**, and CEMD **c**. The obtained embedding is normalized with min-max normalization before visualization

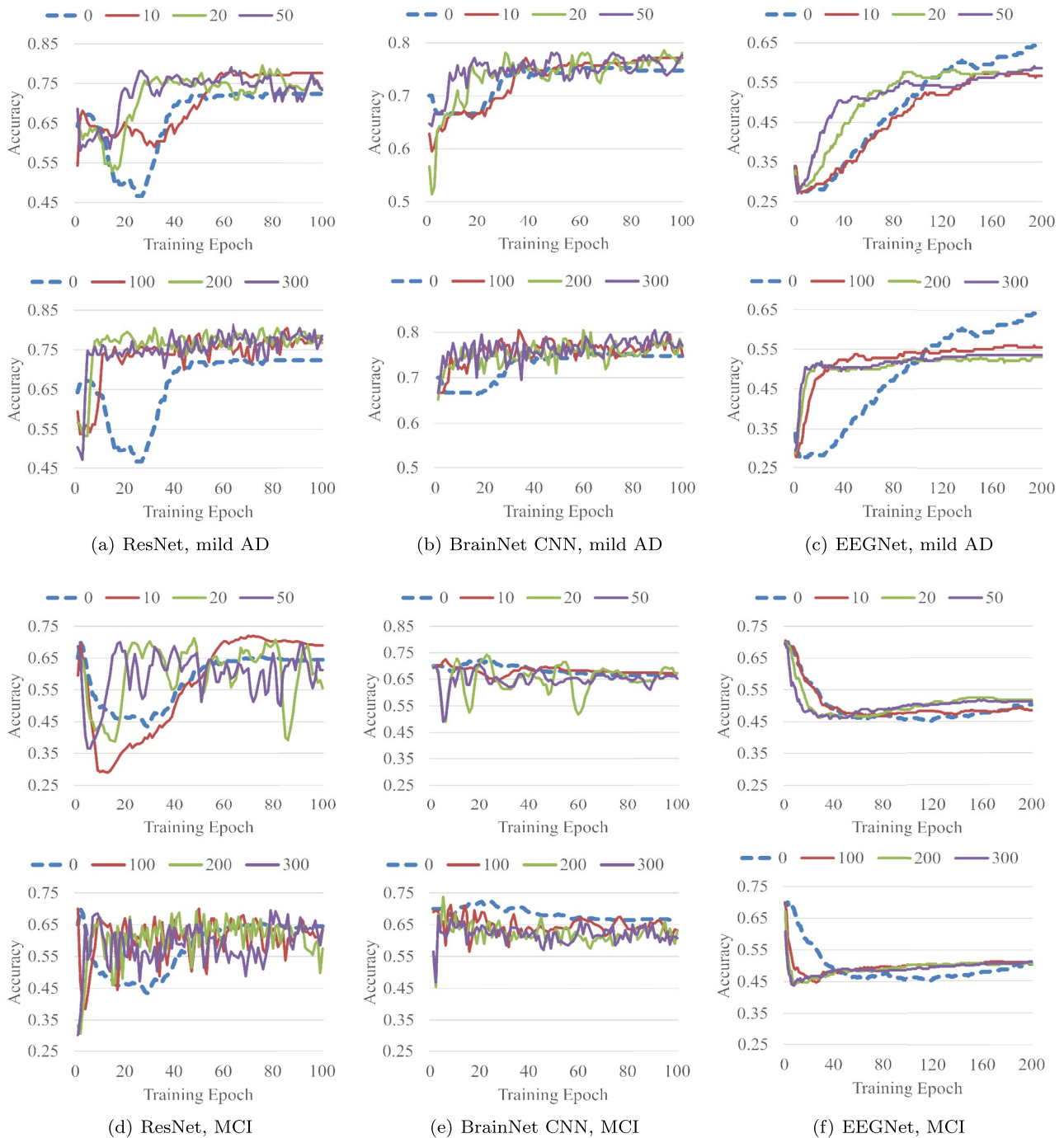


Fig. 9 The testing accuracy averaged across ten folds of the two datasets during the training process. A different number of artificial trials are generated, each one of them shown in a different color in the subplots. For each case, the upper panel contains the experiments with

0 to 50 artificial trials, and the lower panel contains the experiments with 0 to 500 artificial trials. The dashed line represents the 0 case, where no artificial trials are used

increase when 10 artificial trials per class are used. As expected, the improvement is more noticeable in the mild AD database, as the two classes (controls and patients) are more distant from each other when compared to the MCI case, in which the patients are closer to the control subjects.

Summarizing the above experiments, the proposed decomposition and recombination system helps the training of neural networks in small AD datasets, and it seems that just a factor of 2 is enough for that. Having more artificial data does not always provide a better result, as we have seen

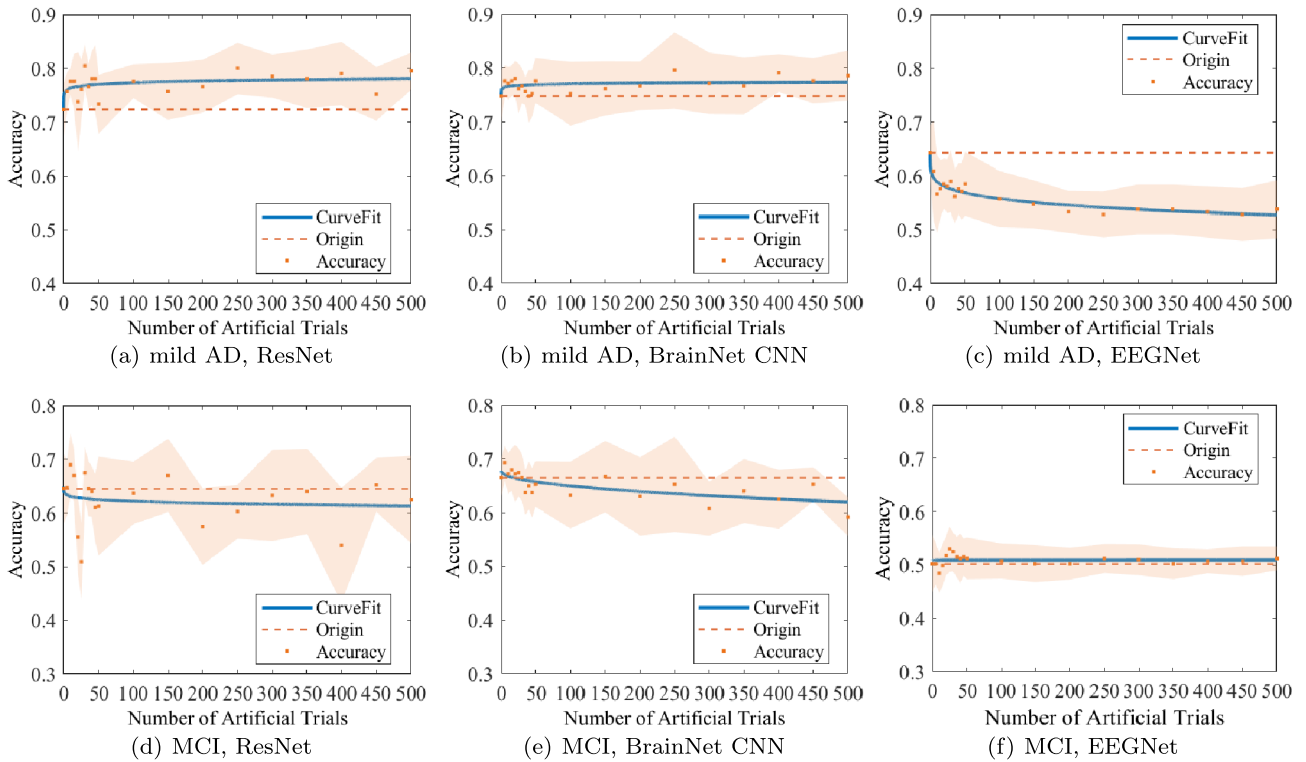


Fig. 10 Accuracy evolution when the number of artificial trials increases from 0 to 500. The trend of the accuracy is fitted with the power function $f(x) = ax^b + c$. The dotted line represents the accuracy without data augmentation

in our experiments. The effects of the data augmentation depend on two factors: (i) the type of neural networks and (ii) the data set. Determining the number of artificial trials is influenced by these two factors, and ascertaining how to obtain an optimal value requires further experiments.

One possible reason for why the proposed data augmentation method does not always improve the accuracy

results is due to the different characteristics of the two datasets. In Fig. 9a, the accuracy of ResNet in the mild AD dataset converges as the number of training epochs increases, and the result is stable in the training, with a small variance around the mean accuracy. However, in Fig. 9d, the accuracy in the MCI dataset still fluctuates in a larger range, especially compared with the mild AD

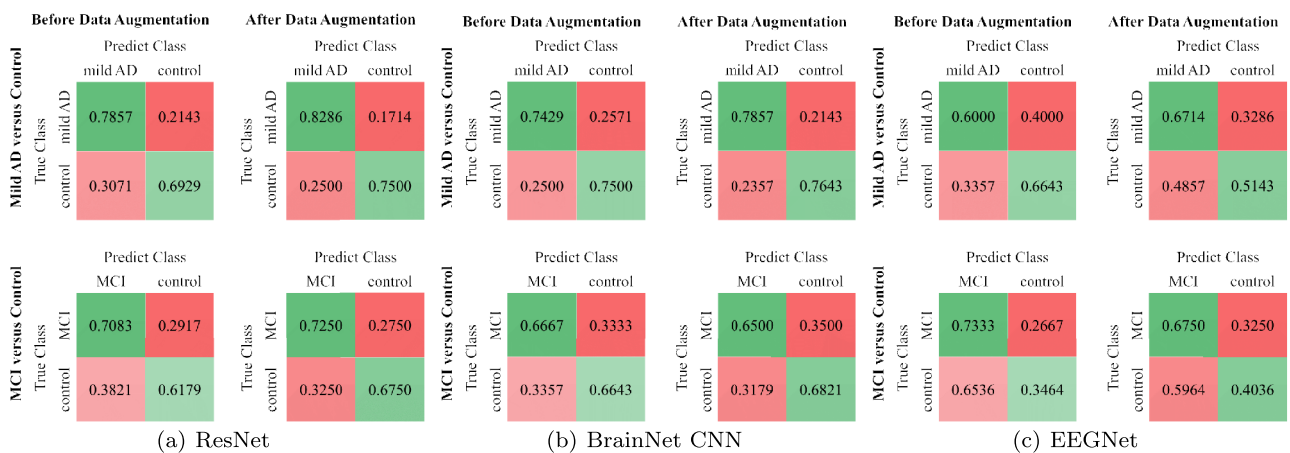


Fig. 11 Comparison of the confusion matrices before and after data augmentation for the two datasets. The confusion matrices are averaged across ten folds and normalized by dividing by the sum of each row

Table 5 Performance measurement before and after data augmentation, calculated using the “confusionmat” function in Fig. 11. The best result in each case is highlighted in bold

ResNet		Accuracy	Sensitivity	Precision
Mild AD	Before	0.7238	0.7393	0.7243
	After	0.7762	0.7893	0.7628
MCI	Before	0.6450	0.6631	0.6450
	After	0.6900	0.7000	0.6775
BrainNet CNN		Accuracy	Sensitivity	Precision
Mild AD	Before	0.7476	0.7464	0.7338
	After	0.7714	0.7750	0.7850
MCI	Before	0.6650	0.6655	0.6469
	After	0.6725	0.6661	0.6517
EEGNet		Accuracy	Sensitivity	Precision
Mild AD	Before	0.6429	0.6321	0.6336
	After	0.5667	0.5929	0.5887
MCI	Before	0.4625	0.5399	0.5399
	After	0.4850	0.5393	0.5337

dataset. This means that the network is more difficult to fit for the MCI dataset or that perhaps the quality of the data is also worse in that case. Although data augmentation improves the accuracy in the MCI dataset very slightly when the number of artificial trials is small, it still helps to train the ResNet: when the accuracy converges, the number of training epochs needed after data augmentation is smaller than without data augmentation, as shown in Fig. 9d. Similar fluctuations can be observed for the BrainNet CNN network in both datasets (Fig. 9b, e). This could explain why data augmentation is not helping in this case.

The proposed decomposition and recombination system has its own limitations. No pre-processing was used to remove artifacts or noise in the databases used in the experiments. Since the proposed method recombines all existing information in the data to enlarge the size of the training data, it is possible that artifacts or noise may also be replicated, which would negatively affect the results. Another aspect that can play a role is the decomposition method used. Here, we combine SEMD and MEMD, but other EMD-based methods have been proposed in the literature. Each method has different properties which impact the frequency mixing effect (overlapping of IMFs) and hence may influence the quality of the artificial frames. Moreover, the number of required artificial trials is unknown, as has been shown, and should be further investigated. More experiments are also needed to determine the number of epochs in the training phase, as our results indicate that the use of artificial trials may help to reduce the number of epochs in training and thus control possible overfitting.

All of these aspects are now under consideration, and we expect to propose more reliable methods in future works.

Conclusion

In this paper, we proposed a decomposition and recombination system for data augmentation of the small AD data set as a way to solve the problem of insufficient data in neural network training.

This system consists of signal decomposition with SEMD-MEMD and a random recombination of the decomposed IMFs. The performance of this system is evaluated using three classifiers on two datasets. The main results show that the proposed system improves the accuracy of ResNet on the mild AD dataset with an increase of 5.24% and on the MCI dataset with an increase of 4.50%. Furthermore, BrainNet CNN results improve on the mild AD dataset with an increase of 2.38% and an increase of 0.75% on the MCI dataset. This work is expected to help the training process of detection methods for early diagnosis of Alzheimer’s disease.

Acknowledgements This work was carried out as part of the doctoral program in Experimental Science and Technology at the University of Vic - Central University of Catalonia. The authors would like to thank Pau Solé-Vilaró for the English proofreading of this manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research was supported by the National Natural Science Foundation of China (Key Program) (No. 11932013) and the Tianjin Science and Technology Plan Project (No. 22PTZWHZ00040). J.S.-C. work is also based upon work from COST Action CA18106, supported by COST (European Cooperation in Science and Technology). C.F.C work is partially supported by grants PICT 2020-SERIEA-00457 and PIP 112202101 00284CO (Argentina).

Data Availability Not applicable. No new data were collected for this study.

Code Availability The code used for the data augmentation is freely available here: <https://github.com/Sole-Casals/DR-EMD>.

Declarations

Informed Consent Not applicable. No new data were collected for this study.

Conflict of Interest The authors declare no competing interests.

Research Involving Human Participants and/or Animals Not applicable. No new data were collected for this study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are

included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- International AD. World Alzheimer report 2019: attitudes to dementia. Alzheimer's Disease International London, UK; 2019.
- Du X, Wang X, Geng M. Alzheimer's disease hypothesis and related therapies. *Translational neurodegeneration*. 2018;7(1):1–7.
- The Alzheimer Association. Alzheimer's disease facts and figures. *Alzheimer's & Dementia*. 2023Apr;19(4):1598–695. Available from: <https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13016>.
- Hansson O, Edelmayer RM, Boxer AL, Carrillo MC, Mielke MM, Rabinovici GD, et al. The Alzheimer's Association appropriate use recommendations for blood biomarkers in Alzheimer's disease. *Alzheimer's & Dementia*. 2022 Dec;18(12):2669–86. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/alz.12756>.
- Musha T, Asada T, Yamashita F, Kinoshita T, Chen Z, Matsuda H, et al. A new EEG method for estimating cortical neuronal impairment that is sensitive to early stage Alzheimer's disease. *Clinical Neurophysiology*. 2002Jul;113(7):1052–8. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1388245702001281>.
- Gallego-Jutglà E, Elgendi M, Vialatte F, Solé-Casals J, Cichocki A, Latchoumane C. Diagnosis of Alzheimer's disease from EEG by means of synchrony measures in optimized frequency bands. In: et al. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE. 2012;2012:4266–70.
- Gallego-Jutglà E, Solé-Casals J, Vialatte FB, Elgendi M, Cichocki A, Dauwels J. A hybrid feature selection approach for the early diagnosis of Alzheimer's disease. *Journal of Neural Engineering*. 2015;12(1).
- Yu H, Lei X, Song Z, Liu C, Wang J. Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification. *IEEE Transactions on Fuzzy Systems*. 2019.
- Mandal PK, Banerjee A, Tripathi M, Sharma A. A comprehensive review of magnetoencephalography (MEG) studies for brain functionality in healthy aging and Alzheimer's disease (AD). *Frontiers in Computational Neuroscience*. 2018;12:60.
- Yang S, Bornot JMS, Wong-Lin K, Prasad G. M/EEG-based biomarkers to predict the MCI and Alzheimer's disease: a review from the ML perspective. *IEEE Transactions on Biomedical Engineering*. 2019;66(10):2924–35.
- Chen X, Zhang H, Gao Y, Wee CY, Li G, Shen D, et al. High-order resting-state functional connectivity network for MCI classification. *Human Brain Mapping*. 2016;37(9):3282–96.
- Chen X, Zhang H, Lee SW, Shen D, Initiative ADN, et al. Hierarchical high-order functional connectivity networks and selective feature fusion for MCI classification. *Neuroinformatics*. 2017;15(3):271–84.
- Zhang Y, Zhang H, Chen X, Liu M, Zhu X, Lee SW, et al. Strength and similarity guided group-level brain functional network construction for MCI diagnosis. *Pattern Recognition*. 2019;88:421–30.
- Sarraf S, Tofighi G. Classification of Alzheimer's disease using fMRI data and deep learning convolutional neural networks. 2016. arXiv preprint <http://arxiv.org/abs/1603.08631> arXiv:1603.08631.
- Kim D, Kim K. Detection of early stage Alzheimer's disease using EEG relative power with deep neural network. In: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE. 2018;2018:352–5.
- Duan F, Huang Z, Sun Z, Zhang Y, Zhao Q, Cichocki A, et al. Topological network analysis of early Alzheimer's disease based on resting-state EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2020.
- Vialatte FB, Solé-Casals J, Maurice M, Latchoumane C, Hudson N, Wimalaratna S, et al. Improving the quality of EEG data in patients with Alzheimer's disease using ICA. In: International Conference on Neural Information Processing. Springer. 2008;979–86.
- Vialatte FB, Solé-Casals J, Cichocki A. EEG windowed statistical wavelet scoring for evaluation and discrimination of muscular artifacts. *Physiol Measure*. 2008;29(12):1435.
- Sanchez-Poblador V, Monte-Moreno E, Solé-Casals J. ICA as a preprocessing technique for classification. In: International Conference on Independent Component Analysis and Signal Separation. Springer. 2004;11657–2.
- Solé-Casals J, Vialatte FB. Towards semi-automatic artifact rejection for the improvement of Alzheimer's disease screening from EEG signals. *Sensors*. 2015;15(8):17963–76.
- Solé-Casals J, Caiafa CF, Zhao Q, Cichocki A. Brain-computer interface with corrupted EEG data: a tensor completion approach. *Cognitive Computation*. 2018;10(6):1062–74.
- Caiafa CF, Solé-Casals J, Marti-Puig P, Zhe S, Tanaka T. Decomposition methods for machine learning with small. *Incomplete or Noisy Datasets Applied Sciences*. 2020;10(8481):1–20.
- Alzubaidi L, Bai J, Al-Sabaawi A, Santamaría J, Albahri AS, Al-dabbagh BSN, et al. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *J Big Data*. 2023;10(1):46. Available from: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00727-2>.
- Hazra D, Byun YC. SynSigGAN: generative adversarial networks for synthetic biomedical signal generation. *Biology*. 2020;9(12):441.
- Bhattacharyya A, Singh L, Pachori RB. Fourier-Bessel series expansion based empirical wavelet transform for analysis of non-stationary signals. *Digit Signal Process*. 2018;78:185–96.
- Zhao X, Solé-Casals J, Li B, Huang Z, Zhao Q. Classification of epileptic EEG signals by CNN and data augmentation. In: ICASSP 2020. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2020.
- Zhang Z, Duan F, Solé-Casals J, Dinares-Ferran J, Cichocki A, Yang Z, et al. A novel deep learning approach with data augmentation to classify motor imagery signals. *IEEE Access*. 2019;1.
- Li B, Zhang Z, Duan F, Yang Z, Zhao Q, Sun Z, et al. Component-mixing strategy: a decomposition-based data augmentation algorithm for motor imagery signals. *Neurocomputing*. 2021;465:325–35. Available from: <https://www.sciencedirect.com/science/article/pii/S0925231221013308>.
- Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings Mathematical Physical & Engineering Sciences*. 1971;1998(454):903–95.
- Haradal S, Hayashi H, Uchida S. Biosignal data augmentation based on generative adversarial networks. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Honolulu, HI: IEEE; 2018;368–71. Available from: <https://ieeexplore.ieee.org/document/8512396/>.
- Gallego-Jutglà E, Solé-Casals J, Vialatte FB, Dauwels J, Cichocki A. A theta-band EEG based index for early diagnosis of Alzheimer's disease. *Journal of Alzheimer's Disease*. 2015;43(4):1175–84.

32. Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society*. 1992;40(9):922-35. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1532-5415.1992.tb01992.x>.
33. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *International Psychogeriatrics*. 1997;9(S1):173-6.
34. Venkatesh M, Jaja J, Pessoa L. Comparing functional connectivity matrices: a geometry-aware approach applied to participant identification. *NeuroImage*. 2020;207:116398. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1053811919309899>.
35. Rodriguez G, Arnaldi D, Picco A. Brain functional network in Alzheimer's disease: diagnostic markers for diagnosis and monitoring. *Int J Alzheimer's Dis*. 2011.
36. Kawahara J, Brown CJ, Miller SP, Booth BG, Chau V, Grunau RE, et al. BrainNetCNN: convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*. 2017;146(1038):1049.
37. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016. p. 770-8.
38. Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: a compact convolutional network for EEG-based brain-computer interfaces. *J Neural Eng*. 2018;15(5):056013. Available from: <http://arxiv.org/abs/1611.08024>.
39. Ho MC, Chen TC, Huang CF, Yu CH, Chen JM, Huang RY, et al. Detect AD patients by using EEG coherence analysis. *J Med Eng*. 2014.
40. Hochreiter S. Recurrent neural net learning and vanishing gradient. *International Journal Of Uncertainty, Fuzziness and Knowledge-Based Systems*. 1998;6(2):107-16.
41. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*. 2011;12:2825-30.
42. McInnes L, Healy J. UMAP: uniform manifold approximation and projection for dimension reduction. *The Journal of Open Source Software*. 2018;3(29):861.
43. McInnes L, Healy J, Saul N, Grossberger L. UMAP: uniform manifold approximation and projection. *The Journal of Open Source Software*. 2018;3(29):861.
44. Inc TM. MATLAB version: 9.9.0 (R2020b). Natick, Massachusetts, United States: The MathWorks Inc. 2020. Available from: <https://www.mathworks.com>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3.2. Graph Empirical Mode Decomposition-Based Data Augmentation Applied to Gifted Children MRI Analysis



Graph Empirical Mode Decomposition-Based Data Augmentation Applied to Gifted Children MRI Analysis

Xuning Chen¹, Binghua Li¹, Hao Jia¹, Fan Feng¹, Feng Duan^{1*}, Zhe Sun^{2*}, Cesar F. Caiafa^{1,3} and Jordi Solé-Casals^{1,4,5*}

¹ Department of Artificial Intelligence, Nankai University, Tianjin, China, ² Computational Engineering Applications Unit, Head Office for Information Systems and Cybersecurity, RIKEN, Saitama, Japan, ³ Instituto Argentino de Radioastronomía, Consejo Nacional de Investigaciones Científicas y Técnicas – Centro Científico Tecnológico La Plata/Comisión de Investigaciones Científicas – Provincia de Buenos Aires/Universidad Nacional de La Plata, Villa Elisa, Argentina, ⁴ Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom, ⁵ Data and Signal Processing Research Group, University of Vic-Central University of Catalonia, Vic, Spain

OPEN ACCESS

Edited by:

Sergey M. Plis,
Georgia State University,
United States

Reviewed by:

Cota Navin Gupta,
Indian Institute of Technology
Guwahati, India
Guang Ling,
Wuhan University of
Technology, China

*Correspondence:

Feng Duan
duanf@nankai.edu.cn
Zhe Sun
zhe.sun.vk@riken.jp
Jordi Solé-Casals
jordi.sole@uvic.cat

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 31 January 2022

Accepted: 27 May 2022

Published: 01 July 2022

Citation:

Chen X, Li B, Jia H, Feng F, Duan F,
Sun Z, Caiafa CF and Solé-Casals J
(2022) Graph Empirical Mode
Decomposition-Based Data
Augmentation Applied to Gifted
Children MRI Analysis.
Front. Neurosci. 16:866735.
doi: 10.3389/fnins.2022.866735

Gifted children and normal controls can be distinguished by analyzing the structural connectivity (SC) extracted from MRI data. Previous studies have improved classification accuracy by extracting several features of the brain regions. However, the limited size of the database may lead to degradation when training deep neural networks as classification models. To this end, we propose to use a data augmentation method by adding artificial samples generated using graph empirical mode decomposition (GEMD). We decompose the training samples by GEMD to obtain the intrinsic mode functions (IMFs). Then, the IMFs are randomly recombined to generate the new artificial samples. After that, we use the original training samples and the new artificial samples to enlarge the training set. To evaluate the proposed method, we use a deep neural network architecture called BrainNetCNN to classify the SCs of MRI data with and without data augmentation. The results show that the data augmentation with GEMD can improve the average classification performance from 55.7 to 78%, while we get a state-of-the-art classification accuracy of 93.3% by using GEMD in some cases. Our results demonstrate that the proposed GEMD augmentation method can effectively increase the limited number of samples in the gifted children dataset, improving the classification accuracy. We also found that the classification accuracy is improved when specific features extracted from brain regions are used, achieving 93.1% for some feature selection methods.

Keywords: GEMD, MRI, gifted children, structural connectivity, BrainNetCNN

INTRODUCTION

Intelligence can be seen as the ability to recognize and understand reality and use knowledge and experience to solve problems such as memory, observation, imagination, thinking, and judgment. Gifted children are regarded to have higher intelligence and perform better in attention, language, mathematics, verbal working memory, shifting, and social problem-solving (Bucaille et al., 2022). At the same time, gifted children demonstrate high working memory capacity and more effective executive attention (Aubry et al., 2021). They also have significant differences in cognitive flexibility function and problem-solving and reasoning (Rocha et al., 2020).

Gifted children have higher intelligence and learn faster than others, probably due to differences in neurophysiology (Gross, 2006). Neurological differences mean that gifted children may experience neurodevelopmental trajectories different from normal children, leading to a greater connection of neuronal pathways (Navas-Sánchez et al., 2014). Gifted children have larger subcortical structures and more robust white matter microstructural organization between those structures in regions associated with explicit memory (Kuhn et al., 2021). They are also characterized by highly developed functional interactions between the right hemisphere and excellent cognitive control of the prefrontal cortex, enhanced frontoparietal cortex, and posterior parietal cortex (Wei et al., 2020). Ma et al. found that gifted children have network topological properties of high global efficiency and high clustering with a low wiring cost and a higher level of local connection density (Ma et al., 2017). Gifted children's structural brain network has a more integrated and versatile topology than normal children (Solé-Casals et al., 2019).

Based on previous work on the brain neuroscience of gifted children, we believe that it is significant to identify gifted children through the structure of their brains. In the past decades, many neuroscientists have tried to understand the brain mechanisms and proposed many types of neuroimaging techniques, such as magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), and diffusion tensor imaging. In recent years, deep learning algorithms have achieved good results in processing these types of signals. Abdelaziz Ismael et al. proposed an enhanced deep learning approach, residual networks, for brain cancer MRI images classification and achieved 99% accuracy (Abdelaziz Ismael et al., 2020). Sarraf et al. used convolutional neural network (CNN) architectures Lenet-5 and GoogleNet to classify fMRI data of Alzheimer's disease subjects and normal controls, and the accuracy of the test dataset reached 96.85% (Sarraf and Tofighi, 2016). The BrainNetCNN is proposed to predict clinical neurodevelopmental outcomes by brain networks (Kawahara et al., 2017). It utilizes structural brain networks' topological locality to create edge-to-edge (E2E), edge-to-node (E2N), and node-to-graph (N2G) convolutional filters, which makes it perform well on human brain data classification. Leonardsen et al. proposed that neural network is able to identify subject brain from its MRI (Leonardsen et al., 2022).

The deep learning technology is notable for its impressive performance and generalization capability, but the number of effective samples in the medical imaging dataset is usually small, leading to performance degradation. The training model needs large amount of data to avoid overfitting (Caiafa et al., 2020). However, obtaining enough MRI data is not easy. The acquisition and preprocessing of brain data are more difficult than image and voice data, for example. It is difficult to find gifted children in our daily life. The number of gifted children is small, especially those whose IQ test score is higher than 140. In this work, we use a sample of 29 children, from which the MRI was obtained. The brain was parcellated into 308 regions and from each region 7 morphometric features were extracted. Hence, we have a total of 2,156 features per subject (7 morphological features by 308 brain regions). Training a model in such a small and high-dimensional MRI dataset is complicated.

Therefore, we focus on MRI data augmentation to improve model training. Data augmentation has proven to be useful in MRI, improving the accuracy of schizophrenia classification by 5% (7–8% relative improvement using augmentation) (Ulloa et al., 2015). Also, Nguyen et al. proposed a data augmentation method synthesizing a new fMRI image by performing a T1-based coregistration to another subject's brain in native space. This method was tested on antidepressant treatment response fMRI and demonstrated a 26% improvement in predicting response using augmented images (Nguyen et al., 2020). Previous work proves that increasing the amount of neuroimaging data through an appropriate data augmentation method can significantly improve the accuracy of deep learning classification.

In our MRI dataset, we propose to use a data decomposition method, graph empirical mode decomposition (GEMD) (Tremblay et al., 2014). GEMD is an adaptation to graph signals of the well-known empirical modal decomposition (EMD) (Huang et al., 1998). EMD has some variants, such as GEMD, masking EMD, ensemble-EMD (EEMD), and multivariate EMD (MEMD). Masking EMD, EEMD, and MEMD can primarily alleviate the mode mixing problem, and masking EMD and MEMD can perform spatiotemporal reconstruction of active sources (Muñoz-Gutiérrez et al., 2018). GEMD improves many aspects of the critical points of EMD, namely, extrema, interpolation, and stopping criterion (Tremblay et al., 2014). Because a parcellation of 308 brain regions is used, which can help to build a brain region connection graph, GEMD is the best choice for our work, as we will base our data augmentation on the decomposition-recombination strategy first presented in Dinarès-Ferran et al. (2018) for EEG signals. To our knowledge, this is the first time this technique has been used on MRI data. To compare the results of the proposed method, we also generate artificial samples through a more classical approach, the synthetic minority over-sampling technique (SMOTE) (Chawla et al., 2002).

In this work, the BrainNetCNN is used as a deep learning classifier. The main motivation for using a deep learning method is that the MRI data can then be fed directly into the model without the need for any feature selection method. This is an important aspect to keep in mind as feature selection methods are usually very database-dependent, and the results could drop if the database is changed. We train the BrainNetCNN for the classification task, showing that a well-trained classification model can increase the classification accuracy from 55.7 to 78% when using artificial data. Moreover, in Zhang et al. (2021), a hybrid selection method of morphological features and brain regions on the same gifted children dataset was derived. They used a completion method, simultaneous tensor decomposition, and completion (STDC), for outlier correction. After tensor completion, several feature selection methods were applied to the training set to explore which morphometric features and brain regions could perform better in the classification step. Based on their methodology, we used GEMD to generate artificial data on Zhang et al.'s work to achieve an accuracy of 93.1% on the F-score (FS), combined feature selection, and rank FS method.

The rest of the article is organized as follows. The materials and methods and the details of the experiments are introduced. Then,

the experimental results are discussed, followed by discussion. Finally, the conclusions are summarized.

MATERIALS AND METHODS

The overall experimental process is shown in **Figure 1**. In this section, we introduce the six parts in order. The details of the data are first described. Then, we show the brain region atlas and the morphometric features. After that, the basic algorithm principle of GEMD will be provided. Then, the data augmentation with GEMD is introduced. The following is the structural connectivity (SC) analysis, which converts MRI images into a correlation matrix. Finally, we introduce a deep learning network, the BrainNetCNN, as a classifier.

Gifted Children MRI Dataset

The MRI dataset of gifted children contains 29 healthy, right-handed male subjects without neurological diseases (Solé-Casals et al., 2019). We refer to this dataset as the UVic-gifted children dataset (UVic-GC dataset). There is no significant age difference between the two groups. Gifted children have a high IQ and outstanding performance in various tasks such as spatial, numerical, reasoning, verbal, and memory (Gras et al., 2010). The criteria for gifted group included having an IQ in the very superior range (≥ 140). Gifted children also had a performance above the 90th percentile in three of the following aptitudes, namely, spatial, numerical, abstract reasoning, verbal reasoning, and memory. More details on the dataset can be found in Solé-Casals et al. (2019). **Table 1** summarizes the details of the dataset. Using similar procedure and scanning parameters, all participants underwent examinations in a 3 T MRI scanner (Magnetom Trio Tim, Siemens Medical Systems, Germany). The raw (anonymized) MRI data are available in the OpenNeuro repository (<https://openneuro.org/datasets/ds001988>).

Brain Region Atlas and Morphometric Features

In our study, the brain is divided into 308 cortical regions following previous work (Romero-Garcia et al., 2012). The parcellation atlas is based on the Desikan-Killiany Atlas (68 cortical areas). Each area defined in the Desikan-Killiany atlas is subdivided into spatially contiguous areas through a backtracking algorithm available in FreeSurfer (Desikan et al., 2006). The size of each area is approximately equal to 500 mm².

The original feature matrix includes seven morphological features measured in each of the 308 brain regions. **Figure 2** shows the morphological features such as gray matter volume, cortical thickness, surface area, intrinsic curvature, mean curvature, curvature index, and fold index.

Graph Empirical Mode Decomposition

Empirical modal decomposition can decompose a signal into a set of intrinsic mode functions (IMFs), each covering different frequency bands by interpolating the extremes in the time series (Huang et al., 1998). The IMFs have two characteristics, namely, (1) the number of its zero crossings must be equal or differ up to one compared to its number of extrema and (2) IMFs' upper

and lower envelopes must be symmetric to zero. When all the IMFs of the original signal are extracted, the iterative process is terminated. GEMD is an adaptation of the classical EMD for graph signals (Tremblay et al., 2014). It improves many aspects of the critical points of EMD, namely, extrema, interpolation, and stopping criterion.

For the graph creation, the set of N regions is used as nodes for the graph. A weighted graph parameter δ is used to define edges in the graph. Only pairs of regions (i, j) at a distance d_{ij} , shorter than δ , are connected by an edge, with weight $w_{ij} = \exp(-d_{ij}^2/2\delta^2)$. The distance d_{ij} is the Euclidean distance in the features space. In that case, we get a graph $G = (N, E)$, where E is the set of edges. The adjacency matrix A , which contains all the weights w_{ij} connecting the nodes, is also needed. We use the 3D coordinate points of 308 brain regions to calculate the adjacency matrix for the 308 brain regions graph.

For the definition of local extrema, a node n will be a local minimum (or maximum) if for all its neighbors in G , $x(n) < x(m)$ [or $x(n) > x(m)$, where $x(n)$ and $x(m)$ represent the value of one of the features in the n th and m th brain regions]. Once the extremes have been obtained, the graph signal is interpolated to get the lower and upper graph envelopes needed to derive the IMFs.

To maintain the hypothesis-free nature of the classical EMD method, interpolation is regarded as a discrete partial differential equation on the graph (Grady and Schwartz, 2003). As the envelope is a slowly changing component, the interpolation signal s needs to minimize the total graph change, $s' L s$, where L is the graph Laplacian matrix under the constraint that the graph signal value of the known vertex remains unchanged. Let K denote the set of vertices of the known graph signal, and U denote the set of unknown vertices. Then, to calculate the new, interpolated, graphical signals, we need to solve the following equation minimize $s' L s$ subject to:

$$s(K) = x(K) \quad (1)$$

Through simple rearrangement of vertices, s can be rewritten as $s' = [s'_K \ s'_U]$ in its equivalent vector expression, where s_K and s_U are the vector representations of $s(K)$ and $s(U)$, and the rearranged Laplacian matrix $L = \begin{bmatrix} L_K & R \\ R' & L_U \end{bmatrix}$. Finally, the graph interpolation is a Dirichlet problem on the graph, and its solution depends on the following linear equation (Kalaganis et al., 2020):

$$L_U s_U = -R s_K \quad (2)$$

We refer the reader to Grady and Schwartz (2003) and Kalaganis et al. (2020) for a detailed explanation of the graph interpolation method. With the mentioned elements, the sifting process can be modified easily. We set the parameter of the stopping criterion, which was defined in Tremblay's work (Tremblay et al., 2014), as follows: stop the loop (steps 4–8 in the following algorithm) as soon as the energy of the average envelope z (computed in the step 6) is lower than the energy of the analyzed signal x_i divided by 1,000.

After defining the graph extremum and interpolation process, the classic EMD algorithm can easily be extended to graph

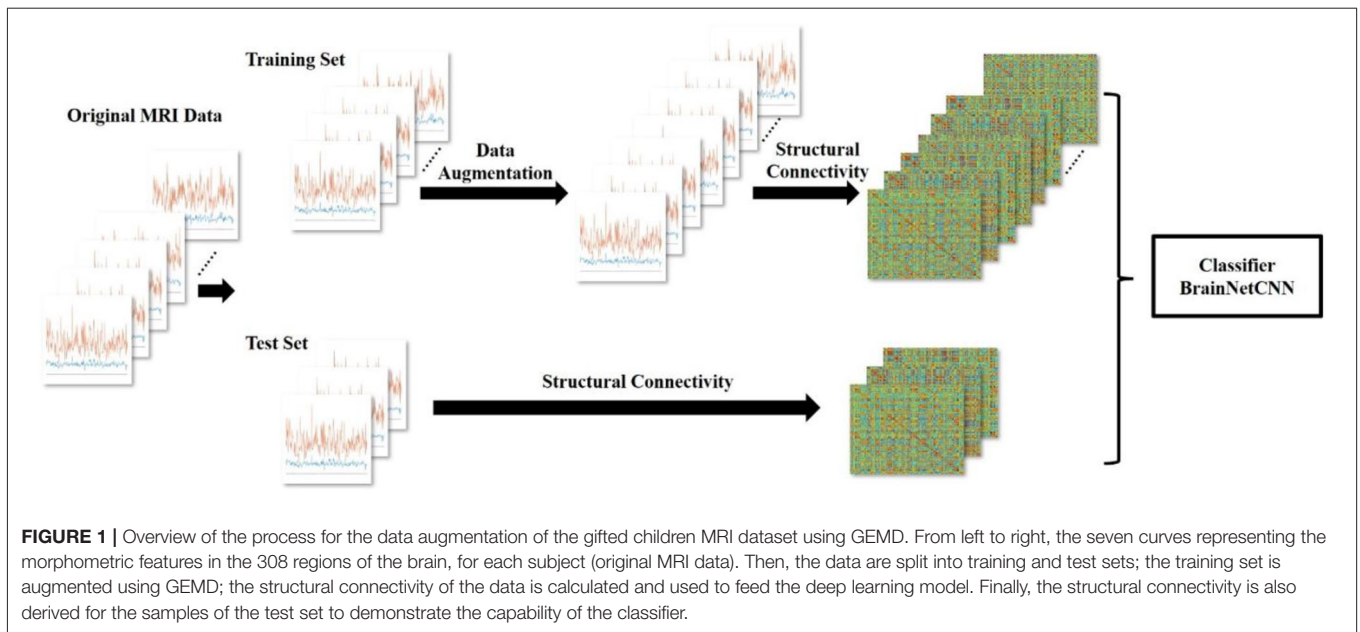


TABLE 1 | Membership information of gifted children MRI dataset.

Group	Gifted group	Control group
Average IQ	148.80 ± 2.93	122.71 ± 3.89
Average age	12.03 ± 0.54	12.53 ± 0.77
Sample size	15	14

signals. The process of data decomposition with GEMD is shown in **Figure 3**. The GEMD algorithm (Tremblay et al., 2014) is defined as follows:

- Step 1: Create the adjacency matrix A for the graph G ;
- Step 2: Initialize $m = x_i$;
- Step 3: While m does not meet the stopping criterion, repeat step 4 to step 8;
- Step 4: Detect the local extreme of m ;
- Step 5: Interpolate the upper and lower extremes of m and get the envelope e_{\max} and e_{\min} ;
- Step 6: Calculate the average envelope $z = \frac{e_{\min} + e_{\max}}{2}$;
- Step 7: Subtract the average envelope from the signal: $m = m - z$;
- Step 8: Set $d_{i+1} = m$ and $x_{i+1} = x_i - m$;
- Step 9: If m meets the stopping criteria: stop the decomposition and terminate, return stored IMFs, and get [MathType-mtef1-eqn-3.mtf].

Data Augmentation

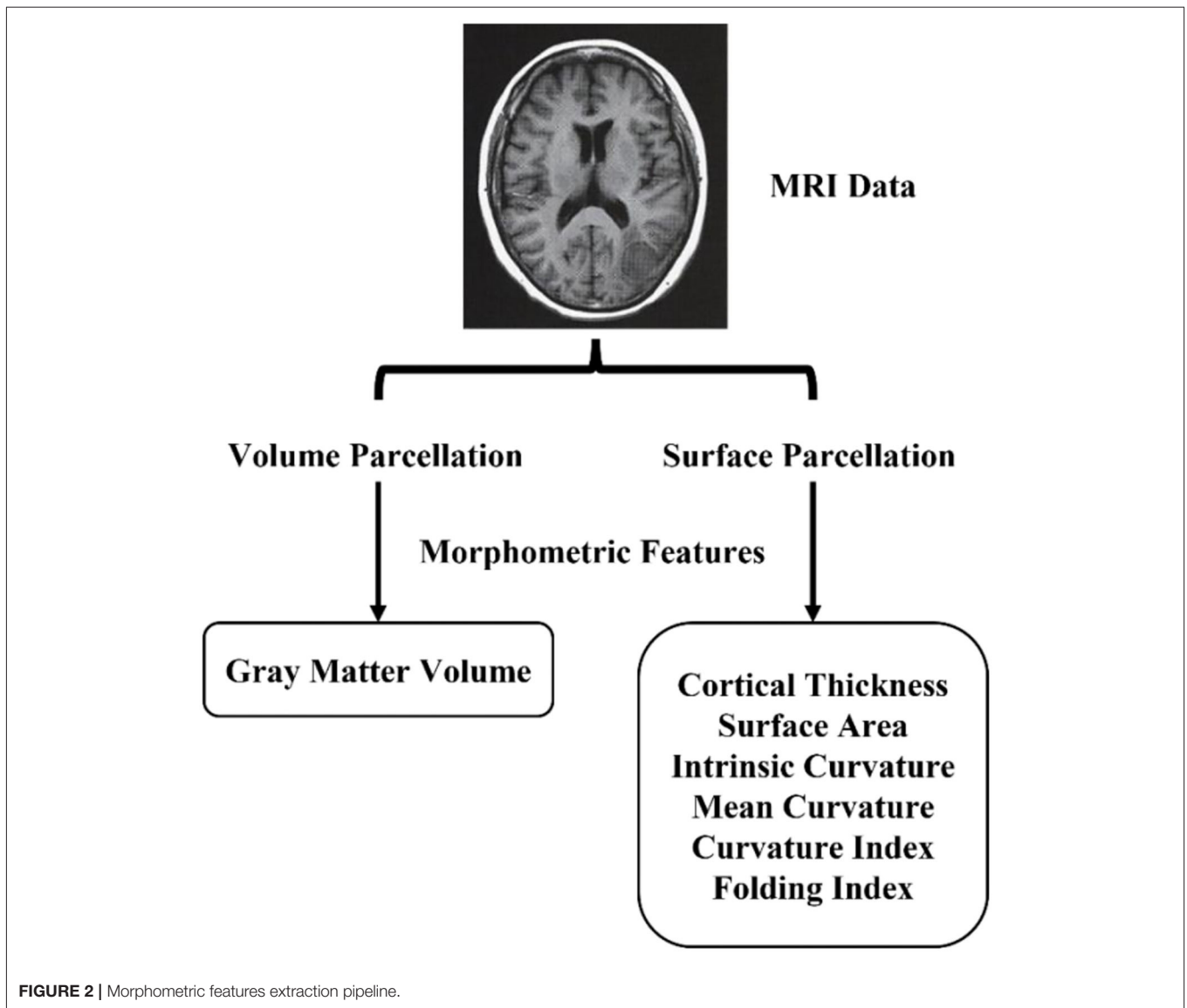
The MRI dataset contains $P = 29$ subjects. Therefore, the training set can be regarded as a three-dimensional tensor $T \in R^{B \times F \times P}$ (B : number of brain regions; F : number of features; P : number of subjects). If the number of subjects in the training set is too small, the model will tend to be overfitted. To overcome the overfitting problem in the UVic-GC dataset, we propose to

increase the training set through GEMD. The data augmentation procedure is based on a decomposition-recombination strategy, originally proposed in Dinarès-Ferran et al. (2018), and first used in a deep learning context in Zhang et al. (2019). The data augmentation process is shown in **Figure 4**. This method has the following steps:

- *Step 1: Data decomposition.*
- Create the adjacency matrix A for the graph G . In our work, A is obtained by calculating the Euclidean distance among the 308 regions.
- Organize the MRI data of all subjects and get the concatenated tensor $T \in R^{B \times F \times P}$.
- Decompose T with GEMD and get $T_{IMF} \in R^{M \times B \times F \times P}$, where M is the total number of IMFs ($M = 5$ in our experiments).
- *Step 2: Artificial data generation.*
- Randomly select M subjects from one of the groups (gifted group or control group).
- Take one IMF from each subject so that you end up with one IMF from each category (IMF₁ to IMF₅), i.e., each subject contributes with one IMF to create the new artificial data. The artificial data of the n_{th} feature is the sum of the M IMFs.

Structural Connectivity Analysis

After creating artificial samples, we use the original subjects and the artificial to perform the classification. For that purpose, we calculate the SC between features in all the regions. The SC matrix (one matrix per sample, i.e., original subjects and artificial subjects created *via* GEMD) will be used later as the input data for the deep learning classification system. SC represents the data correlation between two brain regions (Qi et al., 2019). Pearson's correlation or coherence is usually used to compute the correlation. We use Pearson's correlation and z -score to obtain



the SC in this work. We correlate the seven values (morphometric features) of one region with the seven values (morphometric features) of another region. We perform these correlations for all possible pairs, obtaining a 308×308 matrix per subject. Assuming two brain region data x and y , Pearson's correlation (Kotu and Deshpande, 2019) between x and y can be expressed as follows:

$$c(x, y) = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \quad (3)$$

where S_{xy} is the covariance of x and y , which is defined as,

$$S_{xy} = \sum_{i=1}^n (x_i - x)(y_i - y) \quad (4)$$

S_{xx} and S_{yy} can be calculated as the variance of x and y , respectively. After we get the Pearson's correlation of MRI data, z-score is used to standardize it. Finally, a three-dimensional tensor of dimensions $29 \times 308 \times 308$ is obtained.

This procedure was introduced by Seidlitz et al. (2018) to estimate the inter-regional correlation of multiple MRI features in a single subject instead of estimating the inter-regional correlation of a single feature measured in multiple subjects (which is done with the structural covariance analysis). Therefore, we end up with an SC matrix per subject.

Neural Network Classifier

As the BrainNetCNN (Kawahara et al., 2017) outperforms lots of other methods on structural brain networks datasets, we choose it as a neural network classifier in our experiments. There are three kinds of convolutional filters in BrainNetCNN, namely, E2E, E2N, and N2G filters. They leverage the topological locality of

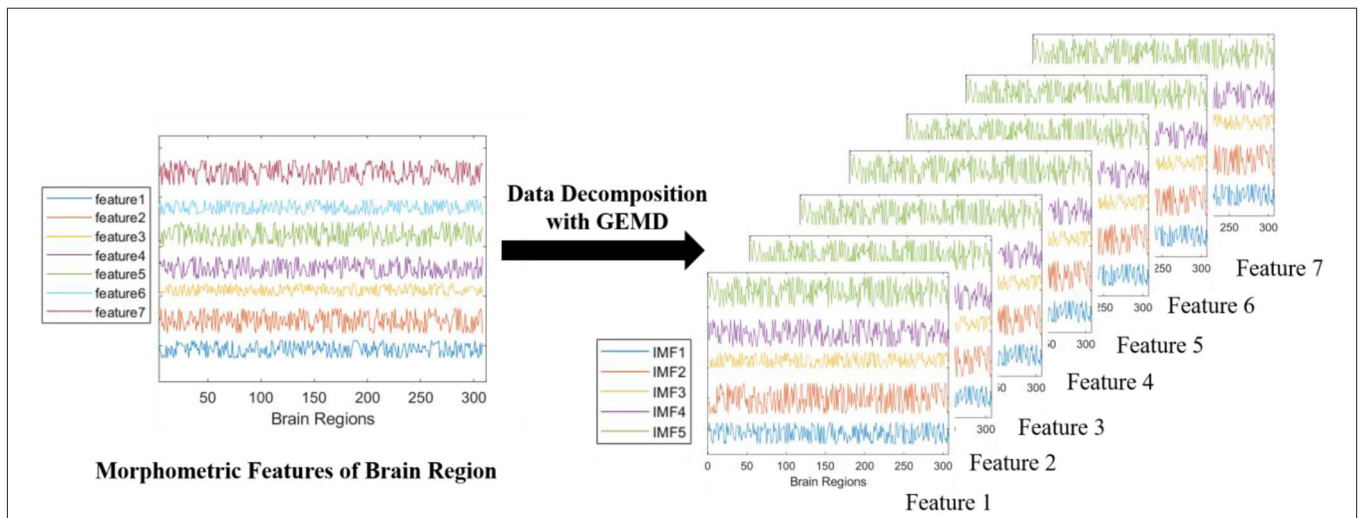


FIGURE 3 | The progress of data decomposition with GEMD. GEMD can decompose feature series (different color lines on the left) into IMFs simultaneously. Here, we use seven morphometric features from MRI data as a decomposition example. Every feature is decomposed into four or five IMFs (different color lines on the right). If only four IMFs are decomposed from raw data, zero-padding will be used to have a total of five IMFs in all the decompositions, so that the data augmentation can proceed successfully.

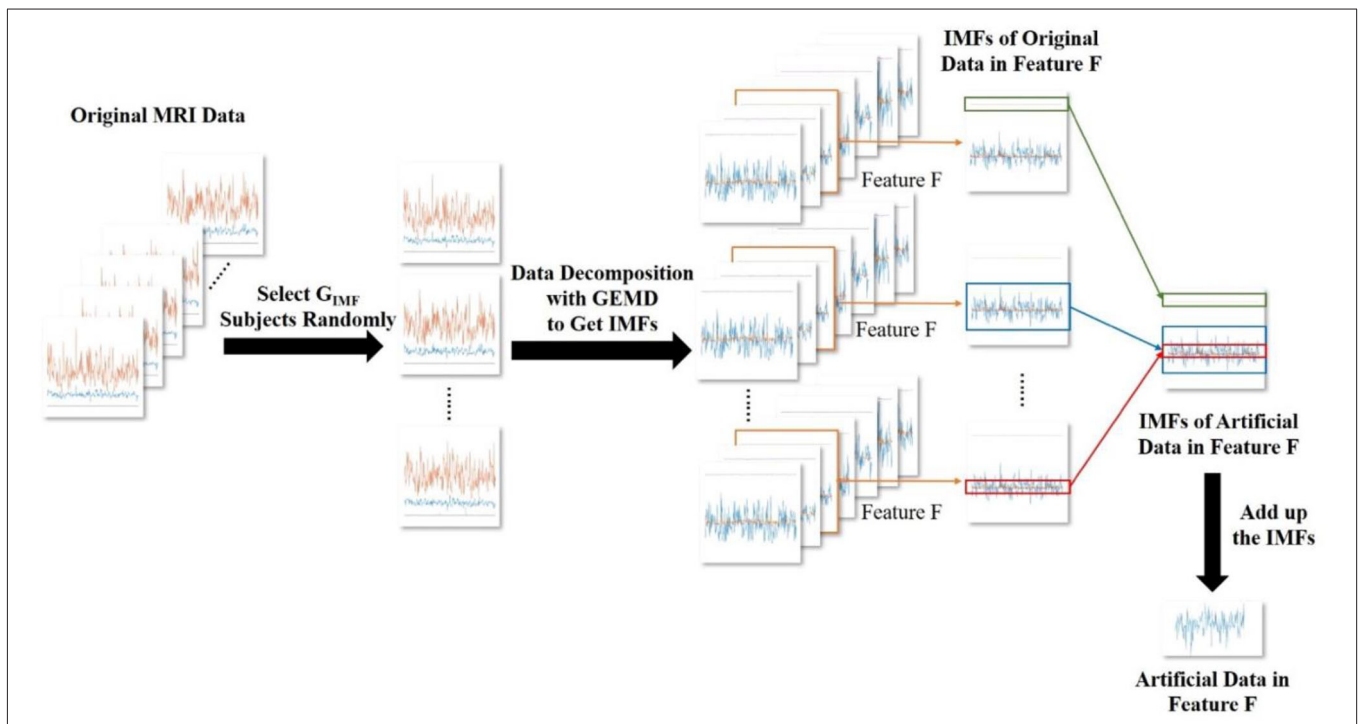


FIGURE 4 | The generation of artificial MRI data. Here, we generate the artificial data in feature F as an example. We randomly select M subjects from the original MRI data. Then, we obtain the IMFs, which are decomposed with GEMD. The IMFs decomposed from feature F of the M randomly selected subjects are recombined. Then, they are added up to obtain the artificial data of the feature F.

structural brain networks. E2E filter convolves the brain network adjacency matrix and weights edges of adjacent brain regions. E2N filter assigns each brain region a weighted sum of its edges. N2G assigns a single response based on all the weighted nodes.

These three filters consist of convolution kernels: kernel $c_1 \in R^{1 \times D}$, $c_2 \in R^{D \times 1}$. The kernel of the E2E filter is $c_{E2E} = c_1 + c_2 \in R^{D \times D}$. D is the number of nodes in a graph, which corresponds to the number of brain regions in this work. The kernels of the

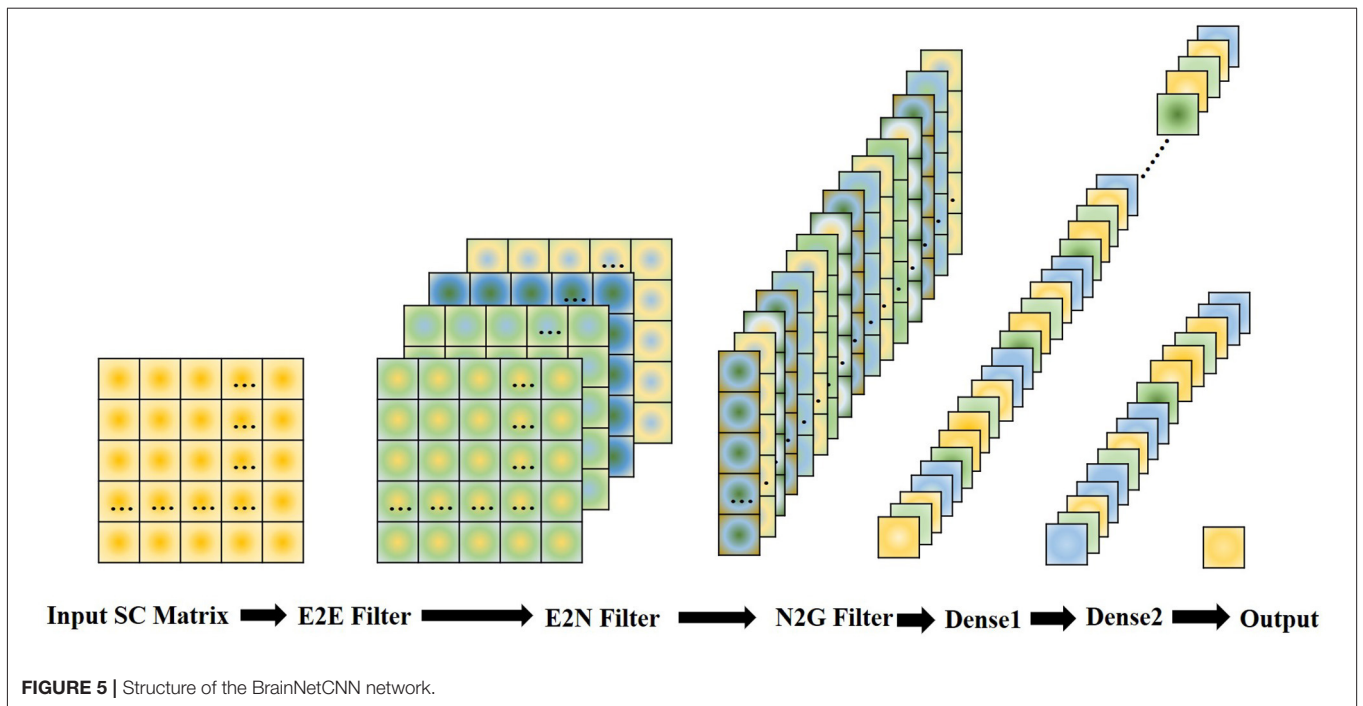


FIGURE 5 | Structure of the BrainNetCNN network.

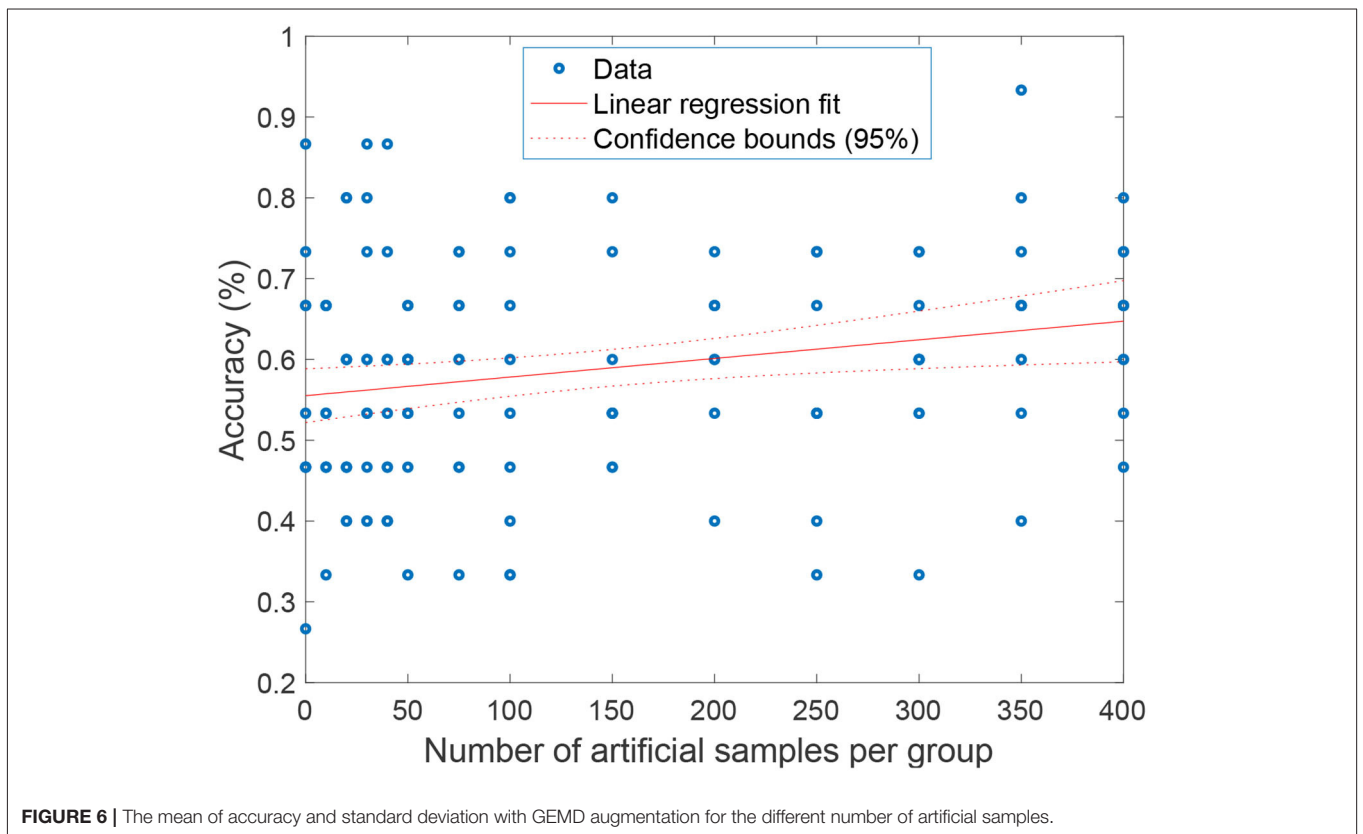
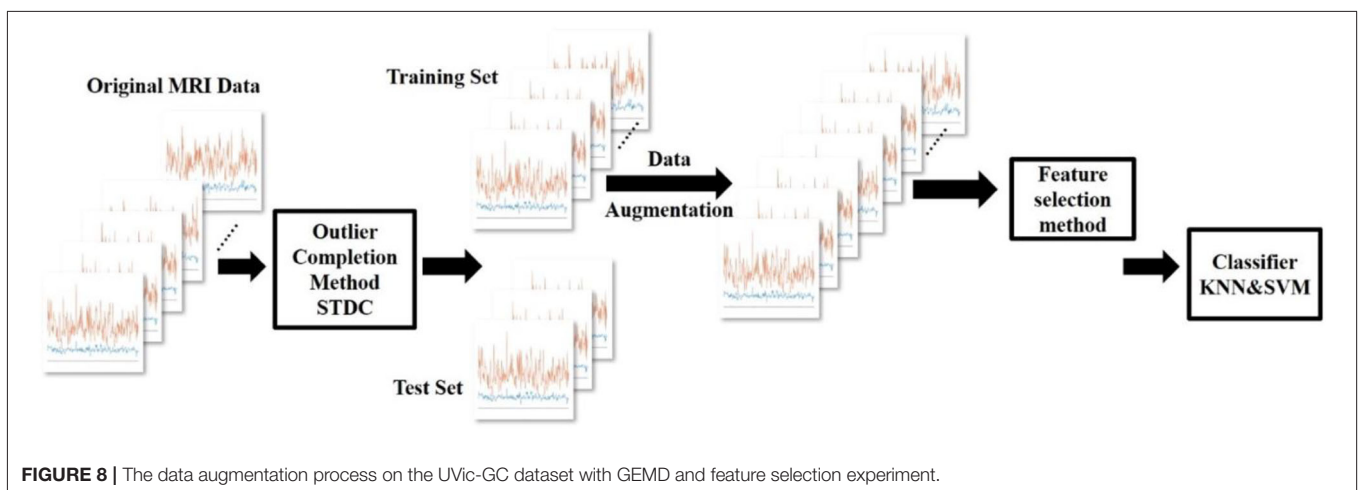
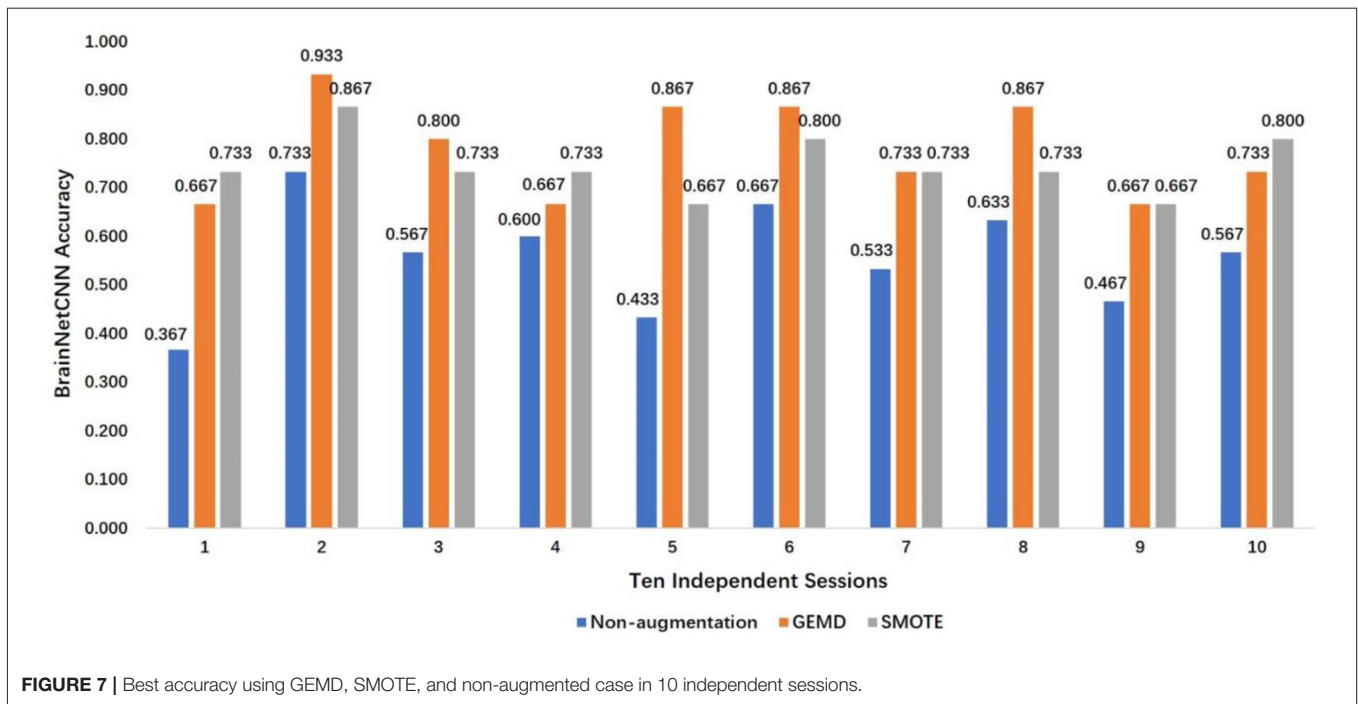


FIGURE 6 | The mean of accuracy and standard deviation with GEMD augmentation for the different number of artificial samples.

E2N filter and N2G filter are $c_{E2N} = c_1$, $c_{N2G} = c_2$. In our experiment, the structure of the BrainNetCNN can be simply expressed as Input (308×308 SC matrix) -> E2E (4 channels)

-> relu -> E2N (16 channels) -> relu -> N2G (32 channels) -> dense1 (16 channels) -> dense2 (1 channels). This structure is shown in **Figure 5**. We use the adaptive moment estimation



(Adam) optimizer, with learning rate $lr = 0.001$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. The network is trained using 300 epochs, and the batch size is 32. Considering the size of the dataset, we applied 10-fold cross-validation and repeated the experiment 10 times to get the average accuracy.

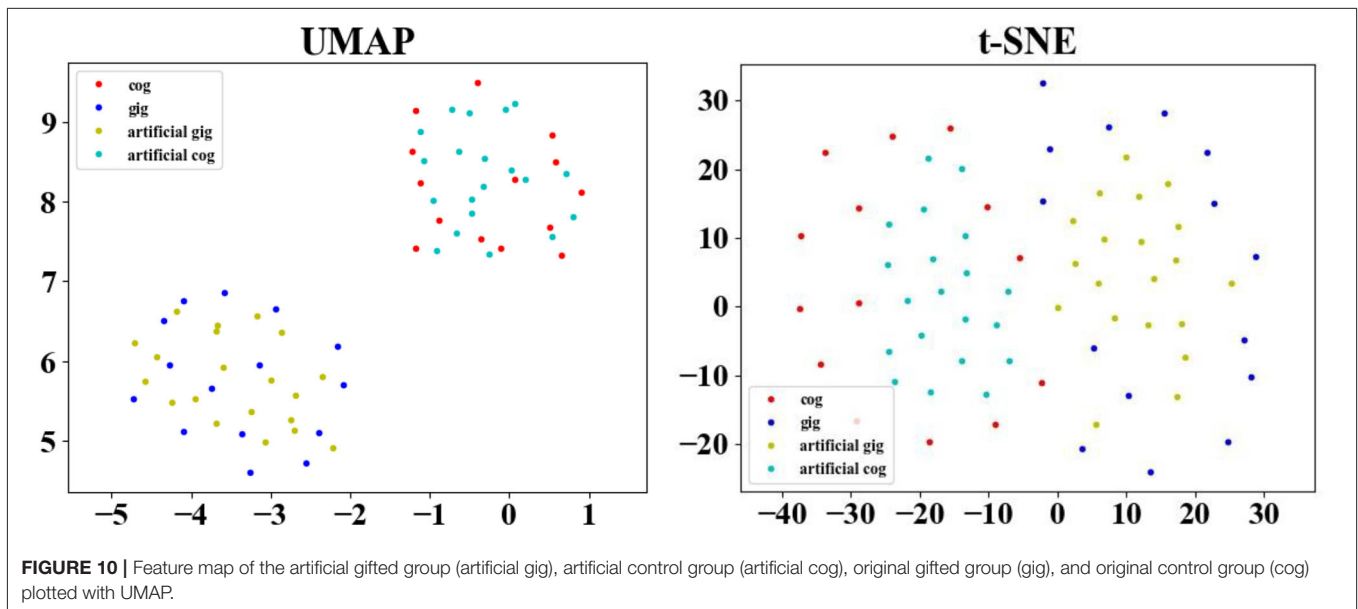
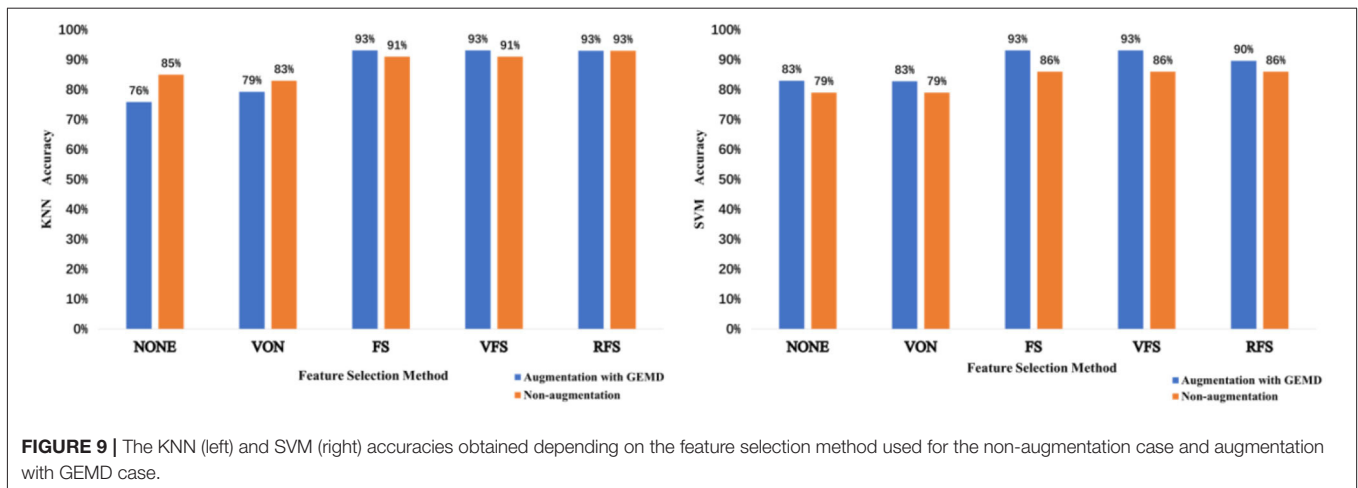
RESULTS

GEMD Performance on BrainNetCNN

We want to prove that the data augmentation with GEMD can improve the performance of the BrainNetCNN in the classification of the UVic-GC dataset. Therefore, we randomly selected 14 subjects (7 from the gifted group and 7 from the control group) as the original MRI data for the training set. The

training set also contains artificial MRI data generated through GEMD from the original data of this training set. The rest of the subjects are used as the test set, containing 15 subjects.

Aiming to study how the number of artificial subjects affects the performance in the training set, we increase the number of artificial samples from 0 to 400 for each group. For each session, the original MRI data are split into the training set and test set. The training set is used to generate the required number of artificial samples. The model is then trained using the original training set and the artificial samples generated from it, and finally the model is tested with the remaining test set. This process is repeated 10 times for each number of artificial samples to get the final average accuracy.



In **Figure 6**, we show the classification accuracy for a different number of artificial samples. As can be seen, the performance of the BrainNetCNN can be improved when adding artificial samples, from 10 artificial samples to 400 artificial samples. The improvement increases when the number of artificial samples increases. Fitting a linear regression model gives us an idea of the expected improvement when adding artificial samples. The model shows a positive trend of gradient $x_1 = 0.00023035$, with a p -value < 0.01 . This means that we should expect a 2.3% increase in the accuracy per 100 artificial samples added. The BrainNetCNN has the best mean accuracy performance at 66.7% when the number of artificial samples is 350, while without GEMD, the mean accuracy is only 56%. This is an increase of 10.7%, slightly better than the 8% predicted by the linear model.

The SMOTE is also used. The results are compared and depicted in **Figure 7**, which presents the best accuracy with GEMD, SMOTE, and non-augmented cases (baseline) in 10

different sessions. The accuracy is always improved, with respect of the non-augmented case, when GEMD and SMOTE are used. This emphasizes the importance of having more data to train the model. Specifically, GEMD shows higher classification accuracy than SMOTE in sessions 2, 3, 5, 6, and 8; while SMOTE has better performance in sessions 1, 4, and 10. In sessions 7 and 9, the accuracies of both GEMD and SMOTE methods are almost the same. In addition, a classification performance of 93.3% is obtained in session 2 by using GEMD, which is the best result obtained with this database so far. The average of the 10 sessions' best accuracy using GEMD achieves 78%, which is better than using SMOTE (74.7%) and the baseline case (55.7%).

GEMD Performance on Feature Selection Methods

In this section, we evaluate the performance of the GEMD using the procedures described in Zhang et al. (2021). In summary,

Zhang et al. proposed an outlier correction on the morphometric features based on the STDC algorithm (Chen et al., 2013) and explored several feature selection methods to classify MRI data from controls and gifted groups. These methods were applied to the UVic-GC dataset with outstanding performance.

According to Zhang et al. (2021), the NONE feature selection method used all the features in the raw feature matrix. The VON feature selection method used only the regions belonging to types 2 and 3 of the von Economo atlas (van den Heuvel et al., 2015), which corresponds to the associative areas of the brain. Choosing the top highest features selected with a threshold, from all the morphometric features and brain regions, was defined as the FS feature selection method. The rank F-score (RFS) method is a variation of the previous one in which, for each region, the FS values are sorted by descending order, where the morphometric features with the highest FS value are the selected ones. Finally, the combination of VON and FS will lead the VFS feature selection method, in which only type 2 and 3 regions are considered when calculating the FS value for morphometric features. Two traditional machine learning methods, KNN and SVM, were used as classifiers (Zhang et al., 2021), with leave-one-out as a cross-validation strategy.

The process of this experiment is shown in **Figure 8**. First, we use the outlier completion method STDC to compute the missing entries from the estimated latent factors. Then, we enlarge the training set of original MRI data by using GEMD. After that, we use feature selection methods NONE, VON, FS, VFS, and RFS to select different features. Finally, the model is trained by KNN and SVM for classification.

From **Figure 9**, we observe that using data augmentation with GEMD generally improves the performance of feature selection methods. For the SVN case (**Figure 9, right**), the GEMD method always improves the accuracy regardless of the feature selection method, while for the KNN case (**Figure 9, left**) only in two cases the accuracy is lower using artificial data. Note that for both KNN and SVM the classification accuracy reaches 93% using FS and VFS, which is the best result with this database, to the best of our knowledge.

DISCUSSION

In our study, we have used GEMD to enlarge the UVic-GC dataset. The motivation for exploring a data augmentation strategy is 2 fold. First, the UVic-GC dataset is small. Second, there are many parameters in the deep neural network that need to be learned from the data. Therefore, overfitting could appear due to insufficient amount of data.

We propose the GEMD augmentation method to solve the problems mentioned above in this work. We analyze the GEMD augmentation result in three aspects, namely, the influence of the number of artificial subjects, the classification accuracy between non-augmentation and augmentation, and the feature selection method used.

It can be seen from **Figure 6** that the accuracy shows an upward trend with the increase in the amount of artificial data. When the number of artificial data reaches 350, the classification accuracy achieves the maximum. Note that the result may vary considerably from experiment to experiment. This is due to the

non-convergence of the BrainNetCNN and the random factor added when selecting the data for each experiment. Prettier but unfair results could be shown by discarding the non-convergent experiments, for example, but we show the full set of results to point out these potential problems.

To clearly illustrate the distribution of the artificial data generated by GEMD, **Figure 10** depicts the original SC matrices, named the original gifted group (gig) and original control group (cog), and 20 artificial SC matrices of the artificial gifted group (artificial gig) and artificial control group (artificial cog). This figure uses Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) and Distributed Stochastic Neighborhood Embedding (van der Maaten and Hinton, 2008) for dimensionality reduction. It can be seen that the artificial data of each group are projected around the original data of the corresponding group, which is a way of showing that the artificial data are meaningful, i.e., the data generated by GEMD are consistent with the distribution of the original data. Furthermore, the two classes (control and gifted) in the two figures can be accurately separated. There is no obvious overlap between the two groups, explaining why the linear classifiers (SVM and KNN) combined with feature selection methods perform very well.

Even if our proposed method can augment the dataset so that the artificial data help improve the classification accuracy, we must highlight that the results of the BrainNetCNN are not stable. This is due to two main factors, the non-convergence of the model and the overfitting that appears despite the amount of artificial data generated. This is the main drawback of the proposed method. We are now investigating it and other possible neural network models with fewer parameters to improve the classification results when using a small number of original MRI subjects in the training dataset and artificial data generated with them. **Figure 10** shows that the artificial data created using the GEMD method are consistent with the original (real) data, which encourages us to use this method and improve the classification model.

CONCLUSIONS

Medical data such as MRI are difficult to obtain, and gifted children are rare in our society. Identifying gifted children from a small set of MRI data is not easy. At the same time, deep neural networks require a large amount of data to improve their performance. They cannot exert their full performance when the dataset is too small. In that case, our work provides a feasible solution by data augmentation. We use the UVic-GC dataset and artificial data generated by GEMD to train the BrainNetCNN neural network. This avoids using a feature selection method as we feed the model directly with the SC data. The results show that GEMD has a significant effect that improves the performance of the classifier. Furthermore, the GEMD data augmentation method can be extended to other similar small datasets. Our future work will focus on the application of GEMD on multisite MRI data, such as the Human Connectome Project data. Due to different scanner settings, parameters, and operators, the distribution of MRI data collected in various regions is different. We expect to be able to adjust the distribution of other datasets by domain adaptation. In that case, we can predict the classification

results of multiple MRI datasets using the trained model after augmentation with GEMD.

DATA AVAILABILITY STATEMENT

The raw anonymized MRI data is available in the OpenNeuro repository: <https://openneuro.org/datasets/ds001988>. The code for replication is available at: <https://github.com/CynthiaChern/GEMD-Based-Data-Augmentation-for-Gifted-Children-MRI-Dataset-Analysis> GraphEMD code is available at https://github.com/fkalaganis/graph_emd.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board (IRB00003099) of the University of Barcelona (Catalonia). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

ZS, CC, and JS-C: conceptualization. FD, ZS, CC, and JS-C: methodology and supervision. XC, BL, HJ, and FF: formal analysis and investigation. XC: writing. XC, BL, HJ, FF,

FD, ZS, CC, and JS-C: writing—review and editing. FD, CC, and JS-C: funding acquisition. FD and JS-C: resources. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported in part by the National Natural Science Foundation of China (Key Program) under Grant No. 11932013 and in part by the Tianjin Natural Science Foundation for Distinguished Young Scholars under Grant No. 18JCQJC46100. JS-C's work was partially based upon work from COST Action CA18106, supported by COST (European Cooperation in Science and Technology) and the University of Vic – Central University of Catalonia (R0947). CC's work was partially supported by grants PICT 2017-3208, PICT 2020-SERIEA-00457, UBACYT 20020190200305BA, and UBACYT 20020170100192BA (Argentina).

ACKNOWLEDGMENTS

We are grateful to Dr. Fotis Kalagnis for providing the GraphEMD code and for fruitful discussions about it. We also thank the reviewers for their suggestions and criticisms that have helped us to improve the paper.

REFERENCES

- Abdelaziz Ismael, S. A., Mohammed, A., and Hefny, H. (2020). An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artif. Intell. Med.* 102, 101779. doi: 10.1016/j.artmed.2019.101779
- Aubry, A., Gonthier, C., and Bourdin, B. (2021). Explaining the high working memory capacity of gifted children: contributions of processing skills and executive control. *Acta Psychol.* 218, 103358. doi: 10.1016/j.actpsy.2021.103358
- Bucaille, A., Jarry, C., Allard, J., Brochard, S., Peudener, S., and Roy, A. (2022). Neuropsychological profile of intellectually gifted children: a systematic review. *J. Int. Neuropsychol. Soc.* 28, 424–440. doi: 10.1017/S1355617721000515
- Caiafa, C. F., Solé-Casals, J., Marti-Puig, P., Zhe, S., and Tanaka, T. (2020). Decomposition methods for machine learning with small, incomplete or noisy datasets. *Appl. Sci.* 10, 8481. doi: 10.3390/app10238481
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi: 10.1613/jair.953
- Chen, Y.-L., Hsu, C.-T., and Liao, H.-Y. M. (2013). Simultaneous tensor decomposition and completion using factor priors. *IEEE Transac. Pattern Anal. Mach. Intell.* 36, 577–591. doi: 10.1109/TPAMI.2013.164
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Dinarès-Ferran, J., Ortner, R., Guger, C., and Solé-Casals, J. (2018). A new method to generate artificial frames using the empirical mode decomposition for an EEG-based motor imagery BCI. *Front. Neurosci.* 12, 308. doi: 10.3389/fnins.2018.00308
- Grady, L. J., and Schwartz, E. L. (2003). *Anisotropic Interpolation on Graphs: The Combinatorial Dirichlet Problem*. Boston, MA: Citeseer.
- Gras, R. M. L., Bordoy, M., Ballesta, G. J., and Berna, J. C. (2010). Creativity, intellectual abilities and response styles: Implications for academic performance in the secondary school. [Creatividad, aptitudes intelectuales y estilos de respuesta: implicaciones para el rendimiento académico en secundaria]. *Ann. Psychol.* 26, 212–219. Available online at: <https://doi.org/10.6018/analesps>
- Gross, M. U. M. (2006). Exceptionally gifted children: long-term outcomes of academic acceleration and nonacceleration. *J. Educ. Gifted* 29, 404–429. doi: 10.4219/jeg-2006-247
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London.* 454, 903–995. doi: 10.1098/rspa.1998.0193
- Kalaganis, F. P., Laskaris, N. A., Chatzilari, E., Nikolopoulos, S., and Kompatsiaris, I. (2020). A Data Augmentation Scheme for Geometric Deep Learning in Personalized Brain-Computer Interfaces. *IEEE Access* 8, 162218–162229. doi: 10.1109/ACCESS.2020.3021580
- Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V., Grunau, R. E., et al. (2017). BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage* 146, 1038–1049. doi: 10.1016/j.neuroimage.2016.09.046
- Kotu, V., and Deshpande, B. (2019). “Chapter 4 - Classification,” in *Data Science (Second Edition)*, eds. V. Kotu and B. Deshpande. *Morgan Kaufmann* p. 65–163. doi: 10.1016/B978-0-12-814761-0.00004-6
- Kuhn, T., Blades, R., Gottlieb, L., Knudsen, K., Ashdown, C., Martin-Harris, L., et al. (2021). Neuroanatomical differences in the memory systems of intellectual giftedness and typical development. *Brain Behav.* 11, e2348. doi: 10.1002/brb3.2348
- Leonardsen, E. H., Peng, H., Kaufmann, T., Agartz, I., Andreassen, O. A., Celius, E. G., et al. (2022). Deep neural networks learn general and clinically relevant representations of the ageing brain. *NeuroImage* 256, 119210. doi: 10.1016/j.neuroimage.2022.119210
- Ma, J., Kang, H. J., Kim, J. Y., Jeong, H. S., Im, J. J., Namgung, E., et al. (2017). Network attributes underlying intellectual giftedness in the developing brain. *Sci. Rep.* 7, 11321. doi: 10.1038/s41598-017-11593-3
- Muñoz-Gutiérrez, P. A., Giraldo, E., Bueno-López, M., and Molinas, M. (2018). Localization of active brain sources from EEG signals using empirical

- mode decomposition: a comparative study. *Front. Integr. Neurosci.* 12, 55. doi: 10.3389/fnint.2018.00055
- Navas-Sánchez, F. J., Alemán-Gómez, Y., Sánchez-Gonzalez, J., Guzmán-De-Villoria, J. A., Franco, C., Robles, O., et al. (2014). White matter microstructure correlates of mathematical giftedness and intelligence quotient. *Hum. Brain Map.* 35, 2619–2631. doi: 10.1002/hbm.22355
- Nguyen, K., Chin Fatt, C., Treacher, A., Mellema, C., Trivedi, M., and Montillo, A. (2020). *Anatomically Informed Data Augmentation for Functional MRI With Applications to Deep Learning*. (Houston, TX: SPIE). doi: 10.1117/12.2548630
- Qi, P., Ru, H., Gao, L., Zhang, X., Zhou, T., Tian, Y., et al. (2019). Neural Mechanisms of Mental Fatigue Revisited: New Insights from the Brain Connectome. *Engineering* 5, 276–286. doi: 10.1016/j.eng.2018.11.025
- Rocha, A., Almeida, L., and Perales, R. G. (2020). Comparison of gifted and non-gifted students' executive functions and high capabilities. *J. Educ. Gifted Young Sci.* 8, 1397–1409. doi: 10.17478/jegys.808796
- Romero-Garcia, R., Atienza, M., Clemmensen, L. H., and Cantero, J. L. (2012). Effects of network resolution on topological properties of human neocortex. *NeuroImage*. 59, 3522–3532. doi: 10.1016/j.neuroimage.2011.10.086
- Sarraf, S., and Tofighi, G. (2016). Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks. *arXiv preprint arXiv:1603.08631*.
- Seidlitz, J., Váša, F., Shinn, M., Romero-Garcia, R., Whitaker, K. J., Vértes, P. E., and NSPN Consortium (2018). Morphometric similarity networks detect microscale cortical organization and predict inter-individual cognitive variation. *Neuron*. 97, 231–247 doi: 10.1016/j.neuron.2017.11.039
- Solé-Casals, J., Serra-Grabulosa, J. M., Romero-Garcia, R., Vilaseca, G., Adan, A., Vilaró, N., et al. (2019). Structural brain network of gifted children has a more integrated and versatile topology. *Brain Struct. Function* 224, 2373–2383. doi: 10.1007/s00429-019-01914-9
- Tremblay, N., Borgnat, P., and Flandrin, P. (2014). "Graph Empirical Mode Decomposition", in: *2014 22nd European Signal Processing Conference (EUSIPCO)* (Lisbon: IEEE), p. 2350–2354.
- Ulloa, A., Plis, S., Erhardt, E., and Calhoun, V. (2015). "Synthetic structural magnetic resonance image generator improves deep learning prediction of schizophrenia", in: *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)* (Boston, MA: IEEE), p. 1–6. doi: 10.1109/MLSP.2015.7324379
- van den Heuvel, M. P., Scholtens, L. H., Barrett, L. F., Hilgetag, C. C., and de Reus, M. A. (2015). Bridging cytoarchitectonics and connectomics in human cerebral cortex. *J. Neurosci.* 35, 13943–13948. doi: 10.1523/JNEUROSCI.2630-15.2015
- van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605. Available online at: <https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf>
- Wei, M., Wang, Q., Jiang, X., Guo, Y., Fan, H., Wang, H., et al. (2020). Directed connectivity analysis of the brain network in mathematically gifted adolescents. *Comput. Intell. Neurosci.* 2020:10. doi: 10.1155/2020/4209321
- Zhang, J., Feng, F., Han, T., Duan, F., Sun, Z., Caiafa, C. F., et al. (2021). A hybrid method to select morphometric features using tensor completion and F-score rank for gifted children identification. *Sci. China Technol. Sci.* 64, 1863–1871. doi: 10.1007/s11431-020-1876-3
- Zhang, Z., Duan, F., Solé-Casals, J., Dinares-Ferran, J., Cichocki, A., Yang, Z., et al. (2019). A novel deep learning approach with data augmentation to classify motor imagery signals. *IEEE Access.* 7, 15945–15954. doi: 10.1109/ACCESS.2019.2895133

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chen, Li, Jia, Feng, Duan, Sun, Caiafa and Solé-Casals. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

3.3. Improving Pre-movement Pattern Detection with Filter Bank Selection



PAPER

Improving pre-movement pattern detection with filter bank selection

RECEIVED
19 January 2022REVISED
10 October 2022ACCEPTED FOR PUBLICATION
28 October 2022PUBLISHED
16 November 2022Hao Jia¹ , Zhe Sun^{2,*} , Feng Duan^{3,*} , Yu Zhang^{4,5,*} , Cesar F Caiafa^{6,*} and Jordi Solé-Casals^{1,7,*} ¹ Data and Signal Processing Research Group, University of Vic-Central University of Catalonia, Vic, Catalonia, Spain² Computational Engineering Applications Unit, Head Office for Information Systems and Cybersecurity, RIKEN, Saitama, Japan³ Tianjin Key Laboratory of Brain Science and Intelligent Rehabilitation, College of Artificial Intelligence, Nankai University, Tianjin, People's Republic of China⁴ Department of Bioengineering, Lehigh University, Bethlehem, PA 18015, United States of America⁵ Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA, 18015, United States of America⁶ Instituto Argentino de Radioastronomía, CONICET CCT La Plata/CIC-PBA/UNLP, V. Elisa, Argentina⁷ Department of Psychiatry, University of Cambridge, Cambridge CB2 3EB, United Kingdom

* Authors to whom any correspondence should be addressed.

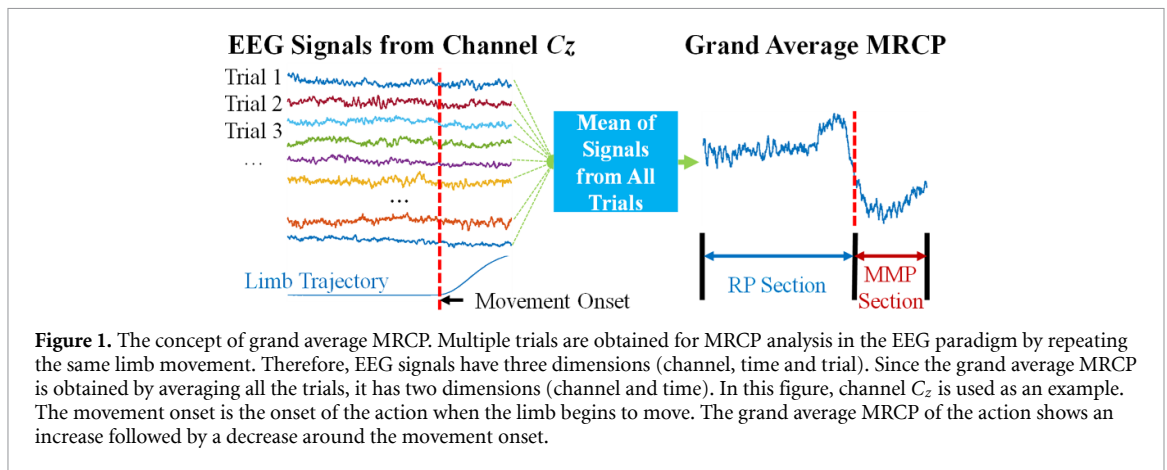
E-mail: zhe.sun.vk@riken.jp, duanf@nankai.edu.cn, yuzhang@lehigh.edu, ccaiafa@fi.uba.ar and jordi.sole@uvic.cat**Keywords:** brain computer interface, movement detection, pre-movement decoding, standard task-related component analysis, filter bank selection**Abstract**

Objective. Pre-movement decoding plays an important role in detecting the onsets of actions using low-frequency electroencephalography (EEG) signals before the movement of an upper limb. In this work, a binary classification method is proposed between two different states. *Approach.* The proposed method, referred to as filter bank standard task-related component analysis (FBTRCA), is to incorporate filter bank selection into the standard task-related component analysis (STRCA) method. In FBTRCA, the EEG signals are first divided into multiple sub-bands which start at specific fixed frequencies and end frequencies that follow in an arithmetic sequence. The STRCA method is then applied to the EEG signals in these bands to extract CCPs. The minimum redundancy maximum relevance feature selection method is used to select essential features from these correlation patterns in all sub-bands. Finally, the selected features are classified using the binary support vector machine classifier. A convolutional neural network (CNN) is an alternative approach to select canonical correlation patterns. *Main Results.* Three methods were evaluated using EEG signals in the time window from 2 s before the movement onset to 1 s after the movement onset. In the binary classification between a movement state and the resting state, the FBTRCA achieved an average accuracy of 0.8968 ± 0.0847 while the accuracies of STRCA and CNN were 0.8228 ± 0.1149 and 0.8828 ± 0.0917 , respectively. In the binary classification between two actions, the accuracies of STRCA, CNN, and FBTRCA were 0.6611 ± 0.1432 , 0.6993 ± 0.1271 , 0.7178 ± 0.1274 , respectively. Feature selection using filter banks, as in FBTRCA, produces comparable results to STRCA. *Significance.* The proposed method provides a way to select filter banks in pre-movement decoding, and thus it improves the classification performance. The improved pre-movement decoding of single upper limb movements is expected to provide people with severe motor disabilities with a more natural, non-invasive control of their external devices.

1. Introduction

Movements of the human limbs lead to potential changes on the human scalp, which can be observed with non-invasive brain-computer interface-based electroencephalography (EEG) signals [1, 2]. In previous studies on movement detection with EEG signals, motor imagery (MI) is one of the most frequently used brain activities in the motor cortex

[3–5]. When the limbs begin moving, the power of EEG signals in alpha rhythm (frequency range: 8–12 Hz) and beta rhythm (frequency range: 13–30 Hz) shows an upward or downward trend, which is called event-related desynchronization/synchronization [6]. When humans imagine a movement of the left or right limb, the power changes in the left/right half of the scalp. However, these power changes in MI occur after the limb moves, which implies that in



MI analysis movement can only be detected after the onset of the imagined movement [6]. Looking at the brain activity through the movement-related cortical potential (MRCP), the movement of human limbs or the resting state before the movement onset can be evaluated. Hence, MRCP is expected to help enhance the restoration of useful motor functions and reduce the time delay of movement detection [7, 8].

MRCP is a type of low-frequency EEG signal (frequency range: 0.5–10 Hz) acquired in the motor cortex [9–11]. MRCP analysis is applied to EEG signals located around the movement onset. The readiness potential (RP) section is the stage that starts from 2 s before the onset and ends on the onset [12], while the movement monitoring potential (MMP) section is the stage that starts on the onset and ends 1 s after that [12]. The pre-movement patterns decoded from the RP section in MRCP signals cannot be observed directly. Grand average MRCP is a way to visualize the pre-movement patterns (figure 1). In grand average MRCP, EEG signals acquired from the motor cortex are averaged across trials. The grand average MRCP of the upper limb movement shows an increase followed by a rapid decrease around the onset compared to the relatively steady grand average MRCP of the resting state. The pre-movement patterns are the features extracted from the RP section based on the grand average MRCP [13, 14].

By analyzing the grand average MRCP, some previous works focused on the binary classification between a movement state and the resting state or the binary classification between two actions. For instance, Jeong *et al* proposed the subject-dependent and section-wise spectral filtering method (SSSF) to extract the amplitude features in MRCP and successfully solved the two-class problem between movement and resting states [15]. This method uses the mean amplitude of MRCP signals in both RP and MMP sections as the features. To optimize the selected features, Jeong *et al* adopted a cross-validation and testing method to select the best frequency range for each subject. Ofner *et al* proposed the discriminative spatial pattern method (DSP) [16], which calculates

Table 1. Binary classification results in pre-movement decoding.

Method	Movement vs Resting	Movement vs Movement
STRCA [13]	0.8287 ± 0.1101	0.5970 ± 0.1424
DCNN [14]	0.9030 ± 0.0560	0.6247 ± 0.0070
SSSF [15]	0.7300 ± 0.0783	—
DSP [16]	0.8500 ± 0.0500	0.4400 ± 0.0700

a LDA classifier for every time step. It was shown that the accuracy increases as the time point approaches the onset of the action in both the RP and MMP sections. Mammone *et al* proposed the deep convolutional neural network (DCNN) [14], which decodes pre-movement patterns from time–frequency maps of EEG signals at the source level. Duan *et al* proposed a pre-movement pattern decoding method, the standard task-related component analysis (STRCA) [13] consisting of the task-related component analysis (TRCA) spatial filter and the canonical correlation patterns (CCPs) features. All the methods mentioned above, except for the SSSF method, are also a solution to the binary classification between two actions. In table 1, the binary classification results of each method are given.

Although STRCA has a very concise structure, it faces the frequency range selection problem when decoding pre-movement patterns in MRCP analysis. In both the SSSF and DCNN methods, the frequency characteristics of the EEG signals is considered when optimizing the two classification methods. The frequency characteristics are optimized by either using filter bank selection with cross-validation and testing or by constructing a time–frequency map [14, 15]. Considering this, it can be seen that STRCA could be further improved with the filter bank technique.

Filter bank selection aims to solve the feature selection problem among various sub-bands in the frequency domain. It is widely used to analyze brain activity such as in MI and in steady-state visually evoked potential (SSVEP). In SSVEP analysis, the canonical correlation analysis method is a classical method used for detecting stimulus frequencies

[17]. Canonical correlation analysis can measure the similarity between EEG signals and the reference signals, and many methods in SSVEP analysis have been developed based on this technique [18–20]. Filter bank canonical correlation analysis was proposed to incorporate harmonic and fundamental frequency components, which improved the detection of standard canonical correlation analysis in SSVEP [21]. Without the filter bank technique, the canonical correlation analysis faces the problem of selecting frequency components. In MI analysis, the common spatial pattern method is the most classical one [22]. The method extracts the logarithm-variance features from the EEG signals filtered by the spatial filter, and it shows a varying accuracy among the sub-bands in alpha (8–12 Hz) and beta (13–30 Hz) rhythms [3, 23]. Filter bank common spatial pattern is an advanced MI analysis method that was developed by combining the common spatial pattern method and the filter bank technique [24]. The method is able to avoid sub-band selection, thus achieving better and more stable accuracies than the common spatial pattern method.

In both MI and SSVEP analysis, the filter bank technique uses a feature selection method to optimize the extracted common spatial pattern features or the canonical correlation features in each sub-band. The optimal frequency range of the filter bank varies among the subjects due to individual differences. The feature selection method overcomes the frequency range selection problem and enables the classification to achieve a stable and accurate result. When applying the filter bank technique to STRCA, there are three problems to tackle:

- (a) The frequency range setting is unknown, so it is unclear how the starting and stopping frequencies of the sub-bands in the filter bank technique can be selected.
- (b) The feature selection method for STRCA is undetermined.
- (c) The feature arrangement is unclear when applying the feature selection method on STRCA features extracted from all sub-bands.

This study aims to analyze how to incorporate the filter bank technique into STRCA. Two steps are adopted for the improvement of the STRCA method in this work: firstly, three feature range settings are compared to decide how to select the frequency range of each sub-band in pre-movement decoding; secondly, a new filter bank TRCA (FBTRCA) method is proposed to decode the pre-movement patterns for the binary classification between a movement state and the resting state or between two actions.

FBTRCA consists of four steps: frequency bank division, spatial filtering, feature selection and classification. In the first step, the EEG signals are band-passed into multiple sub-bands in the low-frequency domain. In the second step, canonical correlations

are extracted from each of these sub-bands by the STRCA method. In the third stage, a feature selection algorithm is used to select the essential features from the features of all bands automatically. In the fourth step, a classifier is used to classify the selected features. This paper presents a selection of feature selection methods and classifiers for use in FBTRCA, and recommends suitable feature selection and classifiers for MRCP-based brain-computer interface.

In section 2, the EEG dataset and the data pre-processing mechanism used are introduced, and the proposed FBTRCA method is described. In section 3, the proposed method is analyzed in terms of the frequency range settings, the feature selection and when compared to other methods. In section 4, the FBTRCA design and workings in pre-movement decoding is discussed. Finally, section 5 contains the conclusions for this study.

To facilitate the understanding of the contents in this work, the abbreviations are given in table 2.

2. Material and method

2.1. Dataset description

There are two public datasets used in this work, namely dataset I and dataset II [16, 31]⁷. Both datasets follow an offline acquisition paradigm in which a trial lasts 5 s. At the start of a trial, the computer screen displays a cross and emits a beeping sound. The computer screen then shows a cue that indicates the required movement or resting 2 s later. When the cue occurs, the subjects implement movements or remain at rest.

The EEG signals are acquired from 11 channels with active electrodes. These channels are located around the motor cortex. According to the 10/20 international system, 5 out of the 11 electrodes are located at the center of the motor cortex: FC_z , C_3 , C_z , C_4 , CP_z ; while the remaining 6 electrodes are located surrounding the motor cortex: F_3 , F_z , F_4 , P_3 , P_z , P_4 . The EEG signals are filtered with an 8-order Chebyshev bandpass filter from 0.01 Hz to 200 Hz. The sample rate of the EEG signals is 512 Hz, and the signals are downsampled to 256 Hz considering the computational load. A notch filter at 50 Hz is applied to avoid the influence of power line interference.

There are two main differences between the two datasets.

Firstly, dataset I contains signals related to hand movement trajectories acquired using a glove sensor. The onsets of the actions can be located with the movement trajectories of the limb. Dataset II, on the other hand, does not contain information about limb movement trajectories.

⁷ <http://bnci-horizon-2020.eu/database/data-sets>, Dataset I: 25. Upper limb movement decoding from EEG (001-2017) [16]; Dataset II: 26. Attempted arm and hand movements in persons with spinal cord injury (001-2019) [31].

Table 2. Descriptions of abbreviations.

Abbreviation	Full name	Description
EEG	Electroencephalograph	Multi-channel signals acquired from the surface of brain scalp.
MRCP	Movement-related cortical potential	A kind of brain activity related to pre-movement.
MI	Motor imagery	A kind of brain activity related to movement.
SSVEP	Steady state visual-evoked potential	A kind of brain activity evoked by visual stimulus.
RP	Readiness potential	EEG signals in the 2 s window before the movement onset.
MMP	Movement monitoring potential	EEG signals in the 1 s window after the movement onset.
CCA	Canonical correlation analysis	A basic classification method in SSVEP [17].
FBCCA	Filter bank canonical correlation analysis	A method that optimizes CCA by filter bank selection [21].
CSP	Common spatial pattern	A basic classification method in MI [22].
FBCSP	Filter bank common spatial pattern	A method that optimizes CSP by filter bank selection [24].
SSSF	Subject-dependent and section-wise spectral filtering	A binary classification method for movement and resting states [15].
STRCA	Standard task-related component analysis	A binary classification method for movement and resting states [13].
FBTRCA	Filter bank tasked-related component analysis	The method that optimizes STRCA by filter bank selection.
TRCA	Task-related component analysis	The spatial filter used in STRCA [13].
CCP	Canonical correlation pattern	The extracted features in STRCA [13].
CNN	Convolutional neural network	A feature selection method consists of convolutional layers.
MIQ	Mutual information quotient	A feature selection method based on mutual information [25].
MAXREL	Maximum relevance	A feature selection method based on mutual information [26].
MINRED	Minimum redundancy	A feature selection method based on mutual information [26].
MRMR	Minimum redundancy maximum relevance	A feature selection method based on mutual information [26].
QPFS	Quadratic programming feature selection	A feature selection method based on mutual information [27].
CIFE	Conditional infomax feature extraction	A feature selection method based on mutual information [28].
CMIM	Conditional mutual information minimization	A feature selection method based on mutual information [29].
MRMTR	Maximum relevance minimum total redundancy	A feature selection method based on mutual information [30].
SVM	Support vector machine	A binary classifier
LDA	Linear discriminant analysis	A binary classifier
NN	Neural network	A binary classifier

Secondly, both datasets have different actions and number of subjects. Dataset I consists of 7 states with 15 subjects. These states include the resting state *rest* and six actions: *elbow flexion*, *elbow extension*, *supination*, *pronation*, *hand close* and *hand open*. For each action, 60 trials were acquired during the signal acquisition. Dataset II, on the other hand, consists of EEG signals from nine subjects. Each subject was asked to implement five actions, including *supination*, *pronation*, *hand open*, *palmar grasp* and *lateral grasp*. Each action has 72 trials.

In dataset I, the onset can be located from the hand trajectory when the movement is executed. However, the onset cannot be located in dataset II. Therefore, different processing procedures were adopted in the two datasets.

2.1.1. Pre-processing in dataset I

The STRCA method has been evaluated on dataset I in previous studies [13]. Here, we adopt the same pre-processing procedure. In dataset I, the hand movement trajectory is used to locate the movement onsets

of the actions. The 1-order difference of the trajectory is taken, and then the 1-order Savitzky–Golay finite impulse response smoothing filter is used to smooth the signals. The length of the time window in the smoothing filter is set to 31. The starting value of the trajectory is subtracted from the trajectory in each trial. The approximate range of the onsets of the actions is the 3 s time window with a 1 s delay after the cue.

The two motions related to elbow movement, which are *elbow flexion* and *elbow extension*, lead to an increase in the amplitude of the hand trajectory. The trajectory is first changed into the absolute value. The hand trajectories are normalized by dividing them by the maximal absolute value. The location where the normalized trajectory is larger than the threshold of 0.05 is regarded as the movement onset. In trials that contain heavy noise contamination, the onsets of these actions cannot be located, and therefore these are manually removed.

For the other four states, the approximate range of the onset shrinks to a 2 s time window with a

Table 3. Average number of trials across subjects after trial rejection.

Motion	<i>Elbow flexion</i>	<i>Elbow extension</i>	<i>Supination</i>	<i>Pronation</i>	<i>Hand close</i>	<i>Hand open</i>	<i>Resting</i>
Number	60	59	52	51	56	55	59

1 s delay after the cue. The hand trajectory has a lower amplitude and is heavily influenced by noise. In these trials, trajectories are first normalized by dividing them by the maximal absolute value of each trajectory. The function $f(x) = a * \exp(-(\frac{x-b}{c})^2) + d$ is used to fit the smoothed and normalized trajectories by tuning the parameters a, b, c, d . The symbol 'exp' denotes the exponential function. Trials that fulfill $a < 0.05, c > 100$ and $d > 10$ are rejected. The onsets of the actions are determined by a threshold criterion, as the bias d is removed from the fitted function $f(x)$ and the onset is set to the location where the value of $f(x)$ is larger than 0.1.

For the signals in the resting state, the amplitude of the hand trajectory is supposed to be steady and have a small variance. The trials are rejected if the variances of the trajectories are greater than the set threshold of 0.02. The trajectories in the resting state have no movement onset. A fake onset is set to 2.5 s following the beeping sounds. In table 3, the number of trials after eliminating the rejected ones is given. The number of trials of these motions is averaged among all trials in dataset I and is rounded to an integer in the table.

The EEG signals can be divided into the RP and MMP sections with the located onsets or fake onsets. The features extracted from the RP section are the pre-movement patterns. In dataset I, we analyze the classification in two cases. In the first case, the EEG signals are from the RP section, and the results are used to analyze the performance of the proposed method in pre-movement decoding. In the second case, the EEG signals are from the RP and MMP sections. In figure 1, the grand average MRCP shows an increasing trend, so therefore we assume that the EEG signals from both RP and MMP sections may improve the performance compared to the EEG signals from only the RP section.

2.1.2. Pre-processing in dataset II

The onsets of the actions in dataset II cannot be located by movement trajectory. Here, we adopt the same processing procedure in MI as in [32]. EEG signals are extracted from the 2 s time window after the cue. The onset is located within this time window, but the precise location is unknown.

Dataset II has five trial-based actions. Compared to the resting state of dataset I, the EEG signals in the resting state are not trial-based. Subjects were asked to have a long-duration rest after acquiring the EEG signals of the actions. The resting state for dataset II is

generated by dividing the long-duration resting-state EEG signals into multiple trials. Each trial lasts 2 s, and there are 72 trials for each subject in total.

The data obtained was denoted as $\mathbf{X} \in \mathbb{R}^{N_c \times N_s \times N_t}$ or $X(t) \in \mathbb{R}^{N_c \times N_t}, t = 1, \dots, N_s$, where N_c is the EEG channel number, N_s is the sample time and N_t is the number of trials. Before the binary classification tasks, EEG signals were normalized by z-score normalization. When evaluating STRCA and the proposed FBTRCA methods with this dataset, ten-fold cross-validation was applied, and the classification performance was calculated as the mean of these ten-folds. The binary classification was implemented between two motions, e.g. *elbow flexion vs elbow extension* and *elbow flexion vs resting* in dataset I. Therefore, dataset I has 21 motion pairs, while dataset II has 15 motion pairs.

2.2. FBTRCA

2.2.1. STRCA

STRCA is used to classify the EEG signals between a movement state and the resting state with MRCP signals in the RP section [13]. The method consists of two components: (a) spatial filter TRCA and (b) CCP features. The extracted features are classified with the linear discriminated analysis (LDA) classifier. Figure 2 illustrates the structure of STRCA.

2.2.1.1. TRCA

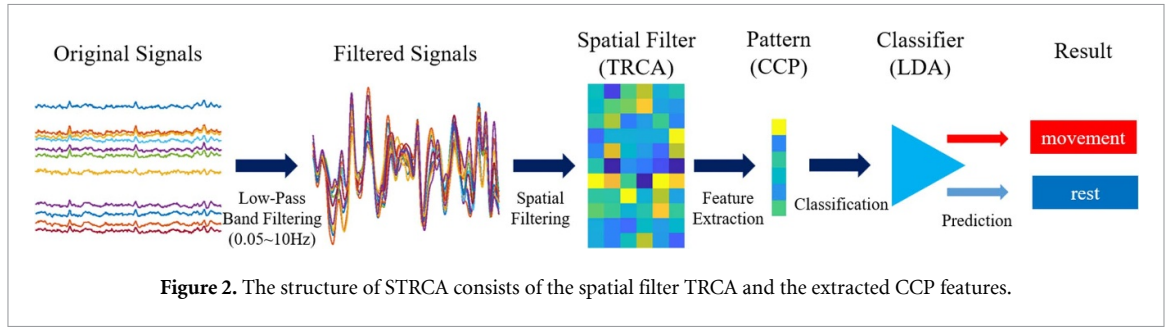
The spatial filter of TRCA is designed by maximizing the reproducibility during the task [33]. In multichannel EEG signals, the training set is supposed to be $X^k(t) \in \mathbb{R}^{N_c \times N_t}$, where k refers to the class of the EEG signals, $k = 1, 2$. $X(t)$ consists of two kinds of signals: (a) task-related signals $s(t) \in \mathbb{R}$ and (b) task-unrelated noise $n(t) \in \mathbb{R}$. The relationship between $X(t)$, $s(t)$ and $n(t)$ is expressed by:

$$X_{i,j}^k(t) = a_{1,i,j}^k s(t) + a_{2,i,j}^k n(t), \quad i = 1, \dots, N_c, j = 1, \dots, N_t. \quad (1)$$

$y(t)$ is the linear sum of EEG signals $X(t)$, and is defined as:

$$y_j^k(t) = \sum_{i=1}^{N_c} w_i^k X_{i,j}^k(t), j = 1, \dots, N_t. \quad (2)$$

In TRCA, the task-related signal $s(t)$ is recovered from $y(t)$. The ideal solution is difficult to calculate but can be approached by maximizing the inter-trial



covariance. The covariance C_{j_1, j_2}^k between the j_1 th trial and the j_2 th trial can be computed using:

$$\begin{aligned} C_{j_1, j_2}^k &= \text{Cov}(y_{j_1}^k(t), y_{j_2}^k(t)) \\ &= \sum_{i_1, i_2}^{N_c} w_{i_1}^k w_{i_2}^k \text{Cov}(X_{i_1, j_1}^k(t), X_{i_2, j_2}^k(t)). \end{aligned} \quad (3)$$

The covariances of all the trials are summed to obtain a combination of all trials:

$$\begin{aligned} \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^{N_t} C_{j_1, j_2}^k &= \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^{N_t} \text{Cov}(y_{j_1}^k(t), y_{j_2}^k(t)) \\ &= \sum_{\substack{j_1, j_2=1 \\ j_1 \neq j_2}}^{N_t} \sum_{i_1, i_2=1}^{N_c} w_{i_1}^k w_{i_2}^k \text{Cov}(X_{i_1, j_1}^k(t), X_{i_2, j_2}^k(t)) \\ &= \mathbf{w}^T S^k \mathbf{w}. \end{aligned} \quad (4)$$

To avoid infinite solutions of \mathbf{w} , the variance of $y_j^k(t)$ is constrained to 1:

$$\begin{aligned} \sum_{j_1, j_2=1}^{N_t} C_{j_1, j_2}^k &= \sum_{j_1, j_2=1}^{N_t} \text{Cov}(y_{j_1}^k(t), y_{j_2}^k(t)) \\ &= \sum_{j_1, j_2=1}^{N_t} \sum_{i_1, i_2=1}^{N_c} w_{i_1}^k w_{i_2}^k \text{Cov}(X_{i_1, j_1}^k(t), X_{i_2, j_2}^k(t)) \\ &= \mathbf{w}^T Q^k \mathbf{w}. \end{aligned} \quad (5)$$

The constrained spatial filter can be obtained by maximizing the generalized eigenvalue equation J , which is expressed as:

$$J = \frac{\mathbf{w}^T S^k \mathbf{w}}{\mathbf{w}^T Q^k \mathbf{w}}. \quad (6)$$

Eigenvectors are obtained by solving the generalized eigenvalue problem. The eigenvectors with the largest eigenvalues are selected as the eigenvectors that are to be used in the spatial filter. Three eigenvectors are adopted in TRCA. These eigenvectors from two classes are then combined into the TRCA spatial filter. The TRCA spatial filter that we obtained is $W \in \mathbb{R}^{N_c \times 6}$.

2.2.1.2. CCP

Using the training set of EEG data, $\mathbf{X}^k \in \mathbb{R}^{N_c \times N_s \times N_t}$, $k = 1, 2$, we can obtain the CCP templates $\hat{X}^k = \sum_{j=1}^{N_t} \mathbf{X}^k / N_t \in \mathbb{R}^{N_c \times N_s}$, $k = 1, 2$ for each of the two classes. The EEG signal of the trial from which we aim to extract features is $X \in \mathbb{R}^{N_c \times N_s}$. Given the TRCA spatial filter W , we extract the CCP after the EEG signals are transformed with W . Three kinds of correlation coefficients are considered in STRCA:

(a) Correlation coefficients between filtered signals:

$$X_k = \hat{X}^k; X_* = X; \quad (7)$$

$$\rho_{1,k} = \text{corr}(X_*^T W, X_k^T W), \quad k = 1, 2; \quad (8)$$

(b) Correlation coefficients between filtered signals with a canonical correlation analysis projection:

$$X_k = \hat{X}^k; X_* = X; \quad (9)$$

$$[A_k, B_k] = \text{cca}(X_*^T W, X_k^T W) \quad (10)$$

$$\rho_{2,k} = \text{corr}(X_*^T W B_k, X_k^T W B_k), \quad k = 1, 2; \quad (11)$$

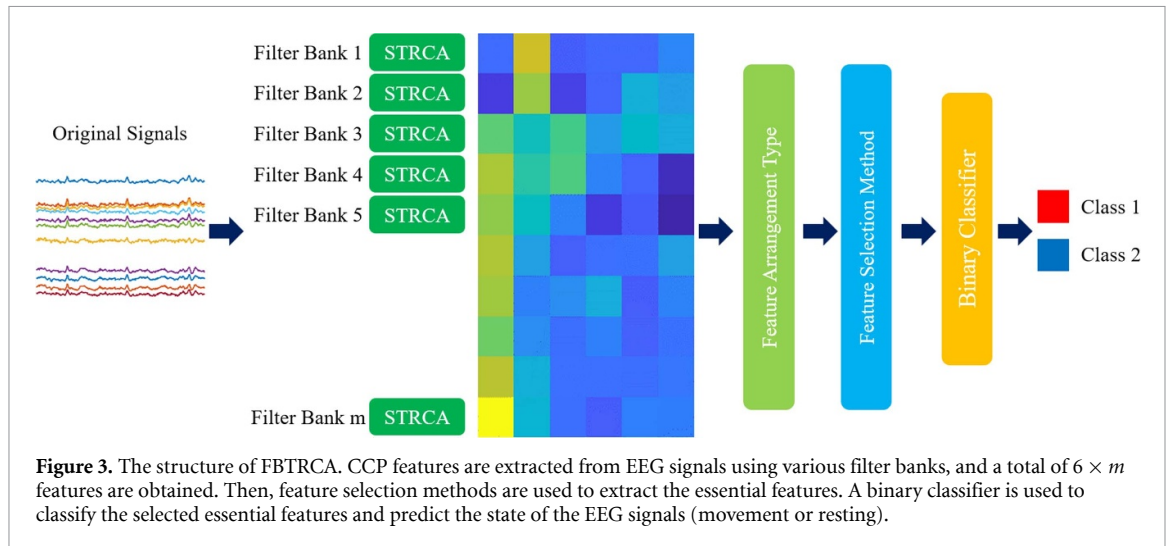
(c) Correlation coefficients between the distances of filtered signals:

$$X_k = \hat{X}^k - \hat{X}^{3-k}; X_* = X - \hat{X}^{3-k}; \quad (12)$$

$$[A_k, B_k] = \text{cca}(X_*^T W, X_k^T W) \quad (13)$$

$$\rho_{3,k} = \text{corr}(X_*^T W A_k, X_k^T W A_k), \quad k = 1, 2. \quad (14)$$

In the above equations, canonical correlation analysis is used to optimize the correlation between the templates \hat{X}^k and X . The function symbols *corr* and *cca* indicate the calculation process of the correlation coefficient and the process of CCA analysis, respectively. For each trial, we can obtain six features, which are referred to as CCP features in the following section.



2.2.2. FBTRCA

This study proposes an FBTRCA method to enhance pattern decoding in MRCP analysis. Figure 3 shows the flowcharts of the proposed method, which consists of three major procedures: (a) filter bank analysis, (b) CCP feature extraction and (c) feature selection.

First, in the filter bank technique, the sub-bands are decomposed with multiple filters that have different pass-bands. In this study, the bandpass filter used for extracting sub-band components from the original EEG signals was an 8-order infinite impulse Butterworth filter.

STRCA is then applied to each sub-band separately, resulting in six CCP features. The number of sub-bands is denoted as m , such that the number of CCP features extracted from all sub-bands is $6 \times m = 6m$. The essential features are extracted from the $6m$ features in all sub-bands using one of the feature arrangement types. The feature arrangement type refers to the arrangement of the $6m$ CCP features when the feature selection method is applied.

Finally, the selected essential features are classified with the binary classifier. This study compares two classifiers, including the LDA and the support vector machine (SVM).

2.2.3. Frequency range settings

In the decomposition of sub-bands, the decomposed EEG signals and classification accuracies vary with different frequency range settings of the filters. In the MI analysis, the frequency range of the filter banks was equipped with equally spaced bandwidths in alpha and beta rhythms, e.g. 4–8 Hz, 8–12 Hz, ..., 36–40 Hz [32]. In the SSVEP analysis, the frequency range of the filter banks started at $n \times 8$ Hz and ended at a fixed frequency, e.g. 8–88 Hz, 16–88 Hz, ..., 80–88 Hz, $n = 1, 2, \dots, 10$ [21].

In MRCP analysis, the frequency range setting is undetermined. Considering that MRCP signals are a type of EEG signal with low frequencies, the

maximum high cut-off frequency is set to 10 Hz. Three frequency arrangement settings are compared, including M_1 , M_2 and M_3 .

- M_1 figure 4(a): The frequency range setting in M_1 is similar to that in FBCSP but with different low cut-off and high cut-off frequencies. The sub-bands in M_1 are equipped with equally spaced bandwidths.
- M_2 figure 4(b): The frequency range setting in M_2 corresponds to the harmonic frequency bands. The high cut-off frequency is twice as high as the low cut-off frequency.
- M_3 figure 4(c): The frequency range setting in M_3 is similar to the best setting in FBCCA. One of the two ends of the sub-bands is a fixed value. The low cut-off is fixed because the MRCP signals are in a low-frequency band. The high cut-off frequencies are arranged as an arithmetic sequence.

2.2.4. Feature arrangement types

From each of the sub-bands, six CCP features are extracted. The total number of features is therefore $6 \times m = 6m$. When selecting essential features from these using feature selection methods, there are two feature arrangement types that were used (figure 5).

Type 1: The feature selection method is applied individually to each feature in the CCP. The feature selection method selects K_1 essential features out of m features, and is applied six times. In the end, $6 \times K_1$ essential features are selected in total. The maximum value of K_1 is 100.

Type 2: The feature selection method is applied to all six features in CCP simultaneously. The feature selection method is applied only once, and K_2 essential features are selected from $6 \times m$ features. The maximum value of K_2 is $6 \times m = 600$.

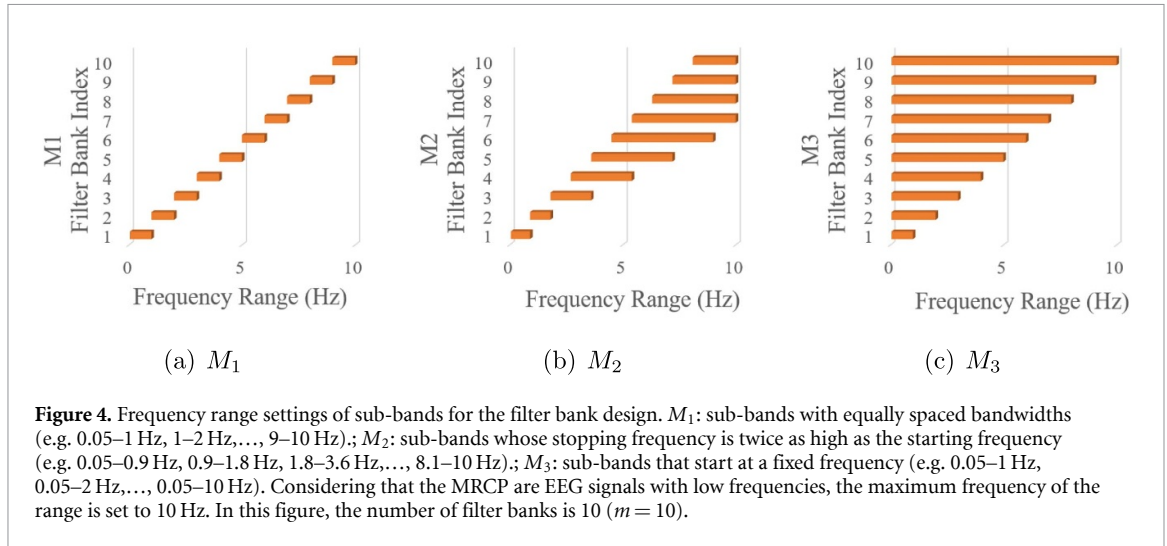


Figure 4. Frequency range settings of sub-bands for the filter bank design. M_1 : sub-bands with equally spaced bandwidths (e.g. 0.05–1 Hz, 1–2 Hz, ..., 9–10 Hz); M_2 : sub-bands whose stopping frequency is twice as high as the starting frequency (e.g. 0.05–0.9 Hz, 0.9–1.8 Hz, 1.8–3.6 Hz, ..., 8.1–10 Hz); M_3 : sub-bands that start at a fixed frequency (e.g. 0.05–1 Hz, 0.05–2 Hz, ..., 0.05–10 Hz). Considering that the MRCPs are EEG signals with low frequencies, the maximum frequency of the range is set to 10 Hz. In this figure, the number of filter banks is 10 ($m = 10$).

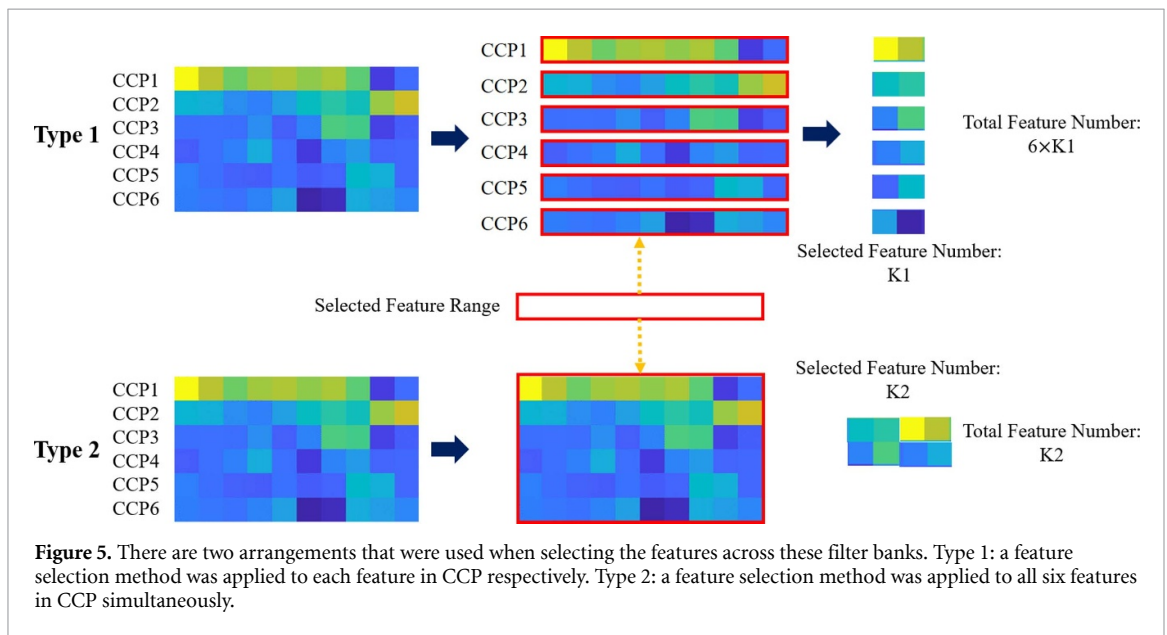


Figure 5. There are two arrangements that were used when selecting the features across these filter banks. Type 1: a feature selection method was applied to each feature in CCP respectively. Type 2: a feature selection method was applied to all six features in CCP simultaneously.

2.2.5. Feature selection methods

Mutual-information-based approaches are a popular feature selection paradigm in data mining. In the FBCSP method of MI analysis, feature selection based on mutual information plays a significant role in optimizing the CSP features in all sub-bands. This study compares eight mutual-information-based feature selection methods to find a suitable one for selecting CCP features in multiple sub-bands. The compared feature selection methods include:

- (a) Mutual information quotient (MIQ) [25]
- (b) Maximum relevance (MAXREL) [26]
- (c) Minimum redundancy (MINRED) [26]
- (d) Minimum redundancy maximum relevance (MRMR) [26]
- (e) Quadratic programming feature selection (QPFS) [27]
- (f) Conditional infomax feature extraction (CIFE) [28]
- (g) Conditional mutual information minimization (CMIM) [29]

- (h) Maximum relevance minimum total redundancy (MRMTR) [30]

2.2.6. Binary classifiers

In STRCA, two binary classifiers have been compared, including the LDA and the SVM. Because the number of features of STRCA is fixed, the simple LDA classifier shows the best performance among the three classifiers explored in [13]. In the proposed FBTRCA method, however, there are more than six features, and the number of features changes due to the feature selection settings. As a result, the kernel-based SVM classifier may show better performance when dealing with hyper-dimension features. Therefore, the two classifiers are also compared in this work.

2.3. Benchmark method

The CNN is used universally for feature selection. A neural network can be adjusted to an unknown function by backpropagation. The convolution unit in CNN can capture the local features of given inputs and thus select the essential features. Considering that

the size of the EEG dataset is small, the CNN method with a simple architecture is used as the benchmark method for feature selection.

In the CNN architecture, the CCP features are the input of the neural network and are regarded as a 6-channel image. The height and width of the images are the numbers of the low cut-off frequencies and the high cut-off frequencies, respectively. A two-dimensional convolution layer is used to extract essential features from these CCP features. This convolution layer has 24 filters, each of size 3×3 . A batch norm layer and a ReLu layer are used to normalize the output of the convolution layer. A 2×2 max pool layer with stride 2 follows the ReLu layer. Finally, a full-connect neural network with a hidden size of 50 is used as the binary classifier. The output of the hidden layer is normalized with the batch norm. The network is trained with an Adam optimizer with a learning rate of 0.001. The maximum training epoch is set to 200.

2.4. Performance measurement

In binary classification, accuracy, F1-score and cross-entropy loss are three prevalent measurements for classification performance. In the classification task, the model will be tested after training. There are four outcomes in the testing result: true positive (TP), true negative (TN), false positive (FP) and false negative (FN) [34]. The definitions of these four outcomes for the binary classification are given as

- TP = the number of cases is correctly identified as one class;
 FP = the number of cases is incorrectly identified as one class;
 TN = the number of cases is correctly identified as the other class;
 FN = the number of cases is incorrectly identified as the other class.

The accuracy measures the ratio of the correctly predicted trials in the testing set, and is calculated through the following expression:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (15)$$

In the calculation of F1-score, two measurements are considered: precision and recall. The precision is defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (16)$$

whereas the recall is defined as

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (17)$$

F1-score is given by combining both precision and recall:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (18)$$

Both the two classes have an F1-score, which are therefore denoted as $F1_1$ and $F1_2$. The macro-average F1-score (macroAVG) is used as a measurement which balances the F1-scores of the two classes:

$$\text{macroAVG} = 0.5 \times (F1_1 + F1_2). \quad (19)$$

Cross-entropy loss refers to the contrast between two random variables. It shows how accurate the classification model is by defining the difference between the estimated probability and the true label. The higher the difference between the two label outputs, the higher the loss. The cross-entropy loss is defined as

$$\text{CrossEntropy} = L(\mathbf{y}, \mathbf{t}) = - \sum_{k=1}^2 \mathbf{t}_k \ln(\mathbf{y}_k), \quad (20)$$

where \mathbf{t} is the true label and \mathbf{y} is the estimated probability. When measuring the classification performance of FBTRCA, the estimated probability is replaced with the predicated label of FBTRCA.

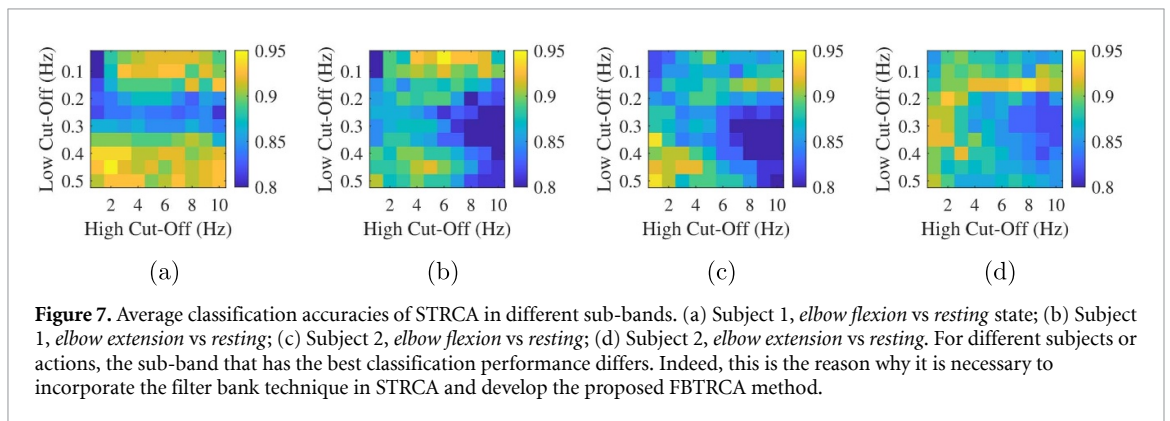
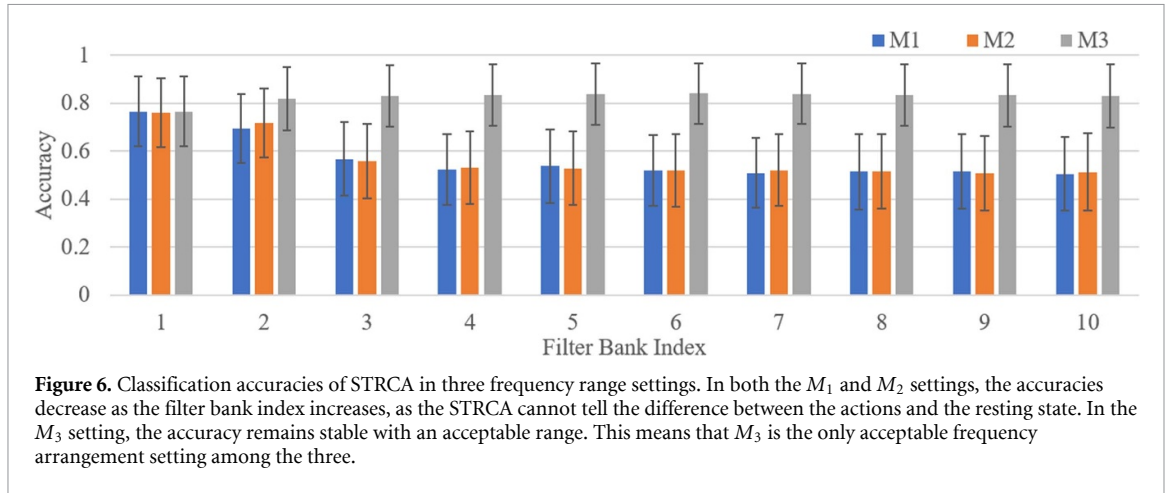
For the performance evaluation, ten-fold cross-validation is used. For each of the subjects and motion pairs, the accuracy, macroAVG and cross-entropy are averaged across ten-folds to get the mean and the deviation. When presenting the classification performance of the motion pairs, we average the means and the deviations of all the subjects. The final evaluation scores for each method are obtained by averaging the means and deviations across all subjects and motion pairs. We also use the two-side t -test to measure the improvement from STRCA to FBTRCA. The p -value is measured from the results of the ten-folds, for each of the subjects and motion pairs.

3. Results

The proposed FBTRCA method is evaluated with the two datasets. EEG signals are divided into the RP and MMP sections in the first dataset, while in the second dataset, the signals are in the 2 s time window after the cue. The result analysis is carried out in three cases: when the EEG signals are from (a) the RP section in dataset I, (b) both the RP and MMP sections in dataset I, and (c) the 2 s time window after the cue in dataset II. The performance of the classification methods is evaluated by ten-fold cross-validation.

This study aims to incorporate the filter bank technique into STRCA and thus propose a new methodology, FBTRCA. Three steps are necessary to achieve this goal:

- (a) decide on the frequency range settings;
- (b) evaluate the parameters K1 and K2 in the two types of feature arrangements;
- (c) compare results achieved through FBTRCA against the benchmark.



In the first step, the properties of each filter bank will be determined, including the number of filter banks and their individual frequency ranges. The second step will evaluate K1 and K2 for each mutual-information-based feature selection method. The effects of LDA and SVM on FBTRCA are also compared in the second step. In the third step, the best performance of the FBTRCA method is compared to those achieved by the CNN and STRCA methods in the three cases presented above.

3.1. Analysis of the frequency range settings

STRCA is applied to filter banks in three different settings: M_1 , M_2 and M_3 . Figure 6 shows the classification accuracies of each filter bank in the three settings. The performance is evaluated with the binary classification between the movement and resting states in the RP section of dataset I. The mean of the 6 (actions) \times 15 (subjects) \times 10 (folds) accuracies is taken to evaluate the classification performance.

In setting M_1 , the accuracy decreases to 0.5 as the filter bank index increases. Similarly, the accuracy of the M_2 setting follows the same trend. The STRCA fails to solve the binary classification of the EEG signals in the sub-bands without low frequencies. Therefore, the two frequency range settings are not suitable for the combination of STRCA and the filter bank technique. In the M_3 setting, however, the accuracies in the filter banks are acceptable.

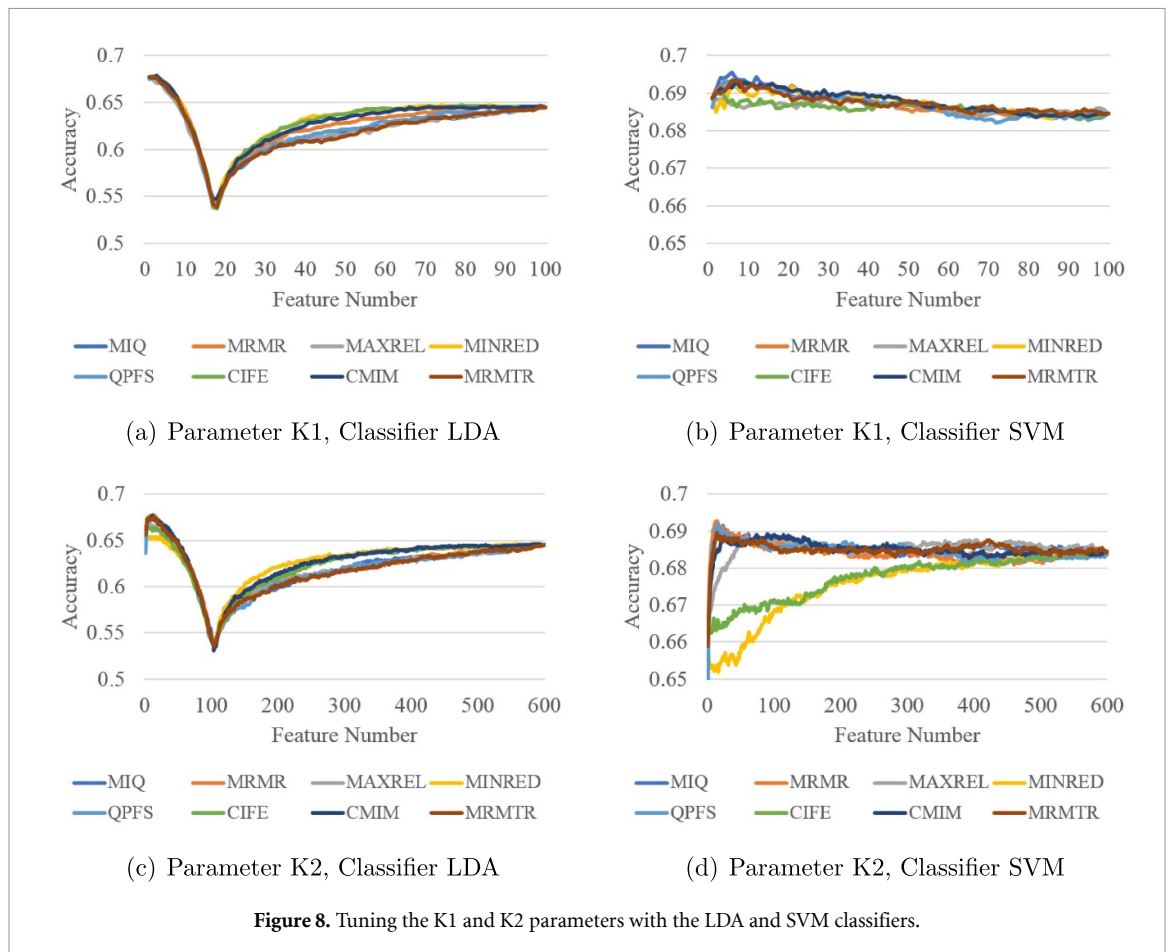
The main difference between M_3 and either M_1 or M_2 is that the frequency ranges of the filter banks in M_3 cover the sub-bands at low frequencies. The sub-bands at low frequencies maintain the information necessary for STRCA.

In the sight of these results, the M_3 frequency range setting is adopted with modifications in the design of the filter banks for the proposed FBTRCA method. The low cut-off frequency is shifted slightly from 0.5 Hz to 0.05 Hz with step 0.05 Hz, and the high cut-off frequency remains the same as that in M_3 . The total adopted number of sub-bands is 10×10 .

In figure 7, the classification accuracies of STRCA in these sub-bands are given. The sub-bands with the highest accuracy vary for different subjects and actions. It is hard to decide on a suitable sub-band for STRCA, and for this reason, FBTRCA is proposed to solve this problem.

3.2. Analysis on feature selection methods

After applying STRCA to 100 sub-bands, 6×100 CCP features are extracted. Then, feature selection methods select the essential features with a certain feature arrangement type. These essential features are classified with binary classifiers. In figure 8, the classification performances of eight mutual-information-based feature selection methods are compared on two feature arrangement types. The essential features are classified with the LDA classifier and the



SVM classifier (linear kernel). The statistics shown in figure 8 are the average accuracies across subjects and motion pairs in the RP section of dataset I. The motion pairs include both (a) the binary classification between two actions and (b) the binary classification between a movement state and the resting state.

The results in figure 8 are analyzed from three different perspectives: (a) binary classifiers, (b) comparison among the mutual-information-based feature selection methods, and (c) parameter searching on K1 and K2.

SVM and LDA are both basic binary classifiers used in machine learning. LDA casts the features into two classes through a linear projection, while SVM converts the features into hyper-space using a linear kernel and then casts the features in hyper-space into two classes. Because of the kernel, SVM is more efficient than LDA when tackling complicated features. In figure 8, it can be seen that SVM has a better classification performance than LDA. The accuracies of LDA decrease sharply in figures 8(a) and (c), and the best accuracy of LDA in figure 8(a) is slightly lower than the accuracy of SVM in figure 8(b). Therefore, the SVM classifier is better in the classification of FBTRCA.

In figure 8, eight feature selection methods based on mutual information are compared. These methods have similar accuracies except for MINRED (yellow

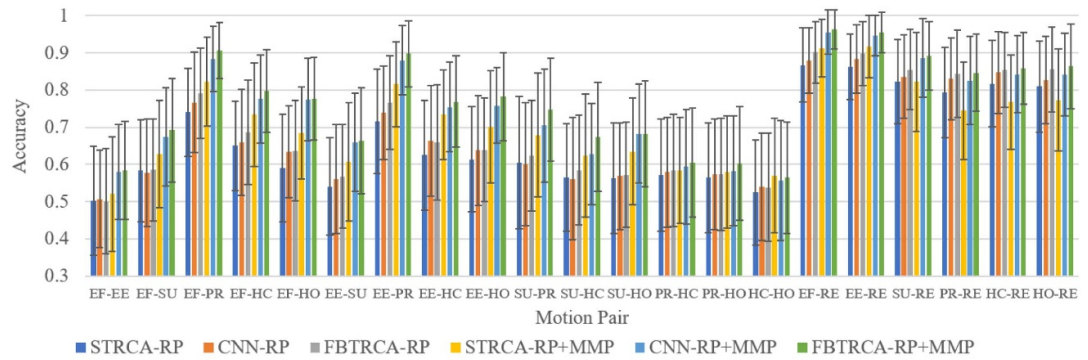
line) and CIFE (green line). The accuracies of the other six methods have a similar (a) best accuracy and (b) changing trend. In the following analysis, the MRMR method is used as the feature selection method based on mutual information.

The ranges of K1 and K2 are 0–100 and 0–600, respectively. Despite their difference in range, their best accuracies are the same in figures 8(b) and (d). There is no significant difference between the two feature arrangement types when selecting essential features using methods based on mutual information.

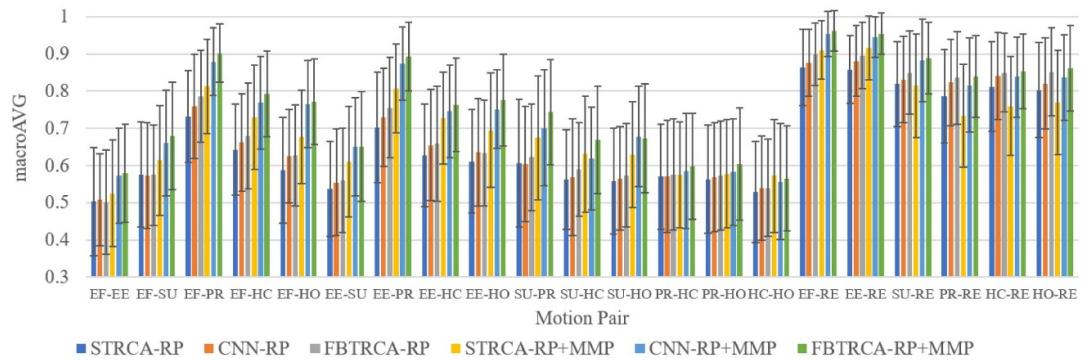
To conclude on the above three points, the best procedure for the FBTRCA method has three steps. First, EEG signals are divided into 10×10 low-frequency banks according to the M_3 frequency range setting. Second, the feature selection method MRMR is used to select essential features. Finally, the selected essential features are classified using the SVM classifier. The number of selected essential features is about 15.

3.3. Comparison against benchmarks

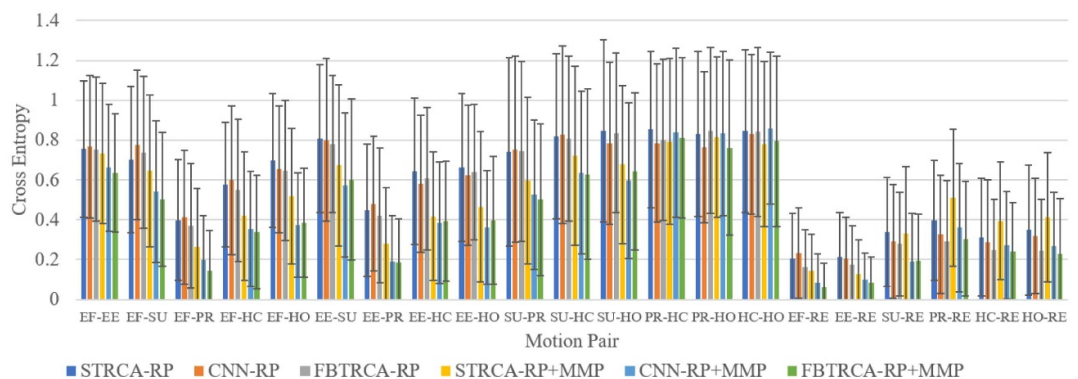
CNN is a universal feature selection method based on machine learning. The proposed FBTRCA method is compared against the STRCA and the CNN methods. The structure of the CNN model is given in section 2.3. The input of the CNN method is $6 \times 10 \times 10$ CCP features, which include CCP features in



(a) Dataset I, Accuracy



(b) Dataset I, MacroAVG



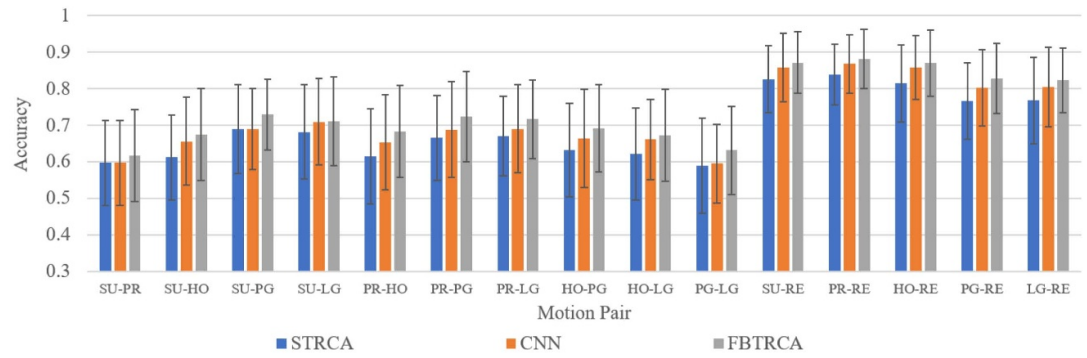
(c) Dataset I, Cross Entropy

Figure 9. Accuracy, macroAVG and cross-entropy loss comparison in dataset I. Three methods including STRCA, CNN and FBTRCA are compared. The classification is evaluated in either the RP section or both the RP and MMP sections. The accuracies are averaged across all subjects and all folds. The x-label represents the motion pair in which the binary classification is applied. The abbreviations of these states are used in the figures to facilitate the presentation; for example, ‘EF’ is short for the movement state *elbow flexion*. Figures (a)–(c) each refer to one of the three measurements respectively.

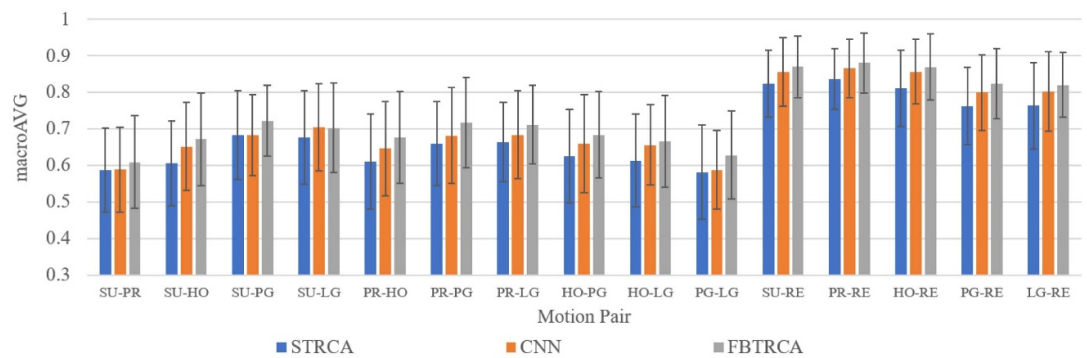
10 × 10 low-frequency filter banks. The input of the STRCA is the EEG signals in 0.05–10 Hz.

A comparison between the STRCA, CNN and FBTRCA methods can be made using the accuracy, macroAVG and cross-entropy loss measurements, which are averaged across all subjects and all folds. In figure 9, the results for the two cases in the dataset I are given, where the classification methods are applied to EEG signals in the RP section or to both the RP and MMP sections. The x-axis refers to the motion pairs, and the abbreviations are short for the names of actions. For example, ‘EF’ is short for ‘*elbow*

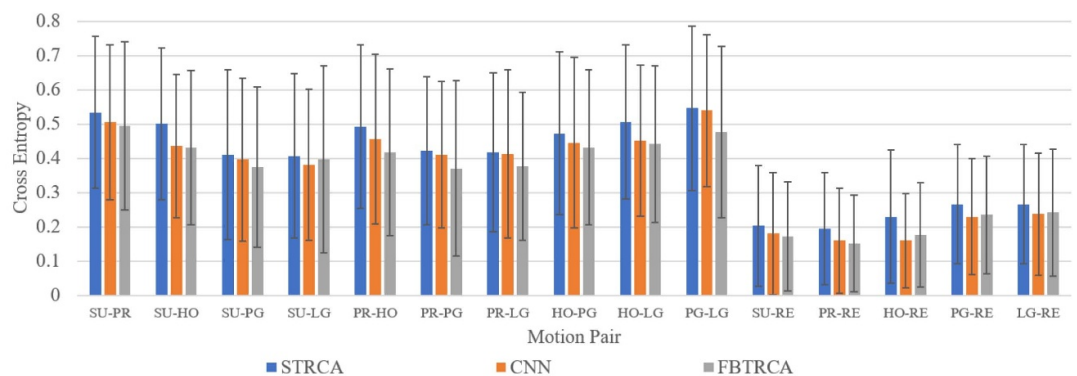
flexion’. Figure 10 presents the classification results of dataset II. The classification methods are applied to the EEG signals in the 2 s time window after the cue. Overall, it can be observed that the FBTRCA method outperforms the STRCA and CNN methods in both datasets. To better present the classification performance, we list the detailed accuracies for all subjects and motion pairs in the datasets I and II (tables 4–15). The detailed accuracies are results averaged from accuracies of ten-folds of cross-validation. The mean and the deviation are listed in two separate tables.



(a) Dataset II, Accuracy



(b) Dataset II, MacroAVG



(c) Dataset II, Cross Entropy

Figure 10. Accuracy, macroAVG and cross-entropy loss comparison in dataset II. Three methods including STRCA, CNN and FBTRCA are compared. The classification is evaluated in the 2 s time window after the cue. The other settings of this figure are the same as those presented in figure 9.

In dataset I, the binary classification is evaluated on the EEG signals in the RP section. The binary classification between actions and the resting state reflects whether the subjects want to move their limbs or stay at rest. The meaning of pre-movement decoding is to detect the movement intention before the limb moves. FBTRCA can improve the performance of pre-movement decoding compared to the previous STRCA method. When EEG signals in both RP and MMP sections are used, all three classification methods show an improved performance compared to the results for signals in only the RP section. The EEG

signals are also different between the two states in the MMP section.

During the acquisition of EEG signals in dataset II, the limb movement trajectories were not recorded, and thus the onsets of actions cannot be located precisely. The 2 s time window after the cue is taken in the classification evaluation. Within this time window, the subjects begin to move their limbs. Therefore, the 2 s time window covers part of both the RP and MMP sections. Although the onsets cannot be located in dataset II, the FBTRCA still improves the classification performance.

Table 4. Mean of classification accuracies of dataset I (RP) across ten-folds (Movement vs Movement). The best case is highlighted in bold.

Subject	Method	Motion pair										Average					
		EF-EE	EF-SU	EF-PR	EF-HC	EF-HO	EE-SU	EE-PR	EE-HC	EE-HO	SU-PR		SU-HC	SU-HO	PR-HC	PR-HO	HC-HO
1	STRCA	0.5167	0.6000	0.7545	0.6364	0.5818	0.4400	0.6273	0.6091	0.4636	0.7333	0.6556	0.5333	0.5400	0.5600	0.5800	0.5888
	CNN	0.5083	0.6600	0.8636	0.6091	0.6545	0.6500	0.7455	0.5727	0.4636	0.6667	0.6222	0.5556	0.5400	0.5700	0.5200	0.6135
	FBTRCA	0.5333	0.6800	0.8455	0.7091	0.6182	0.6000	0.8000	0.5909	0.4818	0.7222	0.6667	0.6111	0.6000	0.5700	0.5200	0.6366
2	STRCA	0.5300	0.5700	0.7500	0.5500	0.4700	0.5600	0.7500	0.5100	0.4800	0.6800	0.5700	0.4600	0.6300	0.6300	0.5000	0.5760
	CNN	0.5300	0.5500	0.7900	0.4800	0.5000	0.5500	0.8000	0.5400	0.6100	0.7400	0.4900	0.5200	0.6100	0.5900	0.4800	0.5853
	FBTRCA	0.5700	0.5800	0.8400	0.5100	0.4800	0.6500	0.8300	0.4800	0.6200	0.7200	0.5500	0.5100	0.6100	0.6300	0.5000	0.6053
3	STRCA	0.4000	0.5700	0.7000	0.8455	0.7900	0.6900	0.6700	0.8909	0.7800	0.6000	0.6556	0.6125	0.6444	0.6500	0.6778	0.6784
	CNN	0.4250	0.6400	0.6900	0.8364	0.8200	0.6500	0.7000	0.8818	0.8200	0.5750	0.6111	0.6875	0.6222	0.5125	0.6556	0.6751
	FBTRCA	0.4583	0.5500	0.7700	0.8545	0.8400	0.6100	0.8100	0.9000	0.8400	0.5750	0.6778	0.6750	0.6444	0.6250	0.7000	0.7020
4	STRCA	0.4417	0.6000	0.8364	0.7750	0.6636	0.5000	0.8364	0.7250	0.7000	0.7600	0.6727	0.6500	0.5727	0.5900	0.5455	0.6579
	CNN	0.5667	0.5091	0.8636	0.7917	0.7273	0.5182	0.8636	0.7750	0.8091	0.7800	0.6727	0.7100	0.5727	0.5900	0.5455	0.6863
	FBTRCA	0.5500	0.4818	0.9182	0.8083	0.7545	0.5636	0.8545	0.8167	0.7818	0.8600	0.8091	0.7300	0.6000	0.6200	0.6182	0.7178
5	STRCA	0.6273	0.4727	0.7000	0.5727	0.4818	0.6000	0.7500	0.5800	0.6000	0.6600	0.5500	0.6000	0.6300	0.5300	0.5500	0.5936
	CNN	0.4636	0.5727	0.7091	0.5636	0.5818	0.5600	0.7100	0.6700	0.5700	0.5400	0.5700	0.5300	0.6000	0.6000	0.5700	0.5874
	FBTRCA	0.5091	0.5182	0.6909	0.5455	0.5909	0.6500	0.7200	0.6400	0.6200	0.5700	0.6000	0.5300	0.5900	0.6200	0.5200	0.5943
6	STRCA	0.5833	0.6545	0.9333	0.6636	0.8182	0.5000	0.8111	0.6000	0.7091	0.6000	0.5700	0.5800	0.5875	0.6000	0.6000	0.6540
	CNN	0.5167	0.6182	0.8667	0.6909	0.7818	0.5545	0.7889	0.6182	0.7455	0.5750	0.5800	0.5400	0.6625	0.6250	0.6200	0.6523
	FBTRCA	0.5083	0.7000	0.9222	0.7000	0.8091	0.5545	0.8000	0.6182	0.7273	0.5625	0.5400	0.5800	0.7000	0.5750	0.5900	0.6591
7	STRCA	0.5167	0.6000	0.7091	0.6182	0.5000	0.4545	0.7364	0.6364	0.6091	0.6900	0.5400	0.5300	0.5700	0.5400	0.5100	0.5840
	CNN	0.5500	0.5818	0.7091	0.5909	0.6182	0.5364	0.8000	0.6818	0.6182	0.6000	0.5200	0.5900	0.5600	0.5300	0.5200	0.6004
	FBTRCA	0.5500	0.6273	0.7727	0.6000	0.6182	0.5000	0.7818	0.6455	0.6636	0.6700	0.5100	0.5400	0.5300	0.5300	0.5300	0.6046
8	STRCA	0.5083	0.6455	0.7636	0.6364	0.6545	0.4818	0.7364	0.5364	0.5636	0.5400	0.5100	0.5800	0.5300	0.5200	0.4100	0.5744
	CNN	0.5083	0.6364	0.7818	0.6545	0.7000	0.5273	0.7000	0.6091	0.6000	0.4900	0.5400	0.4700	0.5500	0.4600	0.4400	0.5778
	FBTRCA	0.5250	0.6091	0.7818	0.6818	0.7000	0.5182	0.7727	0.6364	0.6818	0.5200	0.5200	0.5200	0.5600	0.4800	0.4500	0.5971
9	STRCA	0.5583	0.6545	0.7455	0.7000	0.5727	0.6000	0.6545	0.6182	0.6364	0.5000	0.4800	0.6500	0.4900	0.5800	0.4800	0.5947
	CNN	0.5000	0.6091	0.7273	0.6455	0.5545	0.5727	0.6909	0.6455	0.5455	0.4600	0.4900	0.5900	0.5900	0.7200	0.5200	0.5907
	FBTRCA	0.5667	0.6455	0.7545	0.7000	0.5727	0.5727	0.7091	0.6818	0.6182	0.5000	0.4600	0.6300	0.5100	0.6000	0.5000	0.6014

(Continued.)

Table 4. (Continued.)

Subject	Method	Motion pair												Average			
		EF-EE	EF-SU	EF-PR	EF-HC	EF-HO	EE-SU	EE-PR	EE-HC	EE-HO	SU-PR	SU-HC	SU-HO		PR-HC	PR-HO	HC-HO
10	STRCA	0.4417	0.4273	0.5833	0.6909	0.4818	0.5091	0.6750	0.6636	0.5727	0.6273	0.5900	0.5300	0.5455	0.4909	0.5400	0.5579
	CNN	0.4250	0.4364	0.5750	0.7273	0.4545	0.5273	0.6333	0.7727	0.6364	0.6818	0.6400	0.6100	0.5273	0.4727	0.6400	0.5840
	FBTRCA	0.4583	0.4818	0.6083	0.7182	0.4455	0.4818	0.6750	0.7636	0.5545	0.6364	0.6800	0.5200	0.5364	0.4909	0.5900	0.5761
11	STRCA	0.4400	0.6500	0.7333	0.5400	0.4900	0.5400	0.5444	0.4400	0.5600	0.4778	0.4800	0.6100	0.5222	0.5444	0.5100	0.5388
	CNN	0.5700	0.4800	0.8222	0.5700	0.6000	0.5300	0.7111	0.5900	0.5700	0.5556	0.5600	0.6000	0.5667	0.5778	0.5300	0.5889
	FBTRCA	0.4900	0.5800	0.8000	0.5900	0.4800	0.5200	0.7000	0.5200	0.5200	0.5444	0.5400	0.6000	0.5333	0.5667	0.5000	0.5656
12	STRCA	0.5455	0.5600	0.6900	0.7091	0.6091	0.5333	0.6889	0.6400	0.6100	0.5625	0.5111	0.6111	0.5111	0.4778	0.4600	0.5813
	CNN	0.5727	0.6500	0.8000	0.7182	0.6818	0.5222	0.6889	0.6000	0.7000	0.5875	0.4333	0.6333	0.4778	0.5667	0.4500	0.6055
	FBTRCA	0.4909	0.5900	0.7900	0.7091	0.7273	0.5444	0.7444	0.6900	0.6800	0.6500	0.5111	0.6333	0.6444	0.5333	0.4800	0.6279
13	STRCA	0.5083	0.6000	0.6909	0.5545	0.5000	0.4818	0.6909	0.6273	0.5818	0.4600	0.5200	0.4700	0.5300	0.6600	0.5100	0.5590
	CNN	0.4250	0.5455	0.7636	0.6909	0.5273	0.4818	0.7545	0.7091	0.5000	0.4900	0.5300	0.4800	0.5900	0.5500	0.5200	0.5705
	FBTRCA	0.3917	0.5727	0.7636	0.6818	0.5364	0.6091	0.7182	0.6636	0.5364	0.5300	0.5700	0.4900	0.5400	0.6100	0.4900	0.5802
14	STRCA	0.4833	0.5900	0.7700	0.6455	0.5455	0.5300	0.8200	0.6636	0.6091	0.6125	0.6000	0.5222	0.6444	0.6000	0.4400	0.6051
	CNN	0.5583	0.6200	0.8600	0.7000	0.6273	0.5200	0.7800	0.6636	0.6182	0.6500	0.6000	0.4889	0.6444	0.6667	0.5200	0.6345
	FBTRCA	0.5417	0.6200	0.8500	0.7818	0.6273	0.5000	0.7700	0.6364	0.5636	0.7000	0.6556	0.5333	0.5889	0.6556	0.4100	0.6289
15	STRCA	0.4455	0.5556	0.7400	0.6200	0.6900	0.6900	0.7364	0.6364	0.7364	0.5778	0.5667	0.5000	0.6200	0.4900	0.5600	0.6110
	CNN	0.4909	0.5667	0.6800	0.6200	0.6800	0.7000	0.7182	0.6273	0.7636	0.6111	0.5667	0.5222	0.5700	0.5600	0.5700	0.6164
	FBTRCA	0.4455	0.5889	0.7000	0.6500	0.7100	0.7200	0.7727	0.6909	0.7273	0.5889	0.6000	0.5333	0.5800	0.4700	0.6200	0.6265
Average	STRCA	0.5031	0.5833	0.7400	0.6505	0.5899	0.5407	0.7152	0.6251	0.6141	0.6054	0.5648	0.5626	0.5712	0.5642	0.5249	0.5970
	CNN	0.5074	0.5784	0.7668	0.6593	0.6339	0.5600	0.7390	0.6638	0.6300	0.6002	0.5617	0.5685	0.5789	0.5728	0.5401	0.6112
	FBTRCA	0.5059	0.5884	0.7872	0.6827	0.6340	0.5730	0.7639	0.6649	0.6411	0.6233	0.5927	0.5757	0.5845	0.5718	0.5345	0.6216

Table 5. Deviation of classification accuracies of dataset I (RP) across ten-folds (Movement vs Movement). The best case is highlighted in bold.

Subject	Method	Motion pair														Average	
		EF-EE	EF-SU	EF-PR	EF-HC	EF-HO	EE-SU	EE-PR	EE-HC	EE-HO	SU-PR	SU-HC	SU-HO	PR-HC	PR-HO		HC-HO
1	STRCA	0.1610	0.1700	0.0614	0.1134	0.1973	0.0966	0.1088	0.1548	0.1385	0.1673	0.1330	0.1366	0.1578	0.1174	0.1874	0.1401
	CNN	0.1687	0.1174	0.0643	0.2102	0.0939	0.1434	0.0717	0.1717	0.1629	0.1386	0.1304	0.1171	0.2011	0.1418	0.1549	0.1392
	FBTRCA	0.1427	0.1229	0.0862	0.1472	0.1592	0.1054	0.1032	0.1728	0.1548	0.1677	0.1386	0.0944	0.2261	0.1252	0.2044	0.1434
2	STRCA	0.2111	0.1160	0.2014	0.1650	0.1767	0.1075	0.1958	0.1969	0.2098	0.1476	0.2111	0.1350	0.0949	0.1418	0.1633	0.1649
	CNN	0.1418	0.2068	0.1197	0.1874	0.1700	0.1581	0.0667	0.1955	0.1969	0.1506	0.2132	0.0789	0.1729	0.1663	0.1687	0.1596
	FBTRCA	0.1636	0.1549	0.1075	0.2132	0.1874	0.1780	0.1160	0.1687	0.1549	0.0632	0.2635	0.1197	0.1792	0.1252	0.1563	0.1568
3	STRCA	0.1511	0.0675	0.1054	0.0614	0.1287	0.0876	0.1337	0.0939	0.1398	0.2267	0.0973	0.1811	0.1640	0.1419	0.1330	0.1275
	CNN	0.1441	0.1174	0.1449	0.1032	0.1033	0.1269	0.1247	0.0862	0.1317	0.1972	0.1907	0.1587	0.1405	0.1376	0.1693	0.1384
	FBTRCA	0.1375	0.1354	0.1059	0.0636	0.0516	0.1524	0.0994	0.0904	0.1174	0.1208	0.2369	0.1581	0.1261	0.1179	0.1054	0.1213
4	STRCA	0.2153	0.1150	0.0939	0.1182	0.1289	0.1303	0.1118	0.1245	0.1289	0.1713	0.1150	0.1354	0.1606	0.1287	0.0857	0.1309
	CNN	0.1024	0.1067	0.1154	0.1195	0.1212	0.1289	0.0883	0.1364	0.0904	0.1317	0.1828	0.1101	0.1359	0.1197	0.1050	0.1196
	FBTRCA	0.1427	0.0614	0.0796	0.1245	0.1359	0.1342	0.0977	0.1405	0.1150	0.1506	0.1385	0.1494	0.1828	0.1317	0.1197	0.1269
5	STRCA	0.1450	0.1118	0.1216	0.0748	0.1488	0.1333	0.1080	0.1476	0.1563	0.1713	0.1269	0.1333	0.1059	0.1337	0.1581	0.1318
	CNN	0.1512	0.1289	0.1197	0.1197	0.1227	0.1350	0.1663	0.1252	0.2058	0.1430	0.2058	0.1160	0.1247	0.1155	0.1703	0.1433
	FBTRCA	0.1150	0.1216	0.1615	0.1485	0.1437	0.1179	0.1317	0.1265	0.1932	0.1337	0.1333	0.0949	0.1197	0.1229	0.1874	0.1368
6	STRCA	0.1470	0.1592	0.0777	0.1606	0.1485	0.1372	0.1054	0.1369	0.0717	0.1748	0.1252	0.1619	0.1565	0.1419	0.1764	0.1387
	CNN	0.1097	0.1649	0.1021	0.1300	0.1434	0.1684	0.0820	0.1533	0.1197	0.1581	0.1687	0.1075	0.1959	0.1768	0.1398	0.1414
	FBTRCA	0.1544	0.1548	0.0750	0.1359	0.1385	0.0796	0.1366	0.1757	0.1286	0.1350	0.1430	0.1619	0.1467	0.1972	0.1197	0.1388
7	STRCA	0.0946	0.0878	0.1197	0.1118	0.1437	0.1964	0.1512	0.1286	0.1216	0.1853	0.1838	0.1567	0.1252	0.1647	0.1197	0.1394
	CNN	0.1192	0.1067	0.1592	0.1674	0.0835	0.1738	0.1342	0.1500	0.0939	0.1247	0.0789	0.2025	0.0966	0.1059	0.1989	0.1330
	FBTRCA	0.0896	0.0796	0.1500	0.1227	0.1032	0.1372	0.1227	0.1629	0.1138	0.1252	0.1792	0.1075	0.0949	0.1337	0.1494	0.1248
8	STRCA	0.1641	0.2204	0.1671	0.1286	0.1342	0.0862	0.1985	0.1385	0.1533	0.2011	0.1101	0.1814	0.1567	0.1398	0.1524	0.1555
	CNN	0.1732	0.1545	0.1878	0.1197	0.1054	0.1533	0.1966	0.1425	0.1434	0.1969	0.1265	0.1829	0.1716	0.1265	0.1430	0.1549
	FBTRCA	0.1669	0.1138	0.1777	0.1372	0.0748	0.1606	0.1780	0.1355	0.1372	0.1619	0.1317	0.2098	0.1265	0.1135	0.1080	0.1422
9	STRCA	0.1364	0.1907	0.0835	0.1138	0.1606	0.0878	0.1592	0.1271	0.1355	0.1886	0.1687	0.1434	0.1449	0.1751	0.1619	0.1451
	CNN	0.0962	0.1359	0.1545	0.1450	0.1246	0.1717	0.1150	0.1985	0.2227	0.1647	0.2234	0.1663	0.1449	0.1687	0.0789	0.1541
	FBTRCA	0.1349	0.1684	0.1663	0.1425	0.1289	0.1289	0.0939	0.1728	0.1472	0.1054	0.2413	0.1337	0.2183	0.1054	0.1247	0.1475

(Continued.)

Table 5. (Continued.)

Subject	Method	Motion pair										Average					
		EF-EE	EF-SU	EF-PR	EF-HC	EF-HO	EE-SU	EE-PR	EE-HC	EE-HO	SU-PR		SU-HC	SU-HO	PR-HC	PR-HO	HC-HO
10	STRCA	0.0686	0.1359	0.1361	0.1557	0.1289	0.1497	0.1270	0.1289	0.1054	0.1985	0.1792	0.1567	0.1603	0.1300	0.1265	0.1392
	CNN	0.1386	0.1197	0.1270	0.1286	0.1050	0.1592	0.1721	0.1071	0.1917	0.1780	0.1578	0.1524	0.1858	0.1649	0.1265	0.1476
	FBTRCA	0.1128	0.1359	0.1245	0.1170	0.0904	0.1054	0.1493	0.1150	0.1684	0.1818	0.1619	0.1932	0.1512	0.1369	0.0994	0.1362
11	STRCA	0.1578	0.0707	0.1405	0.1265	0.1524	0.1174	0.1330	0.1506	0.2271	0.1576	0.1549	0.2183	0.1576	0.1848	0.1197	0.1513
	CNN	0.1059	0.1619	0.1589	0.1889	0.1563	0.1337	0.1673	0.1663	0.1418	0.2160	0.2066	0.1563	0.1105	0.1721	0.1418	0.1590
	FBTRCA	0.1524	0.1751	0.1721	0.1524	0.1549	0.1549	0.1177	0.1687	0.1687	0.1223	0.1506	0.1333	0.1021	0.1848	0.1764	0.1524
12	STRCA	0.2010	0.1430	0.1853	0.0939	0.1359	0.1463	0.2271	0.2366	0.1370	0.1693	0.1073	0.0944	0.2042	0.1288	0.1430	0.1569
	CNN	0.1821	0.1650	0.1247	0.1385	0.1500	0.1054	0.1874	0.1247	0.0667	0.1324	0.1430	0.1576	0.0915	0.1772	0.0707	0.1345
	FBTRCA	0.2151	0.2378	0.1449	0.1533	0.1714	0.0973	0.1660	0.1449	0.1687	0.1936	0.1405	0.1391	0.0876	0.1721	0.1549	0.1592
13	STRCA	0.1776	0.1926	0.1150	0.1385	0.1071	0.1821	0.0878	0.1684	0.1067	0.1713	0.1619	0.1337	0.1160	0.1647	0.1101	0.1422
	CNN	0.1329	0.1714	0.2019	0.1227	0.1197	0.2187	0.1054	0.1118	0.1831	0.1969	0.1567	0.1398	0.1729	0.1650	0.1619	0.1574
	FBTRCA	0.1419	0.1870	0.1497	0.1618	0.1572	0.1488	0.1088	0.1425	0.1000	0.2214	0.1059	0.1449	0.2119	0.1370	0.1101	0.1486
14	STRCA	0.0657	0.1197	0.0483	0.1246	0.1660	0.1494	0.1033	0.1054	0.2012	0.1713	0.1829	0.1486	0.2082	0.1673	0.1713	0.1422
	CNN	0.0883	0.1687	0.1075	0.0963	0.1385	0.1398	0.1033	0.1870	0.1408	0.2189	0.1194	0.1903	0.1463	0.1048	0.1989	0.1433
	FBTRCA	0.0708	0.1229	0.0850	0.1067	0.1450	0.1155	0.1160	0.1212	0.2134	0.1687	0.0973	0.1148	0.1963	0.0973	0.1524	0.1282
15	STRCA	0.1000	0.1656	0.1075	0.1229	0.1197	0.1524	0.1512	0.1660	0.0796	0.1640	0.1105	0.1080	0.1476	0.1449	0.1174	0.1305
	CNN	0.1150	0.1430	0.1398	0.1476	0.1135	0.0816	0.1088	0.1738	0.1227	0.1309	0.1523	0.1054	0.1252	0.1838	0.1337	0.1318
	FBTRCA	0.1246	0.1660	0.1333	0.1434	0.1853	0.1476	0.1071	0.1434	0.1355	0.2163	0.1829	0.1021	0.1476	0.1636	0.1549	0.1502
Average	STRCA	0.1464	0.1377	0.1176	0.1206	0.1452	0.1307	0.1401	0.1470	0.1408	0.1777	0.1445	0.1483	0.1507	0.1470	0.1417	0.1424
	CNN	0.1313	0.1446	0.1352	0.1416	0.1234	0.1465	0.1260	0.1487	0.1476	0.1652	0.1637	0.1428	0.1477	0.1484	0.1442	0.1438
	FBTRCA	0.1377	0.1425	0.1280	0.1380	0.1352	0.1309	0.1229	0.1454	0.1478	0.1512	0.1630	0.1371	0.1545	0.1376	0.1415	0.1409

Table 6. Mean of classification accuracies of dataset I (RP) across ten-folds (Movement vs Resting). The best case is highlighted in bold.

Subject	Method	Motion pair						Average
		EF-RE	EE-RE	SU-RE	PR-RE	HC-RE	HO-RE	
1	STRCA	0.8909	0.8727	0.8778	0.8300	0.8800	0.8600	0.8686
	CNN	0.9364	0.8818	0.9000	0.8800	0.8700	0.9000	0.8947
	FBTRCA	0.9364	0.8909	0.9444	0.9100	0.9200	0.9100	0.9186
2	STRCA	0.8600	0.8600	0.8100	0.7900	0.8100	0.8100	0.8233
	CNN	0.8200	0.8900	0.8600	0.8400	0.7900	0.8100	0.8350
	FBTRCA	0.9100	0.9100	0.8800	0.8300	0.8500	0.8800	0.8767
3	STRCA	0.8833	0.9000	0.8600	0.8700	0.8091	0.8600	0.8637
	CNN	0.8833	0.9583	0.8900	0.9000	0.8545	0.8600	0.8910
	FBTRCA	0.9250	0.9667	0.9000	0.8800	0.8545	0.9200	0.9077
4	STRCA	0.9000	0.9364	0.9000	0.9100	0.9000	0.8700	0.9027
	CNN	0.9182	0.9455	0.9100	0.8500	0.9091	0.8400	0.8955
	FBTRCA	0.9273	0.9545	0.9200	0.9000	0.9091	0.8700	0.9135
5	STRCA	0.7091	0.8000	0.7500	0.7900	0.7500	0.7600	0.7598
	CNN	0.8273	0.8800	0.8100	0.7900	0.8000	0.7800	0.8145
	FBTRCA	0.8091	0.8900	0.8100	0.7900	0.8200	0.7700	0.8148
6	STRCA	0.9667	0.9417	0.9182	0.9222	0.9182	0.8727	0.9233
	CNN	0.9583	0.9583	0.9000	0.9000	0.9364	0.8727	0.9210
	FBTRCA	0.9667	0.9500	0.9091	0.9222	0.9182	0.9000	0.9277
7	STRCA	0.8083	0.7500	0.7455	0.6909	0.6636	0.6727	0.7218
	CNN	0.7583	0.7833	0.7182	0.7000	0.6182	0.6545	0.7054
	FBTRCA	0.8000	0.7333	0.7455	0.7182	0.6364	0.7545	0.7313
8	STRCA	0.8667	0.8833	0.8273	0.8182	0.8545	0.8727	0.8538
	CNN	0.9417	0.8667	0.8091	0.8545	0.9000	0.8818	0.8756
	FBTRCA	0.9583	0.9000	0.8727	0.8636	0.8909	0.8909	0.8961
9	STRCA	0.8417	0.7917	0.7364	0.7636	0.6727	0.7455	0.7586
	CNN	0.8667	0.8583	0.7273	0.8182	0.8636	0.8455	0.8299
	FBTRCA	0.8917	0.8417	0.7727	0.8455	0.8273	0.8091	0.8313
10	STRCA	0.9636	0.9182	0.8600	0.7455	0.8400	0.8800	0.8679
	CNN	0.9091	0.8909	0.8600	0.8455	0.8800	0.8400	0.8709
	FBTRCA	0.9455	0.9273	0.8300	0.8545	0.8300	0.8400	0.8712
11	STRCA	0.7900	0.7600	0.7200	0.6333	0.8100	0.7100	0.7372
	CNN	0.8300	0.7800	0.7400	0.7444	0.8400	0.7700	0.7841
	FBTRCA	0.8700	0.8100	0.7800	0.8000	0.8300	0.8100	0.8167
12	STRCA	0.8455	0.7900	0.8111	0.7222	0.7500	0.7400	0.7765
	CNN	0.8273	0.7900	0.8333	0.8000	0.8100	0.7900	0.8084
	FBTRCA	0.9000	0.8200	0.8556	0.7778	0.8400	0.8700	0.8439
13	STRCA	0.9000	0.8917	0.7818	0.7636	0.7909	0.8000	0.8213
	CNN	0.9000	0.8917	0.8364	0.8455	0.8364	0.8273	0.8562
	FBTRCA	0.9500	0.9083	0.8273	0.8909	0.8273	0.8455	0.8749
14	STRCA	0.8667	0.9083	0.8600	0.7500	0.8909	0.8182	0.8490
	CNN	0.9083	0.9500	0.8500	0.7600	0.9091	0.8727	0.8750
	FBTRCA	0.8750	0.9333	0.8200	0.8000	0.9182	0.8636	0.8684
15	STRCA	0.9182	0.9167	0.8800	0.9091	0.9182	0.8727	0.9025
	CNN	0.8909	0.9250	0.8900	0.9273	0.8909	0.8455	0.8949
	FBTRCA	0.9455	0.9333	0.9000	0.8909	0.8909	0.8727	0.9056
Average	STRCA	0.8674	0.8614	0.8225	0.7939	0.8172	0.8096	0.8287
	CNN	0.8784	0.8833	0.8356	0.8304	0.8472	0.8260	0.8501
	FBTRCA	0.9074	0.8913	0.8512	0.8449	0.8508	0.8538	0.8666

We also present the p -values of two-side t -test between the STRCA method and the FBTRCA method (tables 16–18). The p -values are calculated from the accuracies of ten-folds, for each subject and each motion pair. In the three tables, the p -values are highlighted if the value is smaller than 0.1. In the dataset I (RP+MMP section), there is at least one subject

that FBTRCA shows significant improvement than STRCA ($p < 0.1$), except for the classification between *pronation* and *hand close* or between *pronation* and *hand open*.

In summary, the proposed FBTRCA method incorporates the filter bank technique into the basic STRCA method, thus presenting a

Table 7. Deviation of classification accuracies of dataset I (RP) across ten-folds (Movement vs Resting). The best case is highlighted in bold.

Subject	Method	Motion pair						Average
		EF-RE	EE-RE	SU-RE	PR-RE	HC-RE	HO-RE	
1	STRCA	0.1032	0.0977	0.1430	0.1567	0.1229	0.1350	0.1264
	CNN	0.0963	0.0963	0.0820	0.1033	0.0949	0.0943	0.0945
	FBTRCA	0.0748	0.0575	0.0586	0.1101	0.0919	0.0876	0.0801
2	STRCA	0.1265	0.0843	0.1197	0.0994	0.1197	0.0994	0.1082
	CNN	0.1398	0.0738	0.0699	0.1174	0.1370	0.0994	0.1062
	FBTRCA	0.1101	0.0738	0.0632	0.1059	0.0972	0.1398	0.0983
3	STRCA	0.1192	0.0351	0.0699	0.1059	0.0904	0.1430	0.0939
	CNN	0.0805	0.0439	0.1197	0.1155	0.1300	0.0843	0.0957
	FBTRCA	0.0829	0.0583	0.0816	0.1549	0.0977	0.1033	0.0965
4	STRCA	0.0796	0.0963	0.0816	0.0876	0.0671	0.1160	0.0880
	CNN	0.0796	0.0636	0.0738	0.1179	0.0857	0.2011	0.1036
	FBTRCA	0.0835	0.0883	0.0919	0.1054	0.1134	0.2111	0.1156
5	STRCA	0.1703	0.1563	0.1509	0.1287	0.1650	0.1647	0.1560
	CNN	0.0796	0.0789	0.1101	0.1370	0.1155	0.1135	0.1058
	FBTRCA	0.1246	0.1287	0.1449	0.1287	0.0919	0.1160	0.1224
6	STRCA	0.0583	0.0562	0.0904	0.0915	0.1088	0.1067	0.0853
	CNN	0.0439	0.0439	0.1000	0.0820	0.0748	0.0977	0.0737
	FBTRCA	0.0583	0.0430	0.0742	0.1177	0.1170	0.1000	0.0850
7	STRCA	0.1115	0.0786	0.0575	0.0878	0.1138	0.1150	0.0940
	CNN	0.1270	0.1054	0.1088	0.0614	0.1342	0.1472	0.1140
	FBTRCA	0.0896	0.0861	0.0939	0.1088	0.1286	0.1289	0.1060
8	STRCA	0.0703	0.0896	0.0796	0.1485	0.0767	0.0977	0.0937
	CNN	0.1043	0.1125	0.1000	0.1150	0.0904	0.0862	0.1014
	FBTRCA	0.0810	0.0766	0.1150	0.0883	0.1118	0.1032	0.0960
9	STRCA	0.1142	0.0708	0.1938	0.2019	0.1227	0.1118	0.1359
	CNN	0.0583	0.1669	0.1868	0.0857	0.0982	0.1138	0.1183
	FBTRCA	0.0883	0.1329	0.1154	0.1359	0.1385	0.0904	0.1169
10	STRCA	0.0636	0.0671	0.0699	0.1197	0.1430	0.1033	0.0944
	CNN	0.0857	0.0835	0.0966	0.1425	0.1135	0.1430	0.1108
	FBTRCA	0.0636	0.0717	0.0949	0.1227	0.0675	0.1174	0.0896
11	STRCA	0.1197	0.1174	0.1398	0.1391	0.1370	0.1370	0.1317
	CNN	0.0675	0.1549	0.1350	0.1660	0.0966	0.1252	0.1242
	FBTRCA	0.0949	0.1197	0.1033	0.1640	0.1252	0.1370	0.1240
12	STRCA	0.0963	0.1197	0.1288	0.1309	0.1179	0.2119	0.1343
	CNN	0.1000	0.1524	0.1410	0.0876	0.1524	0.2025	0.1393
	FBTRCA	0.0671	0.1229	0.1576	0.1171	0.1075	0.1494	0.1203
13	STRCA	0.0861	0.0791	0.1300	0.1300	0.1663	0.0575	0.1081
	CNN	0.1024	0.0966	0.1342	0.1289	0.1118	0.0904	0.1107
	FBTRCA	0.0805	0.0998	0.1790	0.0835	0.0904	0.0963	0.1049
14	STRCA	0.0896	0.0829	0.1174	0.1080	0.1197	0.1212	0.1065
	CNN	0.0615	0.0583	0.1269	0.1075	0.1050	0.0636	0.0871
	FBTRCA	0.0900	0.0861	0.1398	0.0943	0.1000	0.1071	0.1029
15	STRCA	0.0796	0.0878	0.1317	0.0857	0.0671	0.1150	0.0945
	CNN	0.0939	0.0615	0.0994	0.0717	0.1032	0.0963	0.0877
	FBTRCA	0.0767	0.0861	0.1054	0.1032	0.1032	0.1150	0.0983
Average	STRCA	0.0992	0.0879	0.1136	0.1214	0.1159	0.1223	0.1101
	CNN	0.0880	0.0928	0.1123	0.1093	0.1095	0.1172	0.1049
	FBTRCA	0.0844	0.0888	0.1079	0.1160	0.1054	0.1202	0.1038

comparable classification performance to STRCA in pre-movement decoding.

4. Discussion

In this study, we proposed a new pre-movement decoding method, FBTRCA. This method is

developed by incorporating filter bank selection on the STRCA method. In comparison to MI signals, MRCP signals have the advantage that the movement patterns can be observed before the movement onset. FBTRCA is designed to extract features from the grand average MRCP, which contains information about the movement in MRCP signals. FBTRCA

Table 8. Mean of classification accuracies of dataset I (RP+MMP) across ten-folds (Movement vs. Movement). The best case is highlighted in bold.

Subject	Method	Motion pair														Average	
		EF-EE	EF-SU	EF-PR	EF-HC	EF-HO	EE-SU	EE-PR	EE-HC	EE-HO	SU-PR	SU-HC	SU-HO	PR-HC	PR-HO		HC-HO
1	STRCA	0.4917	0.6700	0.8273	0.7182	0.5909	0.5800	0.7273	0.7273	0.6000	0.8111	0.7111	0.6333	0.6000	0.6200	0.6600	0.6645
	CNN	0.5417	0.6900	0.9091	0.7636	0.7364	0.6400	0.9000	0.7909	0.7364	0.9111	0.7778	0.7556	0.5800	0.6500	0.6300	0.7342
	FBTRCA	0.5500	0.7100	0.9273	0.8091	0.7273	0.6900	0.9182	0.8364	0.7545	0.9222	0.8000	0.7889	0.5900	0.6900	0.6500	0.7576
2	STRCA	0.5300	0.5700	0.8700	0.6000	0.5200	0.5600	0.8100	0.6100	0.4900	0.7400	0.4900	0.5500	0.6400	0.6700	0.5000	0.6100
	CNN	0.5100	0.6300	0.9100	0.5700	0.5000	0.6300	0.8600	0.5300	0.6300	0.7600	0.5200	0.6100	0.7000	0.7200	0.4500	0.6353
	FBTRCA	0.5900	0.6500	0.9400	0.5700	0.5600	0.6100	0.8900	0.6100	0.6600	0.7700	0.5900	0.5900	0.7100	0.7400	0.4200	0.6600
3	STRCA	0.5083	0.7000	0.7500	0.8364	0.7600	0.5800	0.7400	0.9273	0.8400	0.5500	0.6444	0.7000	0.5333	0.6500	0.7111	0.6954
	CNN	0.6917	0.8100	0.7900	0.9455	0.9200	0.7400	0.8400	0.9455	0.9600	0.7125	0.7333	0.8000	0.6000	0.5500	0.6333	0.7781
	FBTRCA	0.7333	0.8500	0.8500	0.9273	0.9100	0.7700	0.8500	0.9545	0.9600	0.6875	0.8000	0.7875	0.5889	0.7000	0.6333	0.8002
4	STRCA	0.5583	0.6182	0.9545	0.8833	0.8000	0.5818	0.9273	0.8500	0.7818	0.9000	0.8455	0.8500	0.5909	0.5700	0.6000	0.7541
	CNN	0.5833	0.7091	0.9909	0.9417	0.8455	0.7000	0.9636	0.9167	0.8545	0.9800	0.9091	0.9400	0.6000	0.6300	0.6273	0.8128
	FBTRCA	0.6417	0.7000	0.9727	0.9500	0.8636	0.6909	0.9636	0.9250	0.8818	0.9800	0.9545	0.9500	0.6000	0.6700	0.6455	0.8260
5	STRCA	0.5818	0.5909	0.8182	0.6818	0.6545	0.7100	0.9100	0.8500	0.8000	0.5900	0.6000	0.6000	0.5700	0.5100	0.5800	0.6698
	CNN	0.6182	0.7364	0.8455	0.8091	0.7727	0.7700	0.9600	0.8800	0.7900	0.6000	0.5700	0.4900	0.6200	0.5200	0.5300	0.7008
	FBTRCA	0.5818	0.6273	0.9091	0.8000	0.7273	0.7300	0.9700	0.8400	0.8300	0.6400	0.5300	0.5000	0.5700	0.5600	0.5600	0.6917
6	STRCA	0.5000	0.6000	0.9333	0.7364	0.8364	0.5455	0.9333	0.6545	0.8091	0.6625	0.5900	0.7100	0.6750	0.6250	0.6400	0.6967
	CNN	0.6250	0.6273	0.9333	0.8182	0.8727	0.6091	0.9111	0.6636	0.8545	0.6625	0.5500	0.6900	0.6125	0.5625	0.6200	0.7075
	FBTRCA	0.6083	0.7091	0.9667	0.7727	0.8818	0.6091	0.9222	0.6000	0.8545	0.7000	0.4800	0.6700	0.6875	0.6000	0.6200	0.7121
7	STRCA	0.4917	0.6182	0.8455	0.7091	0.6909	0.5545	0.8636	0.7182	0.7091	0.6300	0.5500	0.5500	0.5900	0.5200	0.4900	0.6354
	CNN	0.5583	0.6455	0.9364	0.7455	0.8455	0.5273	0.8909	0.7364	0.7091	0.6200	0.6000	0.6100	0.6200	0.5000	0.5900	0.6756
	FBTRCA	0.5417	0.6636	0.9364	0.7636	0.6909	0.5818	0.8909	0.7364	0.7000	0.6800	0.5800	0.5700	0.6600	0.5200	0.4900	0.6670
8	STRCA	0.4583	0.6182	0.8545	0.7727	0.7000	0.5636	0.8818	0.6727	0.6455	0.7700	0.6400	0.7000	0.5000	0.5500	0.5300	0.6572
	CNN	0.4917	0.6545	0.9455	0.7455	0.7818	0.6000	0.9364	0.6727	0.7455	0.8500	0.5800	0.6300	0.5500	0.5200	0.5500	0.6836
	FBTRCA	0.4917	0.7364	0.9364	0.8000	0.7727	0.6364	0.9273	0.7273	0.7636	0.8600	0.7000	0.6800	0.5400	0.6100	0.5300	0.7141
9	STRCA	0.7083	0.7273	0.8364	0.7636	0.7545	0.6455	0.7818	0.7364	0.7273	0.6400	0.4900	0.5600	0.5700	0.6200	0.5900	0.6767
	CNN	0.8417	0.8182	0.9455	0.7818	0.8091	0.6909	0.8909	0.8091	0.8091	0.5900	0.4800	0.7300	0.5400	0.7100	0.5700	0.7344
	FBTRCA	0.8917	0.7545	0.9455	0.8364	0.8091	0.7455	0.9000	0.7727	0.8273	0.6200	0.4500	0.7000	0.5800	0.7000	0.6600	0.7462

(Continued.)

Table 8. (Continued.)

Subject	Method	Motion pair												Average			
		EF-EE	EF-SU	EF-PR	EF-HC	EF-HO	EE-SU	EE-PR	EE-HC	EE-HO	SU-PR	SU-HC	SU-HO		PR-HC	PR-HO	HC-HO
10	STRCA	0.5167	0.5364	0.7250	0.8545	0.6455	0.5545	0.6583	0.7818	0.7636	0.6545	0.6800	0.7200	0.6545	0.4636	0.5600	0.6513
	CNN	0.5083	0.6000	0.7250	0.8091	0.7636	0.5909	0.7917	0.8364	0.7636	0.6182	0.8200	0.7200	0.6000	0.5818	0.5200	0.6832
	FBTRCA	0.4833	0.6000	0.7750	0.8545	0.7364	0.6000	0.7583	0.8455	0.7636	0.6636	0.8000	0.7000	0.6455	0.5455	0.6100	0.6921
11	STRCA	0.4900	0.6600	0.6889	0.6100	0.5800	0.6700	0.7000	0.5700	0.5700	0.6444	0.6700	0.6200	0.5111	0.5778	0.4900	0.6035
	CNN	0.5100	0.6900	0.7667	0.6100	0.6600	0.5900	0.7111	0.6500	0.5800	0.6333	0.5800	0.7200	0.5000	0.6000	0.4400	0.6161
	FBTRCA	0.5300	0.7100	0.8000	0.7000	0.6700	0.6200	0.8000	0.6700	0.6400	0.6444	0.6700	0.7500	0.5222	0.5556	0.3800	0.6441
12	STRCA	0.5636	0.6100	0.7900	0.7182	0.7364	0.6333	0.7778	0.6700	0.7100	0.6250	0.5667	0.5889	0.5667	0.4444	0.5400	0.6361
	CNN	0.6545	0.5600	0.8900	0.7909	0.7818	0.7222	0.8889	0.6600	0.8100	0.6625	0.4667	0.7000	0.5333	0.4889	0.4900	0.6733
	FBTRCA	0.6636	0.6500	0.9200	0.7818	0.8000	0.6778	0.9111	0.7400	0.8400	0.8000	0.7000	0.7444	0.6222	0.5444	0.5100	0.7270
13	STRCA	0.5083	0.5909	0.7636	0.6909	0.5182	0.5545	0.8182	0.7455	0.6636	0.6100	0.6200	0.4600	0.5300	0.6400	0.5200	0.6156
	CNN	0.4500	0.5727	0.8455	0.7182	0.6636	0.6455	0.8182	0.8182	0.6273	0.5900	0.6200	0.5300	0.5700	0.5800	0.6399	
	FBTRCA	0.5250	0.6273	0.8818	0.8182	0.6636	0.6545	0.9091	0.8182	0.6818	0.6500	0.7000	0.5900	0.5200	0.5500	0.5700	0.6773
14	STRCA	0.5250	0.6100	0.8900	0.7545	0.6818	0.6400	0.8300	0.8000	0.6000	0.6750	0.5889	0.6444	0.5556	0.6556	0.4800	0.6621
	CNN	0.6000	0.6900	0.9400	0.8727	0.8000	0.6100	0.9500	0.7636	0.7455	0.7500	0.5889	0.6778	0.6667	0.6889	0.4900	0.7223
	FBTRCA	0.6250	0.7000	0.9300	0.8545	0.8727	0.6100	0.9300	0.8273	0.7727	0.8000	0.6556	0.6556	0.6000	0.7444	0.4700	0.7365
15	STRCA	0.3727	0.7000	0.7900	0.6800	0.8000	0.7300	0.8727	0.6909	0.8091	0.6778	0.6667	0.6333	0.6600	0.5800	0.6600	0.6882
	CNN	0.5182	0.6889	0.8700	0.7200	0.8500	0.8200	0.8818	0.6364	0.7636	0.6333	0.6333	0.6444	0.6200	0.4600	0.6400	0.6920
	FBTRCA	0.4455	0.7222	0.9500	0.7000	0.8600	0.7500	0.9091	0.7182	0.8545	0.7444	0.6667	0.6222	0.6600	0.5500	0.5800	0.7155
Average	STRCA	0.5203	0.6280	0.8225	0.7340	0.6846	0.6069	0.8155	0.7336	0.7013	0.6787	0.6235	0.6347	0.5831	0.5798	0.5701	0.6611
	CNN	0.5802	0.6748	0.8829	0.7761	0.7735	0.6591	0.8796	0.7540	0.7586	0.7049	0.6286	0.6832	0.5942	0.5821	0.5574	0.6993
	FBTRCA	0.5935	0.6940	0.9094	0.7959	0.7697	0.6651	0.8967	0.7748	0.7856	0.7441	0.6718	0.6866	0.6064	0.6187	0.5553	0.7178

Table 9. Deviation of classification accuracies of dataset I (RP+MMP) across ten-folds (Movement vs Movement). The best case is highlighted in bold.

Subject	Method	Motion pair																Average
		EF-EE	EF-SU	EF-PR	EF-HC	EF-HO	EE-SU	EE-PR	EE-HC	EE-HO	SU-PR	SU-HC	SU-HO	PR-HC	PR-HO	HC-HO		
1	STRCA	0.1819	0.0949	0.1629	0.1170	0.1560	0.2044	0.1355	0.1134	0.1671	0.2098	0.1500	0.1391	0.1247	0.1033	0.1713	0.1487	
	CNN	0.1483	0.0876	0.0857	0.1150	0.1572	0.1430	0.0796	0.0862	0.0904	0.0703	0.0907	0.1366	0.1317	0.1269	0.1337	0.1122	
	FBTRCA	0.0805	0.1287	0.0575	0.0904	0.1134	0.0994	0.1000	0.1271	0.0963	0.0915	0.1366	0.1105	0.1449	0.1197	0.1650	0.1108	
2	STRCA	0.1947	0.1703	0.1337	0.1633	0.1135	0.1265	0.1663	0.1101	0.2424	0.1350	0.1853	0.1434	0.1713	0.1494	0.1414	0.1564	
	CNN	0.1197	0.1252	0.0876	0.1947	0.1333	0.1160	0.1350	0.1160	0.1252	0.1265	0.1932	0.1969	0.1563	0.1229	0.2068	0.1437	
	FBTRCA	0.0876	0.1780	0.0699	0.2111	0.1350	0.1197	0.1101	0.1101	0.2221	0.2003	0.2025	0.1729	0.1370	0.0966	0.1814	0.1489	
3	STRCA	0.2058	0.1247	0.1269	0.0575	0.1350	0.1687	0.1713	0.0939	0.1075	0.1687	0.1721	0.1344	0.1021	0.1419	0.1753	0.1391	
	CNN	0.1182	0.0994	0.0738	0.0636	0.0632	0.1174	0.1075	0.0469	0.0516	0.1868	0.1589	0.1467	0.1829	0.1882	0.1486	0.1169	
	FBTRCA	0.1230	0.1434	0.1080	0.0575	0.0876	0.1252	0.1080	0.0479	0.0966	0.1792	0.1366	0.1324	0.0915	0.1467	0.1177	0.1134	
4	STRCA	0.1524	0.1032	0.0643	0.1125	0.1342	0.1828	0.0575	0.1097	0.1434	0.0816	0.1289	0.1354	0.1372	0.1889	0.1067	0.1226	
	CNN	0.0962	0.0575	0.0287	0.0562	0.0862	0.1138	0.0469	0.1039	0.1067	0.0422	0.0742	0.0516	0.1369	0.1160	0.1000	0.0811	
	FBTRCA	0.1364	0.1216	0.0439	0.0703	0.0643	0.1497	0.0636	0.0917	0.1054	0.0422	0.0773	0.0707	0.1300	0.1160	0.1170	0.0933	
5	STRCA	0.1227	0.1728	0.1286	0.1154	0.1472	0.1524	0.1101	0.1179	0.1414	0.1595	0.2404	0.1944	0.1418	0.1524	0.1033	0.1467	
	CNN	0.1954	0.1512	0.1054	0.0796	0.1500	0.1252	0.0516	0.0919	0.0738	0.1563	0.1494	0.1101	0.1317	0.1033	0.1494	0.1216	
	FBTRCA	0.1828	0.0796	0.0742	0.1032	0.1286	0.0949	0.0675	0.1075	0.0949	0.1265	0.1252	0.1944	0.1567	0.1075	0.1647	0.1205	
6	STRCA	0.1800	0.1300	0.0574	0.1938	0.1342	0.1355	0.0777	0.1592	0.1246	0.1773	0.1729	0.1370	0.1054	0.1768	0.1955	0.1438	
	CNN	0.0982	0.1890	0.0777	0.1134	0.0977	0.0862	0.1148	0.1138	0.0977	0.1672	0.1269	0.0994	0.1376	0.1693	0.1989	0.1259	
	FBTRCA	0.1182	0.1808	0.0537	0.1437	0.1216	0.1054	0.0750	0.1973	0.0878	0.1344	0.1814	0.1337	0.1215	0.2555	0.1229	0.1355	
7	STRCA	0.1687	0.1592	0.1138	0.1592	0.1150	0.1938	0.0982	0.1088	0.1271	0.2058	0.1841	0.1509	0.1370	0.1874	0.1287	0.1492	
	CNN	0.1043	0.1170	0.0748	0.1271	0.0862	0.1408	0.0835	0.1738	0.0717	0.1687	0.1414	0.1197	0.1687	0.1155	0.1663	0.1240	
	FBTRCA	0.1195	0.1138	0.0748	0.1300	0.1369	0.0878	0.0835	0.1170	0.1425	0.1476	0.1398	0.1767	0.2221	0.1476	0.1197	0.1306	
8	STRCA	0.1375	0.1032	0.0977	0.1372	0.1289	0.1118	0.0748	0.0977	0.1246	0.1160	0.2119	0.1247	0.1247	0.1581	0.1337	0.1255	
	CNN	0.1544	0.1472	0.0767	0.1342	0.1067	0.1926	0.1216	0.1369	0.0835	0.1434	0.0919	0.1252	0.1650	0.2098	0.1269	0.1344	
	FBTRCA	0.1072	0.1684	0.0748	0.1118	0.1071	0.1545	0.0939	0.1545	0.1369	0.0843	0.1155	0.1814	0.1174	0.1729	0.1418	0.1282	
9	STRCA	0.1128	0.1603	0.1408	0.1227	0.1359	0.1385	0.0878	0.1512	0.1818	0.1430	0.1370	0.1265	0.1494	0.1932	0.1853	0.1444	
	CNN	0.0917	0.0606	0.0636	0.1150	0.0904	0.1150	0.0835	0.1000	0.1088	0.2283	0.1549	0.1418	0.1265	0.1370	0.1829	0.1200	
	FBTRCA	0.0883	0.1216	0.0469	0.1271	0.1317	0.1342	0.0904	0.0883	0.0516	0.1814	0.1354	0.1563	0.1619	0.1155	0.1578	0.1192	

(Continued.)

Table 9. (Continued.)

Subject	Method	Motion pair												Average			
		EF-EE	EF-SU	EF-PR	EF-HC	EF-HO	EE-SU	EE-PR	EE-HC	EE-HO	SU-PR	SU-HC	SU-HO		PR-HC	PR-HO	HC-HO
10	STRCA	0.1459	0.1000	0.1364	0.1300	0.1170	0.1841	0.1208	0.0767	0.1369	0.1907	0.1317	0.1814	0.1408	0.1088	0.1265	0.1352
	CNN	0.1270	0.1300	0.1472	0.1738	0.1497	0.1154	0.0982	0.1118	0.0977	0.2001	0.1229	0.1619	0.1434	0.1369	0.1814	0.1398
	FBTRCA	0.1561	0.1369	0.0883	0.1067	0.1000	0.1434	0.0998	0.0963	0.1227	0.1770	0.1414	0.0943	0.1246	0.1714	0.1595	0.1279
11	STRCA	0.2283	0.1506	0.1463	0.2079	0.1033	0.2452	0.1576	0.2058	0.1494	0.2082	0.1636	0.1476	0.1194	0.1148	0.1595	0.1672
	CNN	0.1969	0.0876	0.1223	0.1197	0.1578	0.1287	0.1304	0.1900	0.1476	0.1892	0.1751	0.1317	0.1757	0.1673	0.2171	0.1558
	FBTRCA	0.1636	0.1197	0.1148	0.1563	0.1494	0.1476	0.1366	0.2003	0.1578	0.1366	0.1829	0.2369	0.0915	0.1656	0.1033	0.1509
12	STRCA	0.1032	0.1969	0.1101	0.1684	0.0904	0.1288	0.1571	0.1494	0.1449	0.1318	0.2124	0.1486	0.1772	0.1889	0.2011	0.1540
	CNN	0.1271	0.1713	0.0994	0.1717	0.0878	0.1907	0.1283	0.1647	0.1663	0.1672	0.1366	0.0915	0.1640	0.1073	0.1370	0.1407
	FBTRCA	0.0963	0.2369	0.0632	0.1434	0.1032	0.2250	0.1148	0.1430	0.1350	0.0874	0.1288	0.0915	0.1304	0.1610	0.2234	0.1389
13	STRCA	0.1329	0.1928	0.1369	0.1227	0.1216	0.1317	0.0958	0.1118	0.1821	0.2234	0.1135	0.1838	0.1494	0.1506	0.1687	0.1478
	CNN	0.0805	0.1870	0.0862	0.1450	0.0963	0.1317	0.0857	0.1355	0.0904	0.1287	0.1476	0.2312	0.1337	0.1581	0.1751	0.1342
	FBTRCA	0.1574	0.1890	0.0862	0.0606	0.1216	0.1472	0.0857	0.1355	0.1071	0.2273	0.1155	0.1729	0.1317	0.1900	0.1418	0.1380
14	STRCA	0.1115	0.1197	0.0876	0.1606	0.1231	0.1578	0.0823	0.1032	0.1434	0.1787	0.1818	0.0703	0.2095	0.1223	0.1549	0.1338
	CNN	0.1165	0.1595	0.0699	0.0767	0.1032	0.1197	0.0527	0.1150	0.1197	0.1768	0.1576	0.0973	0.1656	0.1148	0.1449	0.1193
	FBTRCA	0.1318	0.1054	0.0675	0.1067	0.1150	0.1101	0.0675	0.0796	0.0982	0.1581	0.1610	0.1223	0.1829	0.1177	0.2111	0.1223
15	STRCA	0.1385	0.1818	0.1370	0.1229	0.1054	0.1252	0.1150	0.1067	0.1450	0.1610	0.0907	0.1391	0.1430	0.1398	0.1647	0.1344
	CNN	0.1488	0.2082	0.1160	0.0919	0.0972	0.1317	0.0748	0.1134	0.0977	0.1288	0.1177	0.1554	0.1874	0.2366	0.1506	0.1371
	FBTRCA	0.1938	0.2045	0.0707	0.1054	0.0966	0.1509	0.0742	0.1512	0.1150	0.1391	0.1481	0.1405	0.1075	0.1900	0.1033	0.1327
Average	STRCA	0.1545	0.1440	0.1187	0.1394	0.1240	0.1591	0.1139	0.1210	0.1508	0.1660	0.1651	0.1438	0.1422	0.1518	0.1544	0.1432
	CNN	0.1282	0.1319	0.0877	0.1185	0.1109	0.1312	0.0930	0.1200	0.1019	0.1520	0.1359	0.1331	0.1538	0.1473	0.1613	0.1271
	FBTRCA	0.1295	0.1485	0.0730	0.1150	0.1141	0.1330	0.0914	0.1232	0.1180	0.1408	0.1419	0.1458	0.1368	0.1516	0.1487	0.1274

Table 10. Mean of classification accuracies of dataset I (RP+MMP) across ten-folds (Movement vs Resting). The best case is highlighted in bold.

Subject	Method	Motion pair						Average
		EF-RE	EE-RE	SU-RE	PR-RE	HC-RE	HO-RE	
1	STRCA	0.8636	0.8545	0.8444	0.7300	0.7700	0.8000	0.8104
	CNN	0.9545	0.9000	0.9333	0.8700	0.8400	0.9200	0.9030
	FBTRCA	0.9727	0.9273	0.9444	0.8700	0.8500	0.9200	0.9141
2	STRCA	0.9300	0.9100	0.8600	0.7700	0.8400	0.8200	0.8550
	CNN	0.9500	0.9100	0.8800	0.8300	0.8000	0.7900	0.8600
	FBTRCA	0.9900	0.9400	0.9000	0.8500	0.8800	0.8700	0.9050
3	STRCA	0.9333	0.9417	0.8000	0.7600	0.7545	0.8200	0.8349
	CNN	0.9583	1.0000	0.9000	0.8900	0.8364	0.8700	0.9091
	FBTRCA	0.9583	1.0000	0.9000	0.8900	0.8091	0.9000	0.9096
4	STRCA	0.9545	0.9273	0.9400	0.7800	0.8000	0.8100	0.8686
	CNN	0.9727	0.9636	1.0000	0.8800	0.8818	0.8600	0.9264
	FBTRCA	0.9909	0.9636	1.0000	0.8700	0.8909	0.8800	0.9326
5	STRCA	0.8636	0.9400	0.7500	0.7000	0.7400	0.7400	0.7889
	CNN	0.9364	0.9600	0.8200	0.7800	0.8300	0.8100	0.8561
	FBTRCA	0.9364	0.9800	0.8100	0.7900	0.8100	0.8100	0.8561
6	STRCA	0.9833	0.9833	0.8273	0.9000	0.8455	0.8273	0.8944
	CNN	1.0000	1.0000	0.9455	0.9111	0.9364	0.8545	0.9412
	FBTRCA	0.9917	1.0000	0.9455	0.8889	0.9273	0.8818	0.9392
7	STRCA	0.8417	0.8750	0.7909	0.6545	0.6636	0.6364	0.7437
	CNN	0.9417	0.8583	0.7636	0.7091	0.6818	0.6636	0.7697
	FBTRCA	0.9417	0.8917	0.8455	0.7182	0.6818	0.7091	0.7980
8	STRCA	0.9750	0.9667	0.9364	0.8091	0.8545	0.9000	0.9069
	CNN	1.0000	0.9833	0.9000	0.8545	0.9273	0.9545	0.9366
	FBTRCA	0.9917	0.9833	0.9455	0.8545	0.9182	0.9455	0.9398
9	STRCA	0.9000	0.8833	0.6909	0.7455	0.6364	0.7182	0.7624
	CNN	0.9750	0.9750	0.8455	0.8000	0.8091	0.8636	0.8780
	FBTRCA	0.9833	0.9750	0.8455	0.8273	0.8455	0.8364	0.8855
10	STRCA	0.9000	0.9364	0.8100	0.7636	0.7100	0.8300	0.8250
	CNN	0.9273	0.9364	0.9000	0.8000	0.8100	0.8700	0.8739
	FBTRCA	0.9364	0.9636	0.9200	0.8636	0.8500	0.8900	0.9039
11	STRCA	0.8200	0.8400	0.8300	0.6889	0.7300	0.6500	0.7598
	CNN	0.8900	0.8800	0.8800	0.7778	0.8500	0.7800	0.8430
	FBTRCA	0.9300	0.9300	0.8600	0.7556	0.8600	0.8400	0.8626
12	STRCA	0.9273	0.9300	0.8444	0.6667	0.7900	0.6700	0.8047
	CNN	0.9091	0.9000	0.9111	0.8000	0.8500	0.8000	0.8617
	FBTRCA	0.9455	0.9200	0.8889	0.7556	0.8600	0.8400	0.8683
13	STRCA	0.9250	0.9250	0.7545	0.6909	0.7545	0.7909	0.8068
	CNN	0.9583	0.9750	0.8273	0.7727	0.7636	0.8455	0.8571
	FBTRCA	0.9750	0.9667	0.8455	0.8909	0.8182	0.8727	0.8948
14	STRCA	0.9500	0.9333	0.8300	0.7000	0.8364	0.7636	0.8356
	CNN	0.9833	0.9833	0.8700	0.8100	0.9364	0.9091	0.9154
	FBTRCA	0.9833	0.9750	0.9300	0.8400	0.9000	0.8455	0.9123
15	STRCA	0.9182	0.9167	0.8200	0.8091	0.7818	0.8182	0.8440
	CNN	0.9727	0.9667	0.9200	0.9000	0.8818	0.8273	0.9114
	FBTRCA	0.9818	0.9667	0.9400	0.8818	0.9364	0.8727	0.9299
Average	STRCA	0.9124	0.9175	0.8219	0.7446	0.7672	0.7730	0.8228
	CNN	0.9553	0.9461	0.8864	0.8257	0.8423	0.8412	0.8828
	FBTRCA	0.9672	0.9589	0.9014	0.8364	0.8558	0.8609	0.8968

consists of three modules: the spatial filter TRCA, the CCPs features and the low-frequency bands. The associations between these and the grand average MRCP are detailed below. To better explain the relationship between FBTRCA and grand average MRCP, the grand average MRCP of multiple actions in channel C_z is presented in figure 11.

The spatial filter is a linear transformation of multi-channel EEG signals. It plays the role of channel selection and thus optimizes the spatial characteristic of EEG signals. Our previous work pointed out that SSVEP and MRCP signals have similar spatial distributions, and the STRCA method was proposed by comparing the effects of different spatial filters for

Table 11. Deviation of classification accuracies of dataset I (RP+MMP) across ten-folds (Movement vs Resting). The best case is highlighted in bold.

Subject	Method	Motion pair						Average
		EF-RE	EE-RE	SU-RE	PR-RE	HC-RE	HO-RE	
1	STRCA	0.0773	0.1300	0.1304	0.1252	0.0823	0.1414	0.1144
	CNN	0.0643	0.0904	0.0777	0.0949	0.0966	0.0789	0.0838
	FBTRCA	0.0614	0.0939	0.0586	0.0949	0.0972	0.1033	0.0849
2	STRCA	0.0675	0.0738	0.0966	0.1160	0.1430	0.1317	0.1047
	CNN	0.1080	0.0738	0.1135	0.0949	0.1563	0.1101	0.1094
	FBTRCA	0.0316	0.0516	0.0816	0.0707	0.1033	0.1160	0.0758
3	STRCA	0.0766	0.0686	0.1633	0.1075	0.1488	0.1135	0.1130
	CNN	0.0589	0.0000	0.0943	0.1197	0.1118	0.0949	0.0799
	FBTRCA	0.0708	0.0000	0.0943	0.1197	0.1246	0.0943	0.0839
4	STRCA	0.0479	0.0939	0.0516	0.1476	0.1342	0.1449	0.1033
	CNN	0.0439	0.0636	0.0000	0.1229	0.1054	0.1350	0.0785
	FBTRCA	0.0287	0.0767	0.0000	0.1059	0.0575	0.1874	0.0760
5	STRCA	0.1071	0.0699	0.1900	0.1633	0.1174	0.1578	0.1343
	CNN	0.1359	0.0699	0.1317	0.1135	0.0949	0.0994	0.1075
	FBTRCA	0.0963	0.0422	0.1370	0.1370	0.0994	0.1729	0.1141
6	STRCA	0.0351	0.0351	0.1684	0.0973	0.1138	0.1450	0.0991
	CNN	0.0000	0.0000	0.0767	0.0876	0.0614	0.1227	0.0581
	FBTRCA	0.0264	0.0000	0.0636	0.0907	0.0939	0.1054	0.0633
7	STRCA	0.0998	0.1128	0.1359	0.1408	0.1488	0.1545	0.1321
	CNN	0.0791	0.0791	0.1557	0.1197	0.1437	0.1359	0.1188
	FBTRCA	0.0562	0.0966	0.1606	0.1246	0.1437	0.0835	0.1109
8	STRCA	0.0403	0.0583	0.0439	0.1088	0.0878	0.1246	0.0773
	CNN	0.0000	0.0351	0.0904	0.0878	0.0575	0.0643	0.0559
	FBTRCA	0.0264	0.0351	0.0469	0.0767	0.0904	0.0767	0.0587
9	STRCA	0.1097	0.0896	0.1497	0.1533	0.1485	0.1512	0.1337
	CNN	0.0403	0.0562	0.1717	0.1858	0.1629	0.0773	0.1157
	FBTRCA	0.0351	0.0562	0.1054	0.1385	0.1289	0.1118	0.0960
10	STRCA	0.0796	0.0963	0.1595	0.1434	0.1197	0.1160	0.1191
	CNN	0.0717	0.0614	0.1054	0.1649	0.0876	0.1160	0.1011
	FBTRCA	0.0748	0.0469	0.1033	0.1372	0.0707	0.0994	0.0887
11	STRCA	0.0919	0.1350	0.1494	0.1366	0.1418	0.2273	0.1470
	CNN	0.0994	0.0919	0.1033	0.1171	0.1354	0.1135	0.1101
	FBTRCA	0.0675	0.0675	0.0699	0.1148	0.1265	0.1174	0.0939
12	STRCA	0.1032	0.0823	0.1304	0.1048	0.1370	0.1703	0.1213
	CNN	0.0606	0.0667	0.1148	0.1261	0.0527	0.2000	0.1035
	FBTRCA	0.0469	0.0632	0.1171	0.0876	0.1075	0.1776	0.1000
13	STRCA	0.0615	0.0829	0.1289	0.1300	0.1359	0.0963	0.1059
	CNN	0.0439	0.0403	0.1385	0.1303	0.0878	0.1138	0.0924
	FBTRCA	0.0403	0.0430	0.1138	0.0835	0.0958	0.1227	0.0832
14	STRCA	0.0703	0.0766	0.1636	0.1333	0.1472	0.0767	0.1113
	CNN	0.0351	0.0351	0.0949	0.0994	0.0748	0.0958	0.0725
	FBTRCA	0.0351	0.0403	0.1059	0.1075	0.0671	0.0862	0.0737
15	STRCA	0.0904	0.0556	0.1229	0.1512	0.1067	0.1134	0.1067
	CNN	0.0439	0.0430	0.1135	0.1000	0.1216	0.1088	0.0885
	FBTRCA	0.0383	0.0430	0.0843	0.0748	0.0748	0.0878	0.0672
Average	STRCA	0.0772	0.0840	0.1323	0.1306	0.1275	0.1376	0.1149
	CNN	0.0590	0.0538	0.1055	0.1177	0.1034	0.1111	0.0917
	FBTRCA	0.0491	0.0504	0.0895	0.1043	0.0988	0.1162	0.0847

MRCP signals [13]. The spatial filter TRCA showed a better performance when compared to other spatial filters. Among the compared spatial filters, the discriminative canonical pattern matching is a spatial filter that maximizes the inter-class covariance and minimizes the intraclass covariance. However, even the performance of this filter was worse than

TRCA. Therefore, the spatial filtering does not maximize the intraclass covariance in this work. As given in section 2.2.1, TRCA searches for the projection matrix that maximizes the covariance of each trial in each class. The ideal spatial filter transforms the EEG signals of all trials into a single trial so that the covariances among these trials are maximized. The grand

Table 12. Mean of classification accuracies of dataset II across ten-folds (Movement vs Movement). The best case is highlighted in bold.

Subject	Method	Motion pair										Average	
		SU-PR	SU-HO	SU-PG	SU-LG	PR-HO	PR-PG	PR-LG	HO-PG	HO-LG	PG-LG		
1	STRCA	0.6000	0.5571	0.6429	0.5286	0.5857	0.6286	0.5857	0.6000	0.5929	0.5929	0.5929	0.5914
	CNN	0.5357	0.7071	0.6286	0.5857	0.6000	0.6000	0.6000	0.6429	0.6571	0.6571	0.5429	0.6100
	FBTRCA	0.4929	0.6571	0.6643	0.6143	0.6429	0.6571	0.6071	0.6643	0.6429	0.6429	0.5857	0.6229
2	STRCA	0.5571	0.6571	0.6500	0.7286	0.5714	0.7071	0.6214	0.5714	0.5786	0.5786	0.5143	0.6157
	CNN	0.5500	0.6571	0.6286	0.7714	0.6714	0.7071	0.6857	0.5571	0.5571	0.5571	0.5286	0.6314
	FBTRCA	0.5643	0.7000	0.6929	0.7000	0.6429	0.7071	0.7000	0.6500	0.6643	0.6643	0.5143	0.6536
3	STRCA	0.5071	0.5857	0.7000	0.6786	0.6286	0.6571	0.6214	0.5786	0.5500	0.5500	0.6071	0.6114
	CNN	0.5643	0.6500	0.7214	0.7286	0.6429	0.7143	0.6786	0.6500	0.6071	0.6071	0.6143	0.6571
	FBTRCA	0.5857	0.7571	0.7500	0.6929	0.6643	0.7357	0.6929	0.6500	0.6500	0.6500	0.6714	0.6850
4	STRCA	0.6214	0.7571	0.7786	0.8214	0.7071	0.7571	0.7643	0.6786	0.7357	0.7357	0.6214	0.7243
	CNN	0.6643	0.8143	0.8143	0.8214	0.7714	0.8143	0.8214	0.7357	0.7500	0.7500	0.6786	0.7686
	FBTRCA	0.6571	0.8500	0.8429	0.8929	0.8286	0.8214	0.8429	0.7143	0.7786	0.7786	0.7286	0.7957
5	STRCA	0.7286	0.6143	0.7571	0.7214	0.7214	0.7357	0.8143	0.7286	0.6929	0.6929	0.6357	0.7150
	CNN	0.7071	0.6929	0.7714	0.8000	0.6929	0.7643	0.8643	0.6929	0.6571	0.6571	0.6357	0.7279
	FBTRCA	0.7357	0.6786	0.8071	0.7429	0.7571	0.7929	0.8214	0.7429	0.6857	0.6857	0.6643	0.7429
6	STRCA	0.5571	0.5857	0.8143	0.7143	0.5357	0.6929	0.7143	0.7929	0.6500	0.6500	0.6500	0.6707
	CNN	0.5643	0.5786	0.7714	0.7429	0.6071	0.7214	0.6357	0.8214	0.7143	0.7143	0.6500	0.6807
	FBTRCA	0.6071	0.5500	0.8214	0.7429	0.6143	0.7929	0.6857	0.7929	0.6786	0.6786	0.7000	0.6986
7	STRCA	0.5500	0.5000	0.5357	0.5571	0.5143	0.5429	0.5357	0.5214	0.5429	0.5429	0.4500	0.5250
	CNN	0.5500	0.5429	0.4929	0.5786	0.5643	0.5429	0.5643	0.5571	0.6214	0.6214	0.5071	0.5521
	FBTRCA	0.5929	0.5286	0.5357	0.6000	0.6000	0.5571	0.5143	0.5929	0.5571	0.5571	0.5643	0.5643
8	STRCA	0.6357	0.7214	0.6214	0.6857	0.6500	0.5643	0.6143	0.5714	0.6429	0.6429	0.5929	0.6300
	CNN	0.5714	0.7286	0.6929	0.7071	0.6786	0.5714	0.6357	0.6429	0.6714	0.6714	0.5357	0.6436
	FBTRCA	0.6500	0.7643	0.6929	0.7786	0.7643	0.6500	0.6857	0.7143	0.6857	0.6857	0.5643	0.6950
9	STRCA	0.6143	0.5286	0.7071	0.7000	0.6214	0.7000	0.7571	0.6429	0.6000	0.6000	0.6357	0.6507
	CNN	0.6643	0.5357	0.6857	0.6500	0.6500	0.7571	0.7286	0.6786	0.7143	0.7143	0.6643	0.6729
	FBTRCA	0.6500	0.6071	0.7571	0.6357	0.6643	0.7429	0.7929	0.6500	0.6571	0.6571	0.6714	0.6829
Average	STRCA	0.5968	0.6119	0.6897	0.6817	0.6151	0.6651	0.6698	0.6317	0.6206	0.6206	0.5889	0.6371
	CNN	0.5968	0.6563	0.6897	0.7095	0.6532	0.6881	0.6905	0.6643	0.6611	0.6611	0.5952	0.6605
	FBTRCA	0.6151	0.6770	0.7294	0.7111	0.6865	0.7175	0.7048	0.6857	0.6667	0.6667	0.6294	0.6823

Table 13. Deviation of classification accuracies of dataset II across ten-folds (Movement vs Movement). The best case is highlighted in bold.

Subject	Method	Motion pair										Average
		SU-PR	SU-HO	SU-PG	SU-LG	PR-HO	PR-PG	PR-LG	HO-PG	HO-LG	PG-LG	
1	STRCA	0.1075	0.0878	0.1260	0.1355	0.1157	0.2043	0.1338	0.1223	0.1431	0.1906	0.1367
	CNN	0.1271	0.1139	0.1054	0.1205	0.1313	0.1622	0.1127	0.1065	0.1107	0.1223	0.1213
	FBTRCA	0.1324	0.0999	0.1216	0.1075	0.1347	0.1338	0.0967	0.1169	0.0825	0.1460	0.1172
2	STRCA	0.1054	0.1421	0.0979	0.1205	0.0891	0.1035	0.0955	0.1782	0.1188	0.1536	0.1205
	CNN	0.1617	0.1536	0.0811	0.0878	0.1021	0.1235	0.1075	0.1608	0.1643	0.0768	0.1219
	FBTRCA	0.1139	0.1677	0.1216	0.0940	0.1429	0.1446	0.1251	0.1280	0.1216	0.1251	0.1285
3	STRCA	0.1035	0.1205	0.0878	0.1315	0.1460	0.1251	0.1119	0.1559	0.1262	0.1129	0.1221
	CNN	0.1188	0.1280	0.0979	0.0811	0.1065	0.1388	0.1758	0.1324	0.1078	0.1021	0.1189
	FBTRCA	0.1338	0.1176	0.0907	0.1349	0.1119	0.1067	0.1306	0.1088	0.1595	0.1223	0.1217
4	STRCA	0.0757	0.1223	0.0979	0.0907	0.1407	0.0838	0.1169	0.1078	0.1013	0.1067	0.1044
	CNN	0.0828	0.1075	0.0904	0.1271	0.1677	0.1075	0.1129	0.0894	0.0694	0.1024	0.1057
	FBTRCA	0.0878	0.1366	0.0738	0.0907	0.0904	0.1024	0.0999	0.1214	0.0786	0.0563	0.0938
5	STRCA	0.0999	0.1127	0.1550	0.1485	0.0979	0.0894	0.0768	0.0999	0.1013	0.1280	0.1109
	CNN	0.0855	0.0894	0.0878	0.1608	0.1262	0.1216	0.0855	0.1349	0.1205	0.1235	0.1136
	FBTRCA	0.1119	0.1438	0.0678	0.1513	0.1436	0.1088	0.0772	0.0838	0.1436	0.1651	0.1197
6	STRCA	0.1251	0.1205	0.1513	0.1010	0.1852	0.0828	0.1010	0.0855	0.1280	0.1139	0.1194
	CNN	0.0855	0.1088	0.1107	0.1223	0.1552	0.1188	0.1366	0.1315	0.1166	0.0919	0.1178
	FBTRCA	0.1357	0.1169	0.1129	0.1075	0.1475	0.1280	0.1127	0.1324	0.1129	0.1380	0.1245
7	STRCA	0.1719	0.1117	0.1271	0.1338	0.1460	0.1176	0.1357	0.1431	0.1587	0.1349	0.1380
	CNN	0.1508	0.1313	0.1324	0.1035	0.1324	0.1475	0.0919	0.1421	0.1216	0.1446	0.1298
	FBTRCA	0.1508	0.1313	0.0772	0.1550	0.0768	0.1107	0.1380	0.1067	0.0811	0.1139	0.1141
8	STRCA	0.1188	0.0979	0.1169	0.1355	0.0855	0.1595	0.0964	0.1166	0.1214	0.1013	0.1150
	CNN	0.0891	0.1157	0.1169	0.1088	0.1271	0.1388	0.1631	0.1429	0.0602	0.1129	0.1176
	FBTRCA	0.1407	0.1067	0.0828	0.1366	0.1169	0.1559	0.1223	0.1065	0.1355	0.1280	0.1232
9	STRCA	0.1313	0.1269	0.1324	0.1572	0.1651	0.0738	0.1127	0.1468	0.1355	0.1235	0.1305
	CNN	0.1390	0.1357	0.1756	0.1595	0.1139	0.1223	0.0999	0.1725	0.1166	0.0955	0.1331
	FBTRCA	0.1366	0.1024	0.1176	0.1665	0.1349	0.0904	0.0919	0.1280	0.0999	0.0690	0.1137
Average	STRCA	0.1154	0.1158	0.1214	0.1282	0.1301	0.1155	0.1090	0.1285	0.1260	0.1295	0.1220
	CNN	0.1156	0.1204	0.1109	0.1191	0.1292	0.1312	0.1207	0.1348	0.1098	0.1080	0.1200
	FBTRCA	0.1271	0.1248	0.0962	0.1271	0.1222	0.1202	0.1105	0.1147	0.1128	0.1182	0.1174

Table 14. Mean of classification accuracies of dataset II across ten-folds (Movement vs Resting). The best case is highlighted in bold.

Subject	Method	Motion pair					Average
		SU-RE	PR-RE	HO-RE	PG-RE	LG-RE	
1	STRCA	0.7143	0.7643	0.8071	0.7214	0.6857	0.7386
	CNN	0.7929	0.7643	0.8214	0.7643	0.7357	0.7757
	FBTRCA	0.8571	0.8000	0.8071	0.8000	0.7643	0.8057
2	STRCA	0.9571	0.9357	0.9000	0.9071	0.8643	0.9129
	CNN	0.9429	0.9500	0.9214	0.8929	0.9071	0.9229
	FBTRCA	0.9571	0.9571	0.9357	0.9286	0.9286	0.9414
3	STRCA	0.8214	0.8571	0.7571	0.7500	0.7143	0.7800
	CNN	0.8000	0.8786	0.8357	0.8571	0.8429	0.8429
	FBTRCA	0.8714	0.9000	0.9071	0.8571	0.7857	0.8643
4	STRCA	0.9286	0.9286	0.9071	0.8857	0.8786	0.9057
	CNN	0.9500	0.9500	0.9286	0.9000	0.9214	0.9300
	FBTRCA	0.9786	0.9429	0.9214	0.9286	0.9286	0.9400
5	STRCA	0.8571	0.8643	0.8429	0.7786	0.8214	0.8329
	CNN	0.8929	0.9000	0.8571	0.8643	0.8500	0.8729
	FBTRCA	0.8929	0.9143	0.8929	0.8786	0.8857	0.8929
6	STRCA	0.8571	0.8786	0.8714	0.8571	0.8071	0.8543
	CNN	0.8714	0.8929	0.9357	0.8643	0.7857	0.8700
	FBTRCA	0.8357	0.9000	0.9214	0.9071	0.7929	0.8714
7	STRCA	0.6214	0.6357	0.6786	0.5857	0.6429	0.6329
	CNN	0.7786	0.7643	0.7214	0.6643	0.6857	0.7229
	FBTRCA	0.7357	0.7286	0.7643	0.7000	0.7286	0.7314
8	STRCA	0.8000	0.7929	0.7643	0.7714	0.7500	0.7757
	CNN	0.8286	0.8357	0.8571	0.7500	0.7929	0.8129
	FBTRCA	0.8714	0.8643	0.8429	0.7714	0.8000	0.8300
9	STRCA	0.8786	0.8929	0.8000	0.6429	0.7429	0.7914
	CNN	0.8571	0.8714	0.8429	0.6643	0.7214	0.7914
	FBTRCA	0.8857	0.9000	0.8429	0.7071	0.7643	0.8200
Average	STRCA	0.9124	0.9175	0.8219	0.7446	0.7672	0.8327
	CNN	0.9553	0.9461	0.8864	0.8257	0.8423	0.8912
	FBTRCA	0.9672	0.9589	0.9014	0.8364	0.8558	0.9039

average MRCP is the mean of all the trials in each class. We assume that the ideal single trial is the grand average MRCP, although there might be some biases. The EEG signals of all the trials in each class approach the grand average MRCP of each class during the spatial filtering. Because the inter-trial noises are not correlated, the noise components in EEG signals have lower eigenvalues. The noises are removed from the signals by taking the eigenvectors of the three maximum eigenvalues in the spatial filtering.

In figure 11, the grand average MRCPs of multiple motions are different. In the binary classification, the class of an EEG trial is determined by the distances between the EEG trial and the two grand average MRCPs. The goal is to quantify the relationship between the EEG signals of each trial and the grand average MRCPs of two classes after the spatial filtering. CCPs are in fact used to measure the similarity between EEG signals and the grand average MRCPs. In the calculation of CCPs, three types of correlation coefficients are used to measure the similarity between EEG signals and the grand average MRCPs; these include (a) the correlation between EEG signals and the grand average MRCP, (b) the

canonical correlation between EEG signals and the grand average MRCP, and (c) the canonical correlation between two differences, including the difference between EEG signals and the grand average MRCP and the difference between two grand average MRCPs. Because the three coefficients are measuring the similarity between EEG signals and the grand average MRCP and there are two grand average MRCPs, the number of total CCP features is six.

However, because the differences between the two motions are reflected in their grand average MRCP signals, the similarity between their grand average MRCP signals limits their classification performance. For instance, the classification between *elbow flexion* and *resting* has a high accuracy, but the classification between *elbow flexion* and *elbow extension* achieves an unsatisfactory performance.

The grand average MRCP plays an important role in STRCA and FBTRCA. But the calculation procedure is simple, which is the same as the procedure of the CCP templates given in section 2.2.1. In the training set, the EEG signals of each trial are labeled. The grand average MRCP can be established simply by averaging EEG trials belonging to the same class. Calculating the

Table 15. Deviation of classification accuracies of dataset II across ten-folds (Movement vs Resting). The best case is highlighted in bold.

Subject	Method	Motion pair					Average
		SU-RE	PR-RE	HO-RE	PG-RE	LG-RE	
1	STRCA	0.0952	0.0955	0.1508	0.0979	0.0964	0.1072
	CNN	0.1235	0.0757	0.0772	0.1067	0.1013	0.0969
	FBTRCA	0.0673	0.0999	0.1013	0.0999	0.1119	0.0961
2	STRCA	0.0602	0.0527	0.0602	0.0678	0.1139	0.0710
	CNN	0.0452	0.0588	0.0625	0.0842	0.1067	0.0715
	FBTRCA	0.0369	0.0499	0.0527	0.0476	0.0476	0.0470
3	STRCA	0.0772	0.0825	0.1127	0.1271	0.1347	0.1068
	CNN	0.0940	0.0482	0.1119	0.1260	0.1205	0.1001
	FBTRCA	0.0940	0.0369	0.1067	0.1010	0.0476	0.0773
4	STRCA	0.0952	0.0583	0.1013	0.0904	0.0894	0.0869
	CNN	0.0955	0.0588	0.0673	0.0768	0.0710	0.0739
	FBTRCA	0.0482	0.0563	0.0710	0.0583	0.0753	0.0618
5	STRCA	0.0753	0.0855	0.1054	0.1139	0.0967	0.0954
	CNN	0.0607	0.0964	0.1010	0.0855	0.0855	0.0858
	FBTRCA	0.0772	0.0940	0.0772	0.0894	0.1021	0.0880
6	STRCA	0.0952	0.0678	0.1054	0.1010	0.1431	0.1025
	CNN	0.0452	0.0842	0.0625	0.0786	0.1117	0.0764
	FBTRCA	0.1013	0.1021	0.0855	0.0757	0.1188	0.0967
7	STRCA	0.1545	0.1523	0.1179	0.0738	0.1506	0.1298
	CNN	0.1665	0.0955	0.1407	0.1470	0.1396	0.1379
	FBTRCA	0.1719	0.1710	0.1119	0.0999	0.1421	0.1394
8	STRCA	0.1054	0.0710	0.0828	0.1608	0.1226	0.1085
	CNN	0.1313	0.0828	0.0891	0.0967	0.0710	0.0942
	FBTRCA	0.1107	0.0625	0.0878	0.1205	0.0878	0.0939
9	STRCA	0.0678	0.0842	0.1054	0.1117	0.1223	0.0983
	CNN	0.0825	0.1157	0.0811	0.1349	0.1699	0.1168
	FBTRCA	0.0602	0.0838	0.0738	0.1407	0.1013	0.0920
Average	STRCA	0.0918	0.0833	0.1047	0.1049	0.1189	0.1007
	CNN	0.0938	0.0796	0.0882	0.1040	0.1086	0.0948
	FBTRCA	0.0853	0.0841	0.0853	0.0925	0.0927	0.0880

grand average MRCP with the training set and using the grand average MRCP as a template can be derived from [35–37]. Their grand average MRCP was extracted from the surrogate channel in the training set. In our work, the template is extracted from the channels in the motor cortex. The number of channels in the grand average MRCP is then minimized by the TRCA.

When applying filter bank selection to STRCA, the EEG signals are divided into several frequency ranges. As shown in figure 12(a), the low cut-offs of these frequency ranges are small, while the high cut-offs are sorted in an arithmetic sequence. In MRCP analysis, the grand average MRCP shows an increase followed by a decrease around the movement onset. The grand average MRCP is a low-frequency signal. Figure 12(b) illustrates the power spectrum of the grand average MRCP. As the frequency increases, the time window with high power (yellow part) becomes narrow. Therefore, it is necessary to maintain the low-frequency components, as then the increasing high cut-off of filter banks can introduce more subtle features. For example, in figure 12(a), the 0.5–3 Hz component reflects how the grand average MRCP

changes. As the high cut-off increases to 10 Hz, more local trends are introduced.

The figure 12(a) can be used to explain why the frequency range setting M_3 is better than the other two settings. The correlation coefficients measure the similarity of the unlabeled trial and the grand average MRCPs. As mentioned above, the classification of STRCA and FBTRCA depends on the differences between two grand average MRCPs. The most discriminant features in the grand average MRCP are the increase and decrease around the movement onset (figure 11). Reflected in figure 12(a), the increase and decrease are located in the extremely low-frequency bands (0.5–3 Hz), which indicates the global trend of MRCP signals. When the high cut-off frequency increases, more subtle features are introduced. In the frequency range settings, M_1 and M_2 , the global trend of MRCP signals is removed in the filtering and only subtle or local features are kept. It leads to the results that the increase and decrease trends are removed, and the differences of the increase and decrease between two motions cannot be further analyzed and distinguished in STRCA and FBTRCA.

Table 16. P-value between STRCA and FBTRCA calculated by two-side t-test (Dataset I-RP). The best case is highlighted in bold.

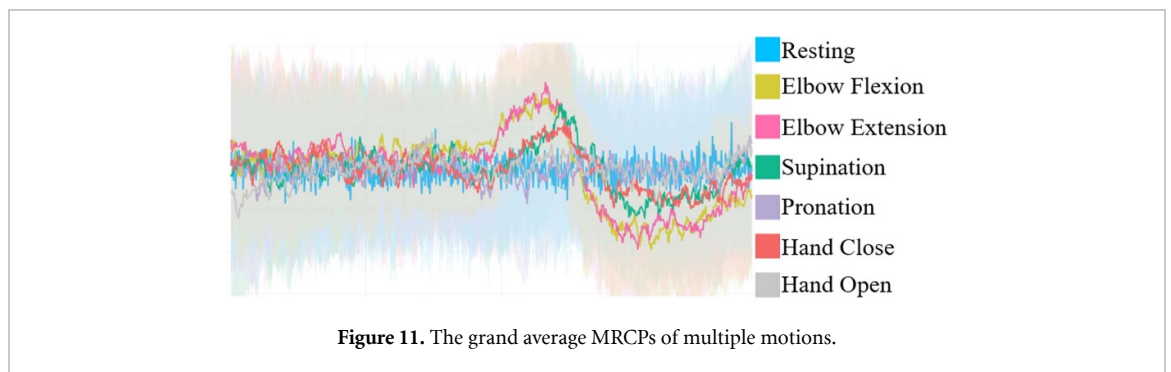
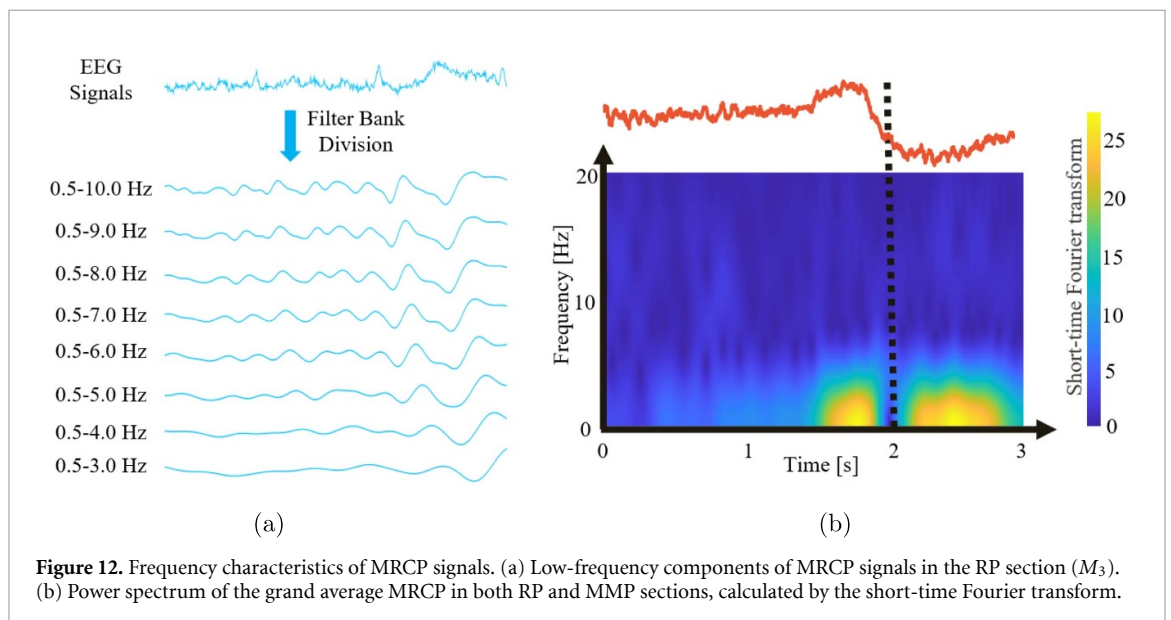
Motion	Subject															Average
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
EF-EE	0.8093	0.6415	0.3785	0.2014	0.0586	0.2806	0.4291	0.8244	0.8923	0.6945	0.4803	0.5653	0.1221	0.0723	1.0000	0.4967
EF-SU	0.2434	0.8720	0.6809	0.0102	0.3955	0.5256	0.4762	0.6485	0.9113	0.3812	0.2564	0.7364	0.7517	0.5872	0.6585	0.5423
EF-PR	0.0142	0.2285	0.1558	0.0499	0.8885	0.7486	0.3082	0.8163	0.8789	0.6733	0.3553	0.1955	0.2388	0.0186	0.4697	0.4027
EF-HC	0.2317	0.6445	0.7486	0.5469	0.6103	0.5914	0.7331	0.4545	1.0000	0.6631	0.4351	1.0000	0.0750	0.0170	0.6215	0.5582
EF-HO	0.6556	0.9036	0.2691	0.1422	0.1127	0.8890	0.0489	0.3618	1.0000	0.4746	0.8859	0.1047	0.5530	0.2557	0.7776	0.4956
EE-SU	0.0023	0.1879	0.1672	0.2962	0.3860	0.2912	0.5560	0.5362	0.5872	0.6432	0.7486	0.8437	0.1041	0.6215	0.6601	0.4421
EE-PR	0.0019	0.2809	0.0161	0.7031	0.5843	0.8409	0.4700	0.6713	0.3630	1.0000	0.0126	0.5402	0.5452	0.3220	0.5427	0.4596
EE-HC	0.8070	0.7187	0.8279	0.1401	0.3419	0.7992	0.8914	0.1201	0.3606	0.0838	0.2778	0.5759	0.6085	0.5979	0.4419	0.5062
EE-HO	0.7851	0.1068	0.3125	0.1515	0.8020	0.7007	0.3140	0.0860	0.7771	0.7756	0.6601	0.3219	0.3388	0.6300	0.8569	0.5079
SU-PR	0.8837	0.4410	0.7618	0.1825	0.2068	0.5979	0.7805	0.8093	1.0000	0.9161	0.3045	0.2962	0.4393	0.2649	0.8984	0.5855
SU-HC	0.8569	0.8535	0.7869	0.0277	0.4017	0.6237	0.7160	0.8558	0.8323	0.2540	0.3913	1.0000	0.4246	0.4077	0.6278	0.6040
SU-HO	0.1558	0.3924	0.4218	0.2257	0.1929	1.0000	0.8697	0.5025	0.7507	0.9003	0.9030	0.6809	0.7521	0.8536	0.4872	0.6059
PR-HC	0.5001	0.7587	1.0000	0.7272	0.4391	0.1145	0.4311	0.6432	0.8120	0.8977	0.8536	0.0739	0.8973	0.5470	0.5520	0.6165
PR-HO	0.8558	1.0000	0.6733	0.6126	0.1346	0.7486	0.8832	0.4915	0.7606	1.0000	0.7911	0.4246	0.4700	0.3760	0.7756	0.6665
HC-HO	0.5025	1.0000	0.6837	0.1356	0.7033	0.8837	0.7450	0.5069	0.7606	0.3388	0.8837	0.7676	0.6893	0.6839	0.3419	0.6418
EF-RE	0.2743	0.3581	0.3760	0.4645	0.1513	1.0000	0.8558	0.0145	0.2878	0.5305	0.1150	0.1589	0.1964	0.8379	0.4453	0.4044
EE-RE	0.6182	0.1753	0.0062	0.6652	0.1769	0.7142	0.6565	0.6601	0.3077	0.7730	0.3581	0.5872	0.6837	0.5165	0.6733	0.5048
SU-RE	0.1892	0.1194	0.2546	0.6132	0.3765	0.8086	1.0000	0.3177	0.6164	0.4313	0.2895	0.4987	0.5241	0.4973	0.7120	0.4832
PR-RE	0.2030	0.3955	0.8681	0.8201	1.0000	1.0000	0.5452	0.4163	0.3018	0.0594	0.0247	0.3306	0.0179	0.2846	0.6733	0.4627
HC-RE	0.4206	0.4228	0.2945	0.8297	0.2564	1.0000	0.6215	0.4073	0.0166	0.8437	0.7372	0.0913	0.5510	0.5872	0.4925	0.5048
HO-RE	0.3388	0.2134	0.2963	1.0000	0.8770	0.5628	0.1515	0.6906	0.1785	0.4291	0.1201	0.1302	0.2163	0.3860	1.0000	0.4394
Average	0.4452	0.5102	0.4752	0.4069	0.4332	0.7010	0.5944	0.5159	0.6379	0.6078	0.4707	0.4725	0.4381	0.4459	0.6528	0.5205

Table 17. P-value between STRCA and FBTRCA calculated by two-side t-test (Dataset I-RP+MMP). The best case is highlighted in bold.

Motion	Subject															Average
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
EF-EE	0.3661	0.3858	0.0082	0.2139	1.0000	0.1291	0.4543	0.5530	0.0008	0.6278	0.6578	0.0379	0.8010	0.0835	0.3471	0.3778
EF-SU	0.4391	0.3180	0.0225	0.1221	0.5530	0.1387	0.4721	0.0747	0.6733	0.2507	0.4218	0.6862	0.6753	0.0913	0.8003	0.3826
EF-PR	0.0838	0.1597	0.0739	0.4697	0.0687	0.1964	0.0490	0.0499	0.0320	0.3433	0.0750	0.0046	0.0329	0.2675	0.0041	0.1274
EF-HC	0.0676	0.7264	0.0024	0.1294	0.0266	0.6394	0.4123	0.6319	0.2095	1.0000	0.2883	0.3750	0.0087	0.1184	0.7007	0.3558
EF-HO	0.0382	0.4825	0.0086	0.1929	0.2546	0.4376	1.0000	0.1869	0.3741	0.0782	0.1346	0.1597	0.0155	0.0021	0.2011	0.2378
EE-SU	0.1433	0.3760	0.0104	0.1615	0.7287	0.2564	0.6900	0.2434	0.1184	0.5456	0.5874	0.5944	0.1268	0.6278	0.7507	0.3974
EE-PR	0.0021	0.2208	0.1030	0.1964	0.1589	0.7486	0.5120	0.2468	0.0083	0.0587	0.1468	0.0439	0.0382	0.0082	0.4118	0.1936
EE-HC	0.0579	1.0000	0.4240	0.1145	0.8451	0.5049	0.7231	0.3579	0.5197	0.1194	0.2853	0.2986	0.2069	0.5165	0.6468	0.4414
EE-HO	0.0208	0.1194	0.0172	0.0925	0.5843	0.3581	0.8820	0.0586	0.1116	1.0000	0.3219	0.0525	0.7886	0.0056	0.4474	0.3240
SU-PR	0.1422	0.6991	0.0943	0.0131	0.4474	0.6005	0.5402	0.0626	0.7873	0.9132	1.0000	0.0026	0.6961	0.1150	0.3349	0.4299
SU-HC	0.1828	0.2643	0.0381	0.0340	0.4247	0.1820	0.6864	0.4419	0.5197	0.0652	1.0000	0.1069	0.1356	0.3968	1.0000	0.3652
SU-HO	0.0126	0.5803	0.1597	0.0531	0.2650	0.5172	0.7886	0.7771	0.0410	0.7606	0.1580	0.0114	0.1206	0.8061	0.8609	0.3942
PR-HC	0.8705	0.3263	0.2163	0.8808	1.0000	0.8086	0.4075	0.4697	0.8875	0.8802	0.8180	0.4351	0.8756	0.6195	1.0000	0.6997
PR-HO	0.1785	0.2295	0.4486	0.1707	0.4077	0.8020	1.0000	0.4286	0.2758	0.2188	0.7313	0.2188	0.2557	0.1150	0.6924	0.4116
HC-HO	0.8957	0.2858	0.2593	0.3760	0.7486	0.7873	1.0000	1.0000	0.3750	0.4474	0.0838	0.7559	0.4822	0.9052	0.2095	0.5741
EF-RE	0.0026	0.0203	0.4583	0.0544	0.1278	0.5560	0.0129	0.2878	0.0345	0.3065	0.0069	0.6182	0.0453	0.1964	0.0553	0.1855
EE-RE	0.1686	0.3061	0.0150	0.3553	0.1387	0.1510	0.7269	0.4486	0.0134	0.4313	0.0756	0.7642	0.1753	0.1451	0.0372	0.2635
SU-RE	0.0401	0.3306	0.1108	0.0017	0.4286	0.0525	0.4230	0.6601	0.0156	0.0838	0.5724	0.4331	0.1118	0.1221	0.0203	0.2271
PR-RE	0.0114	0.0789	0.0199	0.1346	0.1985	0.7947	0.2986	0.2945	0.2266	0.1285	0.2527	0.0544	0.0007	0.0187	0.1897	0.1801
HC-RE	0.0624	0.4825	0.3858	0.0645	0.1673	0.0965	0.7842	0.1278	0.0035	0.0051	0.0442	0.2199	0.2418	0.2295	0.0015	0.1944
HO-RE	0.0439	0.3794	0.1036	0.3624	0.3568	0.3487	0.2069	0.3388	0.0623	0.2301	0.0305	0.0424	0.1145	0.0378	0.2447	0.1935
Average	0.1824	0.3701	0.1419	0.1997	0.4253	0.4336	0.5748	0.3686	0.2519	0.4045	0.3663	0.2817	0.2833	0.2585	0.4265	0.3313

Table 18. *P*-value between STRCA and FBTRCA calculated by two-side *t*-test (Dataset II). The best case is highlighted in bold.

Motion	Subject									Average
	1	2	3	4	5	6	7	8	9	
SU-PR	0.0624	0.8859	0.1592	0.3428	0.8820	0.4029	0.5607	0.8090	0.5585	0.5182
SU-HO	0.0287	0.5452	0.0047	0.1267	0.2806	0.5096	0.6065	0.3618	0.1449	0.2899
SU-PG	0.7033	0.3968	0.2263	0.1145	0.3624	0.9061	1.0000	0.1323	0.3836	0.4695
SU-LG	0.1346	0.5617	0.8132	0.0951	0.7530	0.5480	0.5165	0.1443	0.3864	0.4392
PR-HO	0.3222	0.1964	0.5470	0.0338	0.5241	0.3079	0.1177	0.0225	0.5330	0.2894
PR-PG	0.7157	1.0000	0.1482	0.1419	0.2158	0.0527	0.7829	0.2401	0.2605	0.3953
PR-LG	0.6864	0.1318	0.2057	0.1235	0.8379	0.5579	0.7304	0.1642	0.4474	0.4317
HO-PG	0.2451	0.2723	0.2504	0.4956	0.7331	1.0000	0.2218	0.0104	0.9089	0.4597
HO-LG	0.3510	0.1283	0.1375	0.3045	0.8992	0.6031	0.8027	0.4660	0.2973	0.4433
PG-LG	0.9261	1.0000	0.2378	0.0117	0.6706	0.3886	0.0556	0.5867	0.4351	0.4791
SU-RE	0.0011	1.0000	0.2100	0.1558	0.3087	0.6319	0.1352	0.1567	0.8061	0.3784
PR-RE	0.4246	0.3630	0.1510	0.5843	0.2295	0.5872	0.2160	0.0282	0.8513	0.3817
HO-RE	1.0000	0.1753	0.0068	0.7193	0.2418	0.2593	0.1127	0.0543	0.3061	0.3195
PG-RE	0.0925	0.4240	0.0514	0.2237	0.0425	0.2263	0.0093	1.0000	0.2725	0.2602
LG-RE	0.1099	0.1171	0.1313	0.1929	0.1656	0.8108	0.2069	0.3082	0.6747	0.3019
Average	0.3869	0.4799	0.2187	0.2444	0.4764	0.5195	0.4050	0.2990	0.4844	0.3905

**Figure 11.** The grand average MRCPs of multiple motions.**Figure 12.** Frequency characteristics of MRCP signals. (a) Low-frequency components of MRCP signals in the RP section (M_3). (b) Power spectrum of the grand average MRCP in both RP and MMP sections, calculated by the short-time Fourier transform.

Therefore, the frequency range setting M_3 is used in the analysis of MRCP signals.

In the frequency range setting M_3 of figure 4, the filter bank 10 (e.g. 0.05–10 Hz) covers the ranges of the rest filter banks and thus contains the information from other frequency bands. However, the other

filter banks are still important in the signal processing. This can be explained by the classification accuracies in multiple filter banks. In figure 7, the average classification accuracies of multiple filter banks are presented. When the low cut-off of the filter bank is fixed to a small value, the accuracy of STRCA changes as

the high cut-off frequency increases in an arithmetic sequence. The best accuracies among these bands are different among subjects and motion pairs. A possible reason is the influence of noise in the frequencies greater than the best frequency, and the insufficient information in the frequencies smaller than the best frequency. The best frequency bands cannot be determined directly by giving a certain frequency range. The filter bank selection across these frequency bands in the proposed FBTRCA is necessary when tackling this problem. Therefore, although the filter bank 10 covers all the other filter banks, the other filter banks are still necessary. The filter bank 10 may be influenced by undetected noises and is not in the best frequency range.

During the development of the FBTRCA method, the processing capabilities in SSVEP and MI are considered. This includes the canonical correlation coefficients in SSVEP and the feature selection on filter banks in MI.

SSVEP also uses canonical correlations as features for classification, but there are some differences. In SSVEP, there are two kinds of templates when calculating the correlation coefficients [33]. The first template is the CCP template averaged across trials in each class. The second template is the sinusoidal function. Because the visual stimulus in SSVEP has a specific frequency, the similarity between EEG signals and the sinusoidal function with a specific frequency can predict the frequencies of visual stimuli. Compared to SSVEP, MRCP has a natural drawback: the templates cannot be measured with frequencies. As shown in figure 11, the differences between the grand average MRCPs of a motion pair are different. For example, the differences between the grand average MRCPs of *elbow flexion* and *elbow extension* is smaller compared to the differences between *elbow flexion* and *supination*. We cannot measure how much smaller the differences are, but they can be measured by the frequencies of visual stimuli in SSVEP. Therefore, it is difficult to classify the MRCP signals by adopting the maximum value of correlation coefficients as in SSVEP. The coefficients calculated in STRCA are used as features for further improvement by feature selection.

In section 3, the classification results between EEG signals in the RP section and the RP+MMP sections are compared. The EEG signals in RP+MMP show an improvement over those in only the RP section. The reason is that the grand average MRCPs also show the differences between the two actions in the MMP section. Similarly to the result analysis of dataset II, it is not necessary to locate the onset of the action and divide the EEG signals into RP and MMP sections if one is only considering the classification of two classes. The classification results in the RP section indicate whether this method works in pre-movement decoding. The classification results in both the RP and MMP sections reflect the best classification performance that this method can achieve.

The feature selection methods we used are based on mutual information instead of similarity or sparse learning [38]. Mutual information is preferred in this case because it has shown to have good performances previously in FBCSP. Mutual information is also used in MI analysis when selecting features from filter banks. Because of the efficiency of mutual information in feature selection on filter banks, feature selection methods based on mutual information were first considered. Compared to feature selection with CNN, the feature selection based on mutual information achieved better performance, as shown in figure 9. Compared to mutual information, sparse learning is suitable in the feature selection from both multiple filter banks and time windows [32].

Traditional EEG processing methods have three key points, spectral, spatial and temporal. The proposed FBTRCA method is associated with one point of the traditional method, spectral, i.e. filter banking. According to our research experience on MRCP signals, FBTRCA and STRCA cannot well deal with the spatial optimization. The spatial filtering in STRCA and FBTRCA work as unrelated-components rejection. The introduce of more channels will lead to a decrease of the classification performance of STRCA and FBTRCA. The temporal characteristic used in FBTRCA and STRCA is the correlation, which measures the similarity of two time series. It can be further improved in future work. Besides, the STRCA and FBTRCA can only be used in the binary classification limiting to the scheme of STRCA. Migrating the two binary classification methods to the multi-class classification task is also an important work. In MRCP signals, the grand average MRCP is a special concept in EEG processing, which averages across all trials belonging to a class. However, averaging is a simple approach to finding the center points of all trials and inevitably introduces unexpected influences like outlier points. It is also possible to find a hyper-space that casts the trials to the discriminant points.

5. Conclusion

The proposed FBTRCA method incorporates the filter bank technique and solves the unstable accuracy problem. There are four steps in FBTRCA. First, EEG signals are divided into multiple sub-bands in the low-frequency domain. Second, CCP features are extracted from these sub-bands with the STRCA method. Then, the minimum redundancy maximum relevance method is used to optimize and select the CCP features. Finally, the selected features are classified with the SVM binary classifier. When decoding the pre-movement pattern in the RP and MMP section, the average accuracy increases from 0.8228 ± 0.1149 (STRCA) to 0.8968 ± 0.0847 (FBTRCA) in the binary classification between the actions and the resting state; the average accuracy increases from 0.6611

± 0.1432 (STRCA) to 0.7178 ± 0.1274 (FBTRCA) in the binary classification between two actions.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This work was carried out as part of the doctoral programme in Experimental Science and Technology at the University of Vic-Central University of Catalonia, Vic, Catalonia, Spain.

J S-C work was partially based upon work from COST Action CA18106, supported by COST (European Cooperation in Science and Technology). C F C work was partially supported by Grants PICT 2017-3208, PICT 2020-SERIEA-00457, UBACYT 20020190200305BA and UBACYT 20020170100192BA (Argentina). The authors would like to thank Pau Solé-Vilaró for the English proofreading of this manuscript.


Code availability

The code of this work is freely available here: <https://github.com/plustar/Movement-Related-Cortical-Potential>.


ORCID iDs

Hao Jia  <https://orcid.org/0000-0003-1356-7463>

Zhe Sun  <https://orcid.org/0000-0002-6531-0769>

Feng Duan  <https://orcid.org/0000-0002-2179-2460>

Yu Zhang  <https://orcid.org/0000-0003-4087-6544>

Cesar F Caiafa  <https://orcid.org/0000-0001-5437-6095>

Jordi Solé-Casals  <https://orcid.org/0000-0002-6534-1979>

References

- [1] Schalk G, McFarland D J, Hinterberger T, Birbaumer N and Wolpaw J R 2004 BCI2000: a general-purpose brain-computer interface (BCI) system *IEEE Trans. Biomed. Eng.* **51** 1034–43
- [2] Ramadan R A and Vasilakos A V 2016 Brain computer interface: control signals review *Neurocomputing* **223** 26–44
- [3] Padfield N, Zabalza J, Zhao H, Masero V and Ren J 2019 EEG-based brain-computer interfaces using motor-imagery: techniques and challenges *Sensors* **19** 1423
- [4] Zhang C, Kim Y-K and Eskandarian A 2021 EEG-inception: an accurate and robust end-to-end neural network for EEG-based motor imagery classification *J. Neural Eng.* **18** 046014
- [5] Zhang X, Yao L, Wang X, Monaghan J, Mcalpine D and Zhang Y 2021 A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers *J. Neural Eng.* **18** 031002
- [6] Lisi G, Rivela D, Takai A and Morimoto J 2018 Markov switching model for quick detection of event related desynchronization in EEG *Front. Neurosci.* **12** 24
- [7] McFarland D J, Sarnacki W A and Wolpaw J R 2015 Effects of training pre-movement sensorimotor rhythms on behavioral performance *J. Neural Eng.* **12** 066021
- [8] Sburlea A I, Montesano L and Minguez J 2015 Continuous detection of the self-initiated walking pre-movement state from EEG correlates without session-to-session recalibration *J. Neural Eng.* **12** 036007
- [9] Olsen S, Signal N, Niazi I K, Christensen T, Jochumsen M and Taylor D 2018 Paired associative stimulation delivered by pairing movement-related cortical potentials with peripheral electrical stimulation: an investigation of the duration of neuromodulatory effects *Neuromodulation* **21** 362–7
- [10] Karimi F, Kofman J, Mrachacz-Kersting N, Farina D and Jiang N 2017 Detection of movement related cortical potentials from EEG using constrained ICA for brain-computer interface applications *Front. Neurosci.* **11** 356
- [11] Wang K, Xu M, Wang Y, Zhang S, Chen L and Ming D 2020 Enhance decoding of pre-movement EEG patterns for brain-computer interfaces *J. Neural Eng.* **17** 016033
- [12] Shakeel A, Navid M S, Anwar M N, Mazhar S, Jochumsen M and Niazi I K 2015 A review of techniques for detection of movement intention using movement-related cortical potentials *Comput. Math. Methods Med.* **2015** 1–13
- [13] Duan F, Jia H, Sun Z, Zhang K, Dai Y and Zhang Y 2021 Decoding pre-movement patterns with task-related component analysis *Cogn. Comput.* **13** 1389–405
- [14] Mammone N, Ieracitano C and Morabito F C 2020 A deep CNN approach to decode motor preparation of upper limbs from time-frequency maps of EEG signals at source level *Neural Netw.* **124** 357–72
- [15] Jeong J-H, Kwak N-S, Guan C and Lee S-W 2020 Decoding movement-related cortical potentials based on subject-dependent and section-wise spectral filtering *IEEE Trans. Neural Syst. Rehabil. Eng.* **28** 687–98
- [16] Ofner B, Schwarz A, Pereira J and Müller-Putz G R 2017 Upper limb movements can be decoded from the time-domain of low-frequency EEG *PLoS One* **12** e0182578
- [17] Bin G, Gao X, Yan Z, Hong B and Gao S 2009 An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method *J. Neural Eng.* **6** 046002
- [18] Lin Z, Zhang C, Wu W and Gao X 2006 Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs *IEEE Trans. Biomed. Eng.* **53** 2610–14
- [19] Minpeng X, Xiao X, Wang Y, Hongzhi Q, Jung T P and Ming D 2018 A brain-computer interface based on miniature-event-related potentials induced by very small lateral visual stimuli *IEEE Trans. Biomed. Eng.* **65** 1166–75
- [20] Jiao Y, Zhang Y, Wang Y, Wang B, Jin J and Wang X 2018 A novel multilayer correlation maximization model for improving CCA-based frequency recognition in SSVEP brain-computer interface *Int. J. Neural Syst.* **28** 1750039
- [21] Chen X, Wang Y, Gao S, Jung T-P and Gao X 2015 Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface *J. Neural Eng.* **12** 046008
- [22] Pfurtscheller G and Neuper C 2001 Motor imagery and direct brain-computer communication *Proc. IEEE* **89** 1123–34
- [23] Wang J-G, Shao H-M, Yao Y, Liu J-L and Ma S-W 2021 A personalized feature extraction and classification method for motor imagery recognition *Mobile Net. Appl.* **26** 1–13
- [24] Ang K K, Chin Z Y, Zhang H and Guan C 2008 Filter bank common spatial pattern (FBCSP) in brain-computer

- interface 2008 *IEEE Int. Joint Conf. on Neural Networks (IEEE World Congress on Computational Intelligence)* pp 2390–7
- [25] Ding C and Peng H 2003 Minimum redundancy feature selection from microarray gene expression data *Computational Systems Bioinformatics. Proc. 2003 IEEE Bioinformatics Conf. (CSB2003)* pp 523–8
- [26] Peng H, Long F and Ding C 2005 Feature selection based on mutual information criteria of max-dependency, max-relevance and min-redundancy *IEEE Trans. Pattern Anal. Mach. Intell.* **27** 1226–38
- [27] Rodriguez-Lujan I, Huerta R, Elkan C and Cruz C S 2010 Quadratic programming feature selection *J. Mach. Learn. Res.* **11** 1491–516
- [28] Lin D and Tang X 2006 Conditional infomax learning: an integrated framework for feature extraction and fusion *Proc. 9th European Conf. on Computer Vision—Volume Part I (ECCV'06)* (Berlin: Springer) pp 68–82
- [29] Herman G, Zhang B, Wang Y, Ye G and Chen F 2013 Mutual information-based method for selecting informative feature sets *Pattern Recognit.* **46** 3315–27
- [30] Nguyen X V, Chan J, Romano S and Bailey J 2014 Effective global approaches for mutual information based feature selection *Proc. 20th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD '14)* (New York: Association for Computing Machinery) pp 512–21
- [31] Ofner P, Schwarz A, Pereira J, Wyss D, Wildburger R and Müller-Putz G R 2019 Attempted arm and hand movements can be decoded from low-frequency EEG from persons with spinal cord injury *Sci. Rep.* **9** 7134
- [32] Zhang Y, Nam C S, Zhou G, Jin J, Wang X and Cichocki A 2019 Temporally constrained sparse group spatial patterns for motor imagery BCI *IEEE Trans. Cybernetics* **49** 3322–32
- [33] Nakanishi M, Wang Y, Chen X, Wang Y-T, Gao X and Jung T-P 2018 Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis *IEEE Trans. Biomed. Eng.* **65** 104–12
- [34] Baratloo A, Hosseini M and Negida A 2015 Part 1: simple definition and calculation of accuracy, sensitivity and specificity *Emergency* **3** 2
- [35] Niazi I K, Jiang N, Tiberghien O, Nielsen J F, Dremstrup K and Farina D 2011 Detection of movement intention from single-trial movement-related cortical potentials *J. Neural Eng.* **8** 066009
- [36] Niazi I K, Jiang N, Jochumsen M, Nielsen J F, Dremstrup K and Farina D 2013 Detection of movement-related cortical potentials based on subject-independent training *Med. Biol. Eng. Comput.* **51** 507–12
- [37] Jochumsen M, Niazi I K, Navid M S, Anwar M N, Farina D and Dremstrup K 2015 Online multi-class brain-computer interface for detection and classification of lower limb movement intentions and kinetics for stroke rehabilitation *Brain-Comput. Interfaces* **2** 202–10
- [38] Li J, Cheng K, Wang S, Morstatter F, Trevino R P, Tang J and Liu H 2018 Feature selection: a data perspective *ACM Comput. Surv.* **50** 1–45

3.4. Multi-class Classification of Upper Limb Movements with Filter Bank Task-related Component Analysis

Multi-class Classification of Upper Limb Movements with Filter Bank Task-related Component Analysis

Hao Jia, Fan Feng, Cesar F. Caiafa, Feng Duan, Yu Zhang, Zhe Sun, Jordi Solé-Casals

Abstract

The classification of limb movements can provide with control commands in non-invasive brain-computer interface. Previous studies on the classification of limb movements have focused on the classification of left/right limbs; however, the classification of different types of upper limb movements has often been ignored despite that it provides more active-evoked control commands in the brain-computer interface. Nevertheless, few machine learning method can be used as the state-of-the-art method in the multi-class classification of limb movements.

This work focuses on the multi-class classification of upper limb movements and proposes the multi-class filter bank task-related component analysis (mFBTRCA) method, which consists of three steps: spatial filtering, similarity measuring and filter bank selection. The spatial filter, namely the task-related component analysis, is first used to remove noise from EEG signals. The canonical correlation measures the similarity of the spatial-filtered signals and is used for feature extraction. The correlation features are extracted from multiple low-frequency filter banks. The minimum-redundancy maximum-relevance selects the essential features from all the correlation features, and finally, the support vector machine is used to classify the selected features.

The proposed method compared against previously used models is evaluated using two datasets. mFBTRCA achieved a classification accuracy of 0.4193 ± 0.0780 (7 classes) and 0.4032 ± 0.0714 (5 classes), respectively, which improves on the best accuracies achieved using the compared methods (0.3590 ± 0.0645 and 0.3159 ± 0.0736 , respectively). The proposed method is expected to provide more control commands in the applications of non-invasive brain-computer interfaces.

Index Terms

Brain-computer Interface, Electroencephalogram, Movement-related Cortical Potential, Upper Limb Movement, Pattern Recognition.

I. INTRODUCTION

A NON-INVASIVE brain-computer interface is a framework that bridges the gap between human brains and external computers [1]–[3]. In non-invasive brain computer interface, electroencephalogram (EEG) signals can be recorded from the brain scalp with the acquisition devices. The acquired multi-channel signals can be used to analyze the brain activities and classify the states of the brain, such as left and right limb movements or multiple visual stimuli. These states can be converted to control commands, and thus used to control robots or other external devices. [4], [5].

In current research on brain-computer interfaces, brain activities such as motor imagery and steady-state visual evoked potentials are frequently used in human-robot interactions [6], [7]. In motor imagery, the commands are generated by classifying movements of the left/right hand, a foot or the tongue [8]–[11]. Compared to the imagination of movement, movement execution refers to the actual movement of limbs and can evoke more distinguished activity in brain signals [12]. In the robot controlling with motor imagery, some subjects prefer to executing the movements instead of imagining the movements [13]. The reason is that the movement execution will provide a stronger response in the brain than the movement imagination. In both the imagined movements and the executed movement, these active-evoked commands are controlled by human intent. In steady-state visual evoked potentials, the number of commands depends on the number of visual stimuli, and hence there are more control commands [14]–[16]. However, the steady-state visual evoked potential is evoked by external visual stimuli. When there are no external visual stimuli, the subjects are unable to generate these control commands intentionally. Thus, the passive-evoked commands limit the application of steady-state visual evoked potentials. Movement-related cortical potential (MRCP) is a brain activity related to limb movement [17]. However, current approaches mainly focused on the binary classification between the limb's resting and movement states or two movement states [18]–[27], and very few methods are designed towards multi-class states [12], [28], [29].

Most brain-computer interface studies focus on improving existing classification tasks in motor imagery and steady-state visual evoked potentials [14], [30]–[32]. However, research into less exhausting methods for users has been ignored. Limb movements are active-evoked and are controlled by the intent of the subject. The classification of multiple upper limb movements not only is more friendly to users than visual stimuli, but also has more commands if combined with left and right limb classification. However, there are very few methods for the multi-class classification of limb movements such as elbow flexion and pronation of the single-side limb [33]–[36].

The classification of limb movements can be divided into two cases, double-side limb movements and single-side limb movements, as shown in Fig 1. The double-side limb movement is related to the binary left/right limbs in motor imagery. In

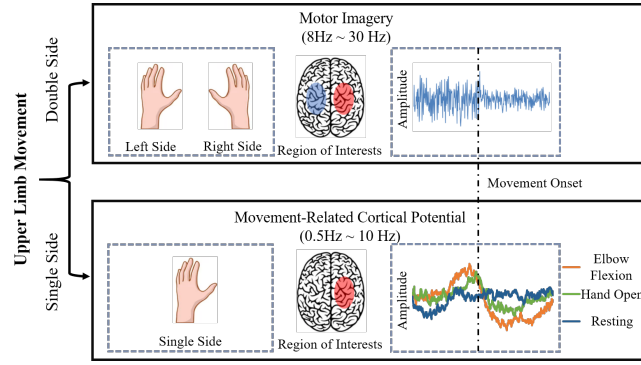


Fig. 1. Differences between double-side and single-side upper limb movement. Motor imagery is used in the classification of double-side movement. In motor imagery, the regions of interest are different between the left side and the right side. The variances of signals change before and after the movement onset. In the single-side classification, the movement-related cortical potential is used. The region of interest is located on a single side of the brain. The amplitude increases and then decreases around the movement onset. The signals' trend also shows differences between motions, such as elbow flexion and hand open. This work focuses on the classification of these single-side motions with movement-related cortical potential.

the multi-class classification, the motions of double-side limb movements are extended to a wide range, including left/right hand movement, left/right foot movement. The single-side limb movement is related to the MRCP signals. The motions include the movement of a single limb, such as elbow flexion, hand close, pronation of the upper limb. The classification of double-side limb movements has obtained many good solutions in the past decades, including the machine-learning methods [37], [38], the deep-learning methods [8], [39], and the source-imaging-based methods [40]. However, the classification of single-side limb movements remains to be developed. Because the data to process in both single-side and double-side movements are the multi-channel EEG signals, the methods in double-side movements can also be used in the single-side movement [25], [36]. However, there are usually performance losses in machine-learning methods and unclear decoding process in deep-learning methods [41]. More related works about the single-side and double-side limb movements are given in the supplementary materials.

In our previous work, we proposed the binary standard task-related component analysis method (bSTRCA) [26]. The bSTRCA follows the processing procedure of spatial filtering and feature extraction. The spatial filter used is the task-related component analysis, and the extracted feature is the canonical correlation coefficient. The bSTRCA method is similar to the matched filter method, as both methods first use spatial filtering to reject the noise in EEG signals and then use a similarity measurement to match the unlabelled EEG signals and the grand average MRCP. However, there are two main differences between the two methods. The first is related to how the spatial filter rejects the noise in the signals. The matched filter method and the bSTRCA method carry out the noise rejection based on the variance and the amplitude of the signals, respectively. MRCP signals are located at the low-frequency band in the frequency domain. In said band, the amplitude of signals mainly reflects the energy change of the signals instead of the variances. The second difference is the role that the similarity measurement plays in the classification. In the matched filter method, the likelihood ratio is the indicator used for classifying the movement and resting states by a threshold criterion. In bSTRCA, correlation coefficients are extracted as features, and a linear discriminant analysis classifier is then used to classify the features. Filter bank selection can further optimize the performance of bSTRCA, and hence the binary filter bank task-related component analysis (bFBTRCA) method was proposed [27]. However, bFBTRCA is not available for multi-class classification because the framework of bSTRCA was initially designed for binary classification.

In this work, we aim to migrate the structure of the bSTRCA method to the multi-class standard task-related component analysis (mSTRCA). Furthermore, we propose the multi-class filter bank task-related component analysis (mFBTRCA) method by incorporating filter bank selection into mSTRCA. The proposed method can be used in the multi-class classification task of single-side limb movements. The proposed method first divides MRCP signals in the low-frequency bands into multiple filter banks. In each filter banks, the multi-channel signals are optimized with the spatial filter. Correlation features are extracted from the optimized features. The correlation features are concatenated and then classified with the support vector machine classifier.

The work firstly explains the decoding of MRCP signals as the rejection of unrelated noises and the measurement of similarity. The mFBTRCA method has a simple structure and shows better performance to other machine-learning and deep-learning methods. This method also extends the use of a TRCA-based method to the context of limb movement.

The structure of this work is as follows. In Section II, the dataset description and details on how the dataset is pre-processed are given. This section also includes a description on the structure of the mFBTRCA method. In Section III, the performance of mFBTRCA is evaluated in the binary classification cases. The proposed method is also compared against other benchmark methods in the multi-class cases. In Section IV, a discussion is given on how the mFBTRCA method uses the information from the MRCP signals, and the bottleneck of mFBTRCA in the multi-class limb movement classifications is also touched upon. Finally, conclusions are given in Section V. A list of acronyms used in this work is included in the supplementary materials

to support the reading of the manuscript.

II. METHOD

A. Dataset Description

Two public EEG datasets (namely datasets I and II) were used to evaluate the performance of the proposed method against the state-of-the-art and baseline methods [12], [42]. In both datasets, the EEG signals were downsampled to 256 Hz , and a notch filter at 50 Hz was applied to avoid the influence of power line interference.

Both datasets have the same acquisition paradigm. Subjects sat on a chair and a screen was in front of the subjects. EEG signals were acquired from the channels on the brain scalp. The channels used in the classification include FC_z , C_3 , C_z , C_4 , CP_z , F_3 , F_z , F_4 , P_3 , P_z and P_4 . At the start of a trial, the screen displayed a cross. Two seconds later, a cue appeared on the screen indicating a motion of the upper limb which the subjects were then supposed to execute. In dataset I, the executed motions include *elbow flexion*, *elbow extension*, *supination*, *pronation*, *hand open*, *hand close* and the *resting* state. Dataset II includes *supination*, *pronation*, *hand open*, *palmar grasp* and *lateral grasp*. In both datasets, The number of trials for each motion were 60 and 72, and the numbers of subjects were 15 and 9, respectively.

Although both datasets have the same paradigm, the time windows of the EEG signals in the two datasets are different. In dataset I, the hand trajectory was simultaneously acquired along with the EEG signals. The movement onset of the executed motions can be located by the hand trajectory. The time window in which the EEG signals will be used for classification purposes lies between one second before the onset and one second after the onset. In dataset II, however, the hand trajectory was not recorded and there is no information about the movement. Therefore, here the time window for the classification corresponds to the two-second window after the cue indicating the start of the executed motions.

The movement onset is located with the movement trajectory in dataset I. The same localization process as in [27] was adopted and the detail about this process is given in the supplementary materials.

B. Binary FBTRCA

The proposed multi-class FBTRCA (mFBTRCA) is developed based on the binary FBTRCA (bFBTRCA). To present the relationship these two methods, we first introduce the structure of bFBTRCA and then detail how mFBTRCA is developed based on bFBTRCA.

The bFBTRCA method is developed by incorporating filter bank selection into the bSTRCA method. The key idea of bFBTRCA is to find the best frequency band in which the bSTRCA method has the best classification performance. Instead of selecting the best frequency band, bFBTRCA selects the best features from features in all frequency bands. The bFBTRCA method first divides EEG signals into multiple filter banks in the low-frequency domain. In each filter bank, bSTRCA calculates the canonical correlation pattern and uses them as the features. In bFBTRCA, the features extracted using bSTRCA in all filter banks are sorted and selected through the minimum redundancy maximum relevance method. Finally, the support vector machine classifies the selected features. The relationship between bSTRCA and bFBTRCA is shown in Fig. 2. The bFBTRCA method has three key points in the classification: (1) the spatial filtering with task-related component analysis, (2) the feature extraction from the canonical correlation pattern, and (3) filter bank selection.

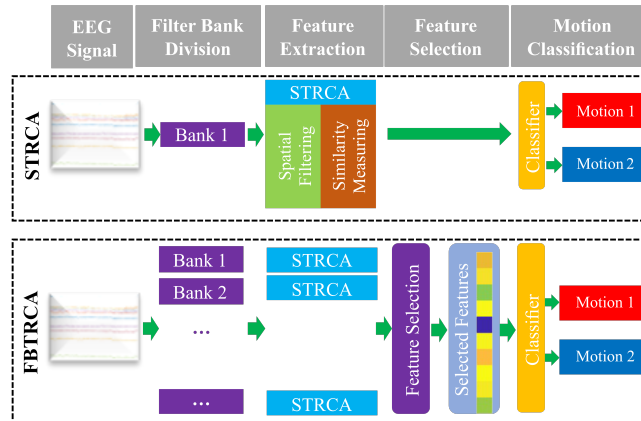


Fig. 2. Relationship between the structure of bSTRCA and bFBTRCA. The STRCA in this figure is either the bSTRCA or the mSTRCA. Both bSTRCA and mSTRCA have two steps: spatial filtering and similarity measuring. The bFBTRCA or mFBTRCA is developed by applying bSTRCA/mSTRCA to multiple banks and enabling feature selection on the features of these banks.

1) *Spatial Filtering*: Because EEG signals are multi-channel signals and the channels are isolated points on the brain scalp, EEG signals naturally have a bad spatial quality. Spatial filtering is commonly used to optimize the spatial quality when processing EEG signals. A spatial filter is used to find a matrix $\mathbf{W} \in \mathbb{R}^{C \times P}$, where C is number of channels and $P \leq C$. By multiplying the given EEG signal $\mathbf{X} \in \mathbb{R}^{C \times T}$ with the matrix \mathbf{W} , the spatial-filtered signal $\mathbf{X}^T \mathbf{W} \in \mathbb{R}^{T \times P}$ is obtained. Here, T is the number of sample points. Compared to the original EEG signal \mathbf{X} , the dimension of the spatial-filtered signals $\mathbf{X}^T \mathbf{W}$ is suppressed. The calculation methods of the spatial filter differ for different brain activities. For example, CSP is widely used in motor imagery, which aims to discern between the channels with the biggest and smallest variances [31]. In steady-state visual evoked potentials, the discriminative canonical pattern matching method is used to maximize the between-class difference and minimize the within-class difference [43]. In the bSTRCA and bFBTRCA methods, the task-related component analysis is used as the spatial filter, thereby extracting the task-related components.

Task-related component analysis optimizes the signals based on covariances of the EEG signals. The training set of EEG signals is given as $\mathcal{X}_k = \{\mathbf{X}_1^k, \mathbf{X}_2^k, \dots, \mathbf{X}_{I_k}^k\}$, where k is the index of classes; for instance, in binary classification, $k = 1, 2$. I_k represents the number of trials of class k . \mathbf{X} are multi-channel EEG signals of size $C \times T$. The task-related component analysis first computes the covariance of the intra-trial and inter-trial of each class. The intra-trial covariance is

$$\mathbf{C}_i^k = \mathbf{X}_i^k (\mathbf{X}_i^k)^T, \quad (1)$$

while the inter-trial covariance is given by

$$\mathbf{C}_{i,j}^k = \mathbf{X}_i^k (\mathbf{X}_j^k)^T + \mathbf{X}_j^k (\mathbf{X}_i^k)^T. \quad (2)$$

The spatial filter is the combination of eigenvectors, which is obtained by solving the following eigen equation:

$$\max_{\boldsymbol{\omega}} J^k = \frac{\boldsymbol{\omega}^T \mathbf{S}^k \boldsymbol{\omega}}{\boldsymbol{\omega}^T \mathbf{Q}^k \boldsymbol{\omega}}. \quad (3)$$

\mathbf{S}^k is the sum of inter-trial covariances of class k

$$\mathbf{S}^k = \sum_{i,j=1, i < j}^{I_k} \mathbf{C}_{i,j}^k, \quad (4)$$

and \mathbf{Q}^k is the sum of the intra-trial covariances of class k

$$\mathbf{Q}^k = \sum_{i=1}^{I_k} \mathbf{C}_i^k. \quad (5)$$

The eigen equation $\max_{\boldsymbol{\omega}} J^k$ can be solved with the generalized Schur decomposition as the generalized eigenvalue problem. The eigenvectors related to the maximal eigenvalues are denoted as $\boldsymbol{\omega}_k \in \mathbb{R}^{C \times P}$, where P is the number of fetched eigenvectors. The spatial filter of task-related component analysis, \mathbf{W} , is the concatenation of eigenvectors of two classes $\mathbf{W} = [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2] \in \mathbb{R}^{C \times 2P}$. The optimized calculation step of the task-related component analysis can be found in [44].

2) *Similarity Measurement*: In MRCP, the grand average MRCP is the mean of EEG signals across trials, denoted as:

$$\hat{\mathbf{X}}^k = \sum_{i=1}^{I_k} \mathbf{X}_i^k / I_k. \quad (6)$$

When measuring the relationship between the grand average MRCP, $\hat{\mathbf{X}}^k$, and each of the trials $\mathbf{X} \in \mathbb{R}^{C \times T}$, bSTRCA and bFBTRCA both use the canonical correlation pattern to measure the similarity. The canonical correlation pattern includes three correlation coefficients:

(1) Correlation between \mathbf{X} and $\hat{\mathbf{X}}^k$:

$$\mathbf{X}_* = \mathbf{X}; \mathbf{X}_k = \hat{\mathbf{X}}^k; \quad (7)$$

$$\rho_{1,k} = \text{corr}(\mathbf{X}_*^T \mathbf{W}, \mathbf{X}_k^T \mathbf{W}); \quad (8)$$

(2) Correlation between \mathbf{X} and $\hat{\mathbf{X}}^k$ after canonical correlation analysis:

$$\mathbf{X}_* = \mathbf{X}; \mathbf{X}_k = \hat{\mathbf{X}}^k; \quad (9)$$

$$[\mathbf{A}_k, \mathbf{B}_k] = \text{cca}(\mathbf{X}_*^T \mathbf{W}, \mathbf{X}_k^T \mathbf{W}); \quad (10)$$

$$\rho_{2,k} = \text{corr}(\mathbf{X}_*^T \mathbf{W} \mathbf{B}_k, \mathbf{X}_k^T \mathbf{W} \mathbf{B}_k); \quad (11)$$

(3) Correlation between $\mathbf{X} - \hat{\mathbf{X}}^k$ and $\hat{\mathbf{X}}^{3-k} - \hat{\mathbf{X}}^k$ after canonical correlation analysis:

$$\mathbf{X}_* = \mathbf{X} - \hat{\mathbf{X}}^k; \mathbf{X}_k = \hat{\mathbf{X}}^{3-k} - \hat{\mathbf{X}}^k; \quad (12)$$

$$[\mathbf{A}_k, \mathbf{B}_k] = cca(\mathbf{X}_*^T \mathbf{W}, \mathbf{X}_k^T \mathbf{W}); \quad (13)$$

$$\rho_{3,k} = corr(\mathbf{X}_*^T \mathbf{W} \mathbf{A}_k, \mathbf{X}_k^T \mathbf{W} \mathbf{A}_k); \quad (14)$$

In the above equations, *corr* corresponds to the two-dimensional Pearson correlation coefficient, and the function symbol *cca* computes the canonical coefficients for the two input data matrices. $\hat{\mathbf{X}}^{3-k}$ denotes the grand average MRCP of the other class ($k = 1, \hat{\mathbf{X}}^{3-k} = \hat{\mathbf{X}}^2; k = 2, \hat{\mathbf{X}}^{3-k} = \hat{\mathbf{X}}^1$). Because the canonical correlation analysis is used in the feature extraction, the EEG signals must be z-normalized before spatial filtering in bSTRCA [45], [46]. The correlations between EEG signals \mathbf{X} and the grand average MRCPs of two classes are calculated in the binary classification. The number of correlation features is six.

3) *Filter Bank Selection*: In bFBTRCA, the filter bank selection consists of two steps: filter bank division and feature selection. After the z-normalization of the original EEG signals, these are divided into subbands in the low-frequency domain. The low cut-off frequencies of these subbands are fixed to 0.5 Hz, while their high cut-off frequencies are in the arithmetic sequence going from 1 Hz to 10 Hz with a 1 Hz step. Therefore, ten filter banks are used in our work.

In each filter bank, bSTRCA is used to extract features. This feature extraction includes spatial filtering and correlation coefficient extraction. The number of features is 6 in each subband, giving a total of 60 features from all subbands.

The adopted feature selection method is the minimum-redundancy maximum-relevance, which is used to select essential features from the total 60 features. Mutual information measures the mutual dependence between two variables, and it quantifies the information from one variable by observing the other variable. In the minimum-redundancy maximum-relevance, relevance is the mutual information between the label and the features, while redundancy is the mutual information between two features. The minimum-redundancy maximum-relevance method optimizes the sequence of features by minimizing the redundancy and maximizing the relevance. The selected features are then classified by the binary support vector machine classifier.

C. Multi-class FBTRCA

The bFBTRCA method is designed to classify two states of limb movement based on the differences between grand average MRCPs of two states. When adapting the bFBTRCA to solve the multi-class classification problem, the spatial filter's structure restricts the framework's extension.

In the spatial filtering of binary classification, eigenvectors $\omega_1 \in \mathbb{R}^{C \times P}$ and $\omega_2 \in \mathbb{R}^{C \times P}$ of two classes are concatenated into the spatial filter $\mathbf{W} \in \mathbb{R}^{C \times 2P}$ used in bFBTRCA. In the K -class classification, the size of the spatial filter \mathbf{W} is $\in \mathbb{R}^{C \times KP}$, where K is the number of classes. In this case, the number of channels KP after spatial filtering is greater than the number of channels C of the original EEG signals. After having been filtered with $\mathbf{W} \in \mathbb{R}^{C \times KP}$, the EEG signals are not full-rank, and therefore contain more redundant information than the original EEG signals before spatial filtering. The framework of bFBTRCA is optimized to fit with the multi-class classification.

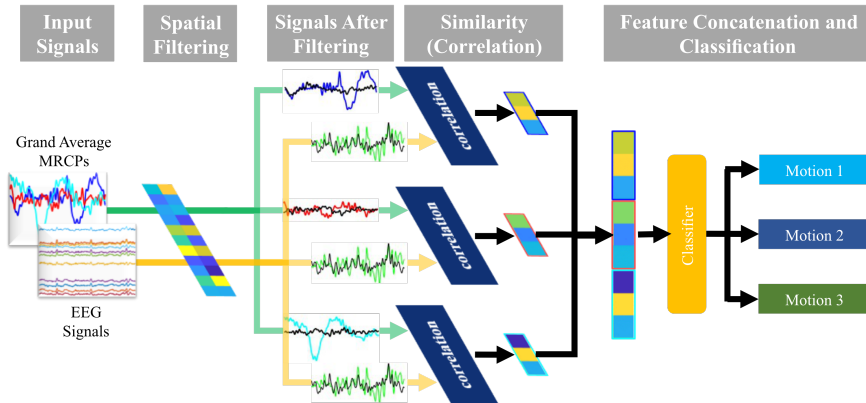


Fig. 3. Structure of the mSTRCA method for the three-class classification problem. This structure can be extended to a classification model for K classes, where $K \in \mathbb{Z}^+$. The optimization of the bSTRCA includes two main points. The first is the optimization of the spatial filter, which avoids the dimensional increase of EEG signals in multi-class classification. The second is the input of the correlation. The same components are removed from the signals after spatial filtering when measuring the similarity between the grand average MRCPs and the EEG signals, as given in Equation 16. The proposed mFBTRCA method incorporates the filter bank selection into mSTRCA. This procedure is given in Fig. 2.

This optimization includes two points: the spatial filter and the similarity measurement. After the optimization, the bSTRCA method can be used in the multi-class classification, which is the multi-class standard task-related component analysis (mSTRCA) method. The mFBTRCA method is developed by applying the filter bank selection to the mSTRCA method. The structure of the mSTRCA is shown in Fig. 3, using three-class classification as an example.

1) *Spatial Filtering*: The bSTRCA is a method developed based on the grand average MRCP. Before optimizing the frame of bSTRCA, it is necessary to clarify the relation between bSTRCA and the grand average MRCP. The grand average MRCP is the mean of EEG trials in the same class. There are three kinds of inputs involved in the calculation of the correlation coefficients in bSTRCA:

- (1) each of the EEG trials before averaging
- (2) the grand average MRCP of one class
- (3) the grand average MRCP of the other class.

The binary classification is based on the differences between the two grand average MRCPs. The features in bSTRCA use the similarities between each EEG trial as well as the two grand average MRCPs with correlation coefficients. The labels of the EEG trials can be predicted by their similarity. However, the noise in EEG signals are not eliminated by taking the mean of all trials. bSTRCA uses the task-related component analysis as a spatial filter to reject the task-unrelated components such as noise from the original EEG signals. Therefore, the spatial filter plays a main role in rejecting noise here, and is not related to discriminating the classes of EEG signals.

In bSTRCA, the eigenvectors of two classes are obtained by solving the eigen equation in Equation 3 and are then concatenated into the used spatial filter. However, it is not necessary to label the eigenvectors in the spatial filter, because the filter is used for noise rejection and task-related component extraction. Since the spatial filter is responding to the noise rejection and is not related to the classification, the spatial filter used in bSTRCA is modified such that it only removes the information about classes in the spatial filter.

The summed-up inter-trial covariance \mathbf{S}^k and the summed-up intra-trial covariance \mathbf{Q}^k are obtained through Equations 4 and 5. The spatial filter for the multi-class classification is found using the eigen equation

$$\max_{\omega} J = \frac{\omega^T \mathbf{S} \omega}{\omega^T \mathbf{Q} \omega}, \quad (15)$$

where $\mathbf{S} = \sum_{k=1}^K \mathbf{S}^k$ and $\mathbf{Q} = \sum_{k=1}^K \mathbf{Q}^k$. K is the number of classes in the multi-class classification.

2) *Similarity Measurement*: In binary classification, the performance is determined by the differences between the grand average MRCP of two motions. When two grand average MRCPs have large differences, this indicates that the classification accuracy of the two motions is higher than that with minor differences. To reduce the similarity between two grand average MRCPs, a possible approach is to remove the mean of the two from both grand average MRCPs. The differences between two grand average MRCPs are then maximized. In the multi-class task, the mean of grand average MRCPs of K motions are removed from the grand average MRCP $\hat{\mathbf{X}}^k$ and the input EEG signals \mathbf{X} in Equation 7, 9 and 12

$$\mathbf{X} \rightarrow \mathbf{X} - \frac{1}{K} \sum_{k=1}^K \hat{\mathbf{X}}^k; \hat{\mathbf{X}}^k \rightarrow \hat{\mathbf{X}}^k - \frac{1}{K} \sum_{k=1}^K \hat{\mathbf{X}}^k. \quad (16)$$

The canonical correlation pattern consists of three correlation coefficients for each class in multi-class classification. The first two correlation coefficients are the same as the ones given in Equations 8 and 11. The third one is given by Equation 14. However, to fit with the needs of multi-class classification, Equation 12 is replaced with

$$\mathbf{X}_* = \mathbf{X} - \hat{\mathbf{X}}^k; \mathbf{X}_k = \frac{1}{K-1} \sum_{kk=1, k \neq k}^K \hat{\mathbf{X}}^{kk} - \hat{\mathbf{X}}^k. \quad (17)$$

In Equation 12, \mathbf{X}_k is the distance between $\hat{\mathbf{X}}^1$ and $\hat{\mathbf{X}}^2$. The distances between \mathbf{X} and the grand average MRCPs, namely $\mathbf{X} - \hat{\mathbf{X}}^1$ and $\mathbf{X} - \hat{\mathbf{X}}^2$, are normalized by the correlation with $\hat{\mathbf{X}}^1$ and $\hat{\mathbf{X}}^2$. In the multi-class classification, $\hat{\mathbf{X}}^k$ is given in Equation 17 to normalize the distance \mathbf{X}_* .

In a K -class classification, there are K grand average MRCPs $\hat{\mathbf{X}}^k$. For each grand average MRCP, three correlation coefficients are calculated. Therefore, there are $3K$ coefficients in each filter bank. In the binary classification, $K = 2$ and there are six features in each filter bank.

3) *Filter Bank Selection*: In the filter bank selection of mFBTRCA, the same setting is used as the one presented in Section II-B3. EEG signals are divided into ten filter banks, and the minimum-redundancy maximal-relevance method is used to optimize the sequence of features and select the best features for classification. The selected features are classified using the multi-class support vector machine method.

D. Comparison Methods

We compare the proposed method to both the state-of-the-art methods and the baseline methods. The state-of-the-art methods refer to the methods that have been proposed and validated in previous researches. Because the neural network is a universal solution to data processing and has no specific model, we summarize previous neural network architectures on EEG processing. The baseline methods are the networks designed under the summarized architecture. The details about these methods are given in the supplementary materials.

1) *State-of-the-Art methods*: The following is a brief introduction to the compared state-of-the-art models on the multi-class classification of limb movements. All the given methods were implemented on the same datasets used in this paper, including mCSP+LDA [37], SpoC+Ridge [47], MDM [48], TSLDA [48], SCNN and DCNN [49], WaveNet [50], HopeFullNet [51].

2) *Baseline Methods*: As EEG signals are multi-channel time series, it is unavoidable to discuss their temporal characteristics. RNN is a universal solution to the feature extraction of time series. Along with the state-of-the-art methods presented above, our model is also compared with models that combine the RNN and CNN layers. In previous EEG signal analyses, CNN was used to extract features from EEG data; the extracted features were then processed by the RNN layers to extract the temporal features, and finally the fully connected layer was used as the classifier [52], [53]. The baseline methods follow these steps in previous analysis. Four models are compared in this work, including (1) C-R-CNN, (2) the convolutional-GRU-convolutional neural network (C-G-CNN), (3) convolutional-LSTM neural network (C-L-CNN) and (4) Graph C-G-CNN (GC-G-CNN).

3) *Parameter Settings*: The performance of the proposed and the compared methods are evaluated by 10-fold cross-validation. The compared methods contain various neural networks, including SCNN [49], DCNN [49], Wavenet [50], HopeFullNet [51] and the baseline methods. The compared neural networks have the same hyper-parameters, including batch size (50), learning rate (0.001) and training epochs (50). The loss function is the cross-entropy, and optimized by Adam optimizer. Both datasets are split based on the 10-fold cross-validation. The performance of all these methods is evaluated by the classification accuracy averaged from the 10 folds. All statistical analyses were conducted without correction for multiple comparisons.

III. RESULT

In this work, the mFBTRCA method is proposed to solve the multi-class classification problem of upper limb movements. Two datasets are used to evaluate and compare the proposed methods' performance against state-of-the-art and baseline methods. The results analysis consists of three parts: (1) the performance comparison between bFBTRCA and mFBTRCA in the binary classification task, (2) the evaluation of a three-class classification including two limb motions and the *resting* state, and (3) the multi-class classification performance evaluation. The first and second parts are evaluated and analyzed with EEG signals in dataset I. In the third part, both datasets are used to analyze the relationship between the classification accuracy and the grand average MRCP of each motion.

A. Structure Comparison

This work optimizes the spatial filter and similarity measurement of bFBTRCA such that the resulting mFBTRCA method can be used in multi-class classification. Before applying mFBTRCA to multi-class classification tasks, it is necessary to compare the performances of bFBTRCA and mFBTRCA in the binary classification task. Therefore, the bFBTRCA and mFBTRCA methods are applied to classify motion pairs in dataset I.

Fig. 4 gives the classification accuracies summarized from 10 folds of 15 subjects and 21 motion pairs. Fig. 6 shows the accuracy comparison of each motion pair. '1' refers to the case that removing the mean of grand average MRCPs (Equation 16) is not used; '2' refers to the case that the Equation 16 is used. In the spatial filter of both bFBTRCA and mFBTRCA, the number of selected eigenvectors P is 3 [27].

We also test the nested cross-validation when determining the selected eigenvectors P in mFBTRCA-Nested. Because the mFBTRCA consists of filter banks, the optimal P in each filter banks may be different. The mSTRCA is used to determine the P for each filter banks. The mFBTRCA-Nested shows a better performance than the other methods. In Fig. 5, the optimal P of Subject 1 in 0.5~10 Hz are visualized by a stacked column chart. In the chart, the optimal P may be different between two arbitrary motions. Therefore, compared to giving a fixed value, determining P by nested cross-validation is a better approach.

In Fig. 4, the use of Equation 16 increases the averaged accuracy about 1% when the maximum accuracy is reached, where the number of selected features is about 10. The nested cross-validation increases the averaged accuracy about 1%. To avoid further discussion on the hyper-parameters P and the number of selected features, we use the nested cross-validation to determine P and set the number of selected features to the maximum. In Table I, the averaged accuracies and p -values are given when the number of selected features is the maximum. The p -values is the results of two-sample t -test between the nested mFBTRCA and the others. The chance level is calculated with the dummy classifier with scikit-learn package in python. Both bFBTRCA and mFBTRCA methods achieve similar classification accuracies in the binary classification task.

TABLE I
ACCURACY AND p -VALUE COMPARISON IN BINARY CLASSIFICATION

	Accuracy	p -value
Chance Level	0.5197±0.0000	0.0619
bFBTRCA-1	0.7318±0.1242	0.5724
bFBTRCA-2	0.7342±0.1263	0.5898
mFBTRCA-1	0.7355±0.1233	0.5431
mFBTRCA-2	0.7376±0.1256	0.5845
mFBTRCA-Nested	0.7487±0.1250	-

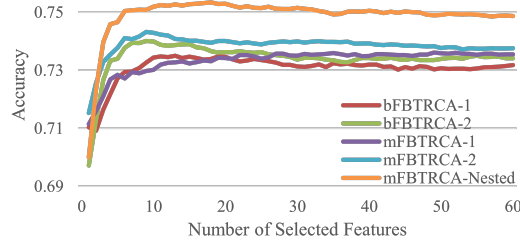


Fig. 4. Averaged accuracies across all subjects and folds. 'FBTRCA-1' denotes that the Equation 16 is not used; 'FBTRCA-2' denotes that the Equation 16 is used. In 'mFBTRCA-Nested', the hyper-parameter P is determined by nested cross-validation and the Equation 16 is used.

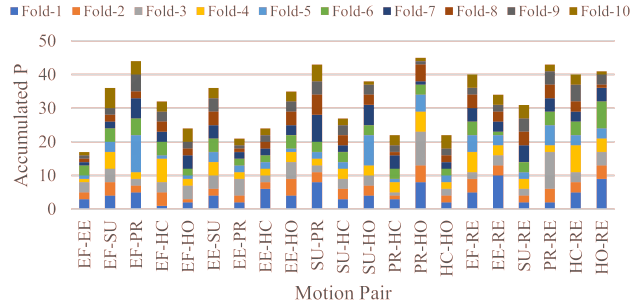


Fig. 5. Hyperparameter P values determined by nested cross-validation in the frequency range $0.5\sim 10$ Hz for Subject 1. The values of P in this figure are calculated for 10 folds and 21 pairs of movement, respectively. The hyperparameter values of P from the 10 folds are accumulated and stacked on the graph in different colors to be adequately visualized.

B. Three-class Comparison

This three-class comparison is carried out based on the classification accuracy between motion pairs of movement states and the *resting* state. In Fig. 7, the classification performances of the proposed mFBTRCA method and the state-of-the-art methods are given. Table II is the p -value between mFBTRCA and the other methods, calculated with two-sample t -test. The p -values between mFBTRCA and the chance level are almost zeros, so they are not presented in this table. As can be observed, the SCNN and DCNN methods have a comparable performance to mFBTRCA. However, the process going from EEG signals to the classification features in SCNN and DCNN is ambiguous due to the interpretability of neural networks; we know that the performance of the deep neural network is good, but do not know how the network utilizes the information in the EEG signals. For the proposed mFBTRCA method, on the other hand, it is clear how the MRCP signals are transformed into features. The method removes noise via the spatial filter and measures the similarity via the correlation coefficients.

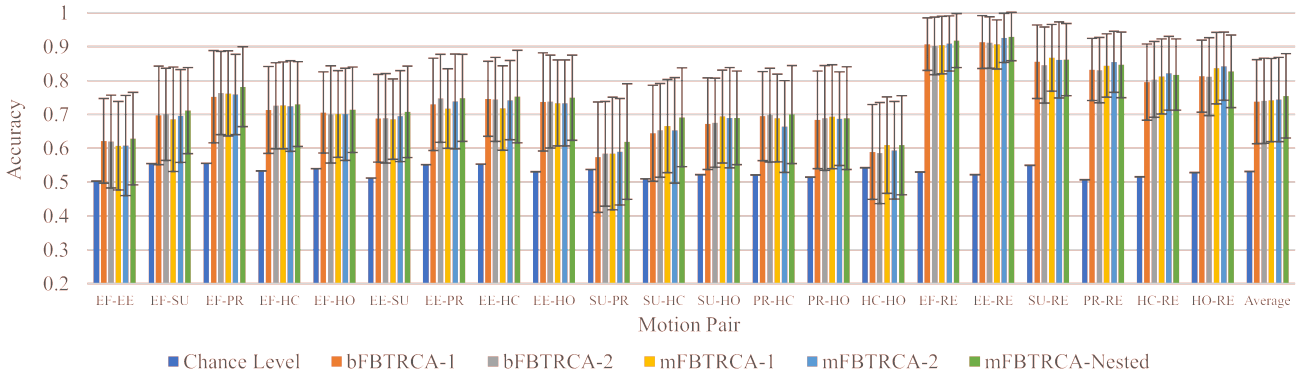


Fig. 6. Comparison of the classification accuracies between the bFBTRCA and mFBTRCA methods in the binary classification task. The abbreviations on the x-axis refer to the motion names; for instance, 'EE' is the abbreviation for *elbow extension*. The evaluation is based on the 21 motion pairs in dataset I. Accuracies are averaged across ten folds of 15 subjects. The mFBTRCA method performs similarly to bFBTRCA in binary classification.

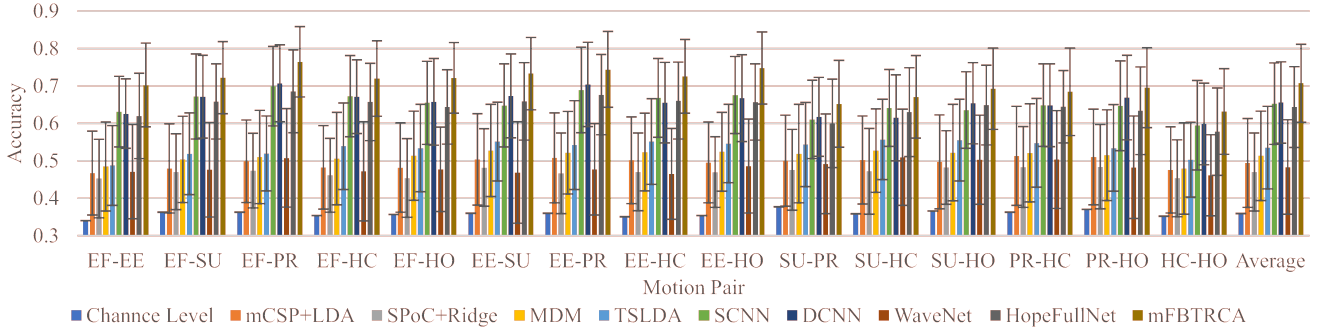


Fig. 7. Three-class classification performance comparison between the proposed mFBTRCA method and the state-of-the-art methods. The three classes in this figure are the motion pairs of limb movements and the *resting* state. The labels on the x-axis represent the name of the motion pairs, and the name for the *resting* state, 'RE', is ignored. The mFBTRCA method has a comparable performance to the SCNN and DCNN methods.

TABLE II
THE p -VALUE OF THREE-CLASS CLASSIFICATION

	EF-EE	EF-SU	EF-PR	EF-HC	EF-HO	EE-SU	EE-PR	EE-HC	EE-HO	SU-PR	SU-HC	SU-HO	PR-HC	PR-HO
mCSP+LDA	0.0143	0.0016	0.0006	0.0079	0.0216	0.0108	0.0030	0.0029	0.0013	0.1504	0.0639	0.0146	0.0735	0.0603
SPoC+Ridge	0.0043	0.0004	0.0031	0.0009	0.0008	0.0013	0.0008	0.0015	0.0045	0.0368	0.0268	0.0199	0.0053	0.0061
MDM	0.0251	0.0018	0.0018	0.0131	0.0069	0.0507	0.0097	0.0160	0.0029	0.0997	0.0871	0.0446	0.0862	0.0626
TSLDA	0.0132	0.0180	0.0192	0.0382	0.0547	0.0133	0.0294	0.0552	0.0663	0.1970	0.1630	0.1339	0.1203	0.0421
SCNN	0.2793	0.3313	0.3570	0.3449	0.3270	0.2106	0.3649	0.3188	0.2157	0.3584	0.3937	0.2610	0.4876	0.3795
DCNN	0.2186	0.3967	0.3110	0.4279	0.3026	0.3527	0.4888	0.2453	0.2080	0.3852	0.3401	0.3884	0.4794	0.5751
WaveNet	0.0218	0.0012	0.0127	0.0021	0.0125	0.0009	0.0015	0.0055	0.0167	0.0637	0.0941	0.0654	0.0586	0.0055
HopeFullNet	0.2924	0.2424	0.3313	0.2140	0.1394	0.2066	0.3284	0.2837	0.1709	0.4383	0.3383	0.4301	0.3918	0.3317

C. Multi-class Comparison

In the results analysis of the multi-class comparison, the performance of the proposed mFBTRCA method is compared to the state-of-the-art methods and the baseline methods with EEG signals from both datasets.

a) *Overall Comparison*: The overall performances of the mFBTRCA method and the compared models are summarized in Table III, where the accuracies are averaged across all subjects and folds of each dataset. In this table, we apply FBTRCA to 10 bands, which have the low cut-off of 0.5 Hz and the high cut-offs of an arithmetic sequence from 1 Hz to 10 Hz with step of 1 Hz . In dataset I, mFBTRCA improves on the classification accuracy of SCNN by 6.03% ($p = 0.1258$). Furthermore, in dataset II, mFBTRCA improves on the classification accuracy of GC-G-CNN by 8.73% ($p = 0.0736$).

TABLE III
COMPARISON WITH THE STATE-OF-THE-ART AND BASELINE METHODS. ALL THE METHODS ARE APPLIED ON THE SAME DATASETS (I AND II).

Method	Performance (Mean \pm Std (p -value))	
	Dataset I (7 Classes)	Dataset II (5 Classes)
Chance Level	-	0.1594 \pm 0.0000 (0.0000)
mCSP+LDA	[37]	0.2313 \pm 0.0551 (0.0037)
SPoC+Ridge	[47]	0.2207 \pm 0.0497 (0.0006)
MDM	[48]	0.2550 \pm 0.0700 (0.0032)
TSLDA	[48]	0.2712 \pm 0.0631 (0.0053)
SCNN	[49]	0.3590 \pm 0.0645 (0.1258)
DCNN	[49]	0.3529 \pm 0.0740 (0.1287)
WaveNet	[50]	0.2027 \pm 0.0569 (0.00005)
HopeFullNet	[51]	0.3377 \pm 0.0774 (0.0835)
C-R-CNN		0.3116 \pm 0.0705 (0.0374)
C-L-CNN		0.3100 \pm 0.0654 (0.0254)
C-G-CNN		0.3121 \pm 0.0696 (0.0184)
GC-G-CNN		0.3177 \pm 0.0745 (0.0464)
mFBTRCA		0.4193 \pm 0.0780

IV. DISCUSSION

The mFBTRCA method is developed by extending our previous bFBTRCA method, which is a binary classification model. However, there is a drawback involved when migrating from binary to multi-class classification for the bFBTRCA. The spatial filter in bFBTRCA is obtained by concatenating the eigenvectors of two classes. In the multi-class classification, the number of classes increases such that the number of eigenvectors in the spatial filters also increase. However, the number of eigenvectors

in the spatial filter should be smaller than the number of channels, as otherwise the spatial filtering will increase the dimension along the channel axis of the EEG signals while keeping the rank unchanged; in other words, it would introduce useless information into the EEG signals.

The bFBTRCA method consists of three modules: spatial filtering, similarity measuring and filter bank selection. In the migration from the bFBTRCA method to the mFBTRCA method, it is assumed that spatial filtering plays the role of noise rejection, while the similarity measuring determines the classes of the EEG signals. This assumption is based on the processing of the grand average MRCP. The grand average MRCP of a class is obtained by averaging the EEG signals of that class. In MRCP signals, the class of the motions can be discriminated by comparing the differences of the grand average MRCPs. The purpose of averaging EEG signals is to remove random noise from the original signals. Therefore, the process of using the grand average MRCP to discern between different motions has two steps, (1) removing irrelevant noise from EEG signals and (2) discriminating classes by comparing the grand average MRCPs. These two steps are spatial filtering and similarity measuring, respectively.

The bFBTRCA method is migrated to mFBTRCA with this assumption. The optimization has two steps: (1) remove the steps related to the class information in the spatial filtering, and (2) reduce the similarity of the grand average MRCPs. The first step involves calculating the spatial filter with the eigenvectors given in Equation 15, and the second step involves removing the mean of all grand average MRCPs from the spatial-filtered EEG signals in Equation 16.

When calculating the correlation between each of the trials and the grand average MRCPs, the two inputs are spatial-filtered by task-related component analysis. Fig. 8 is an example of EEG signals before and after spatial filtering. The signals are sorted in a descending sequence of eigenvalues after spatial filtering. When the index of eigenvalues is greater than three, the signals become flat. The signals in the first three channels show distinguished fluctuation to other flat signals. This is the reason to adjust the spatial filter when immigrating the binary to the multi-class classification. In MRCP analysis, P is determined to be around 3 when the movement onset can be located. In steady-state visual evoked potential, task-related component analysis is also used in the spatial filtering of the EnsembleTRCA method [14]. The spatial filters in mFBTRCA and the EnsembleTRCA are the same. The P in steady-state visual evoked potential is set to 1 because of the high signal-to-noise ratio of steady-state visual evoked potential. However, the setting in MRCP signals is different because of the low signal-to-noise ratio. $P = 1$ cannot cover all the task-related signals. This is the reason why we don't concatenate the spatial filters of different classes as in steady-state visual evoked potential.

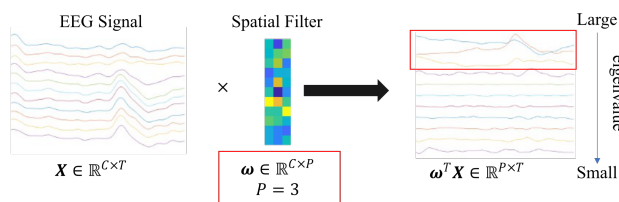


Fig. 8. EEG signals before and after spatial filtering with task-related component analysis.

Although task-related component analysis may not be the most efficient tool to remove the noises, we compared it to some spatial filters in our previous work. It shows the best performance among the compared spatial filters [26]. Discriminative canonical pattern matching is the spatial filter in both steady-state visual evoked potential and MRCP [12], [43], [54]. Discriminative canonical pattern matching aims to find a projection to maximize the differences of two classes. Based on the results in [26], the performance of discriminative canonical pattern matching is much worse than task-related component analysis. Therefore, we concluded that maximizing class differences by spatially filtering should be avoided.

In the calculation of the correlation coefficients, we remove the mean of the grand average MRCPs from both EEG trials and the grand average MRCPs in Equation 16. In mFBTRCA, the classification performance highly relies on the differences of these grand average MRCPs. A possible approach to increase the differences is to remove the common components of these grand average MRCPs. The simplest common component is the mean of these grand average MRCPs. Fig. 9 shows the grand average MRCPs in channel C_z before and after removing the mean of the grand average MRCPs. We then calculate the correlation coefficients of the grand average MRCPs (C_z) of motion pairs. In this figure, it shows that the grand average MRCPs of the movement states are of high similarity. After removing the mean of these grand average MRCP, the correlations between motion pairs are reduced.

Now we can conclude the differences between the proposed mFBTRCA and the EnsembleTRCA. Despite of the same spatial filter in two methods, the differences between mFBTRCA and EnsembleTRCA consist of three points.

(1) The filter banks. The filter bank in MRCP signals is located at the low-frequency bands. In mFBTRCA, we first divide the signals into subbands in the low-frequency bands based on our previous work [27]. In EnsembleTRCA, the filter banks are not used.

(2) Added-up spatial filters instead of concatenation. As mentioned above, limited to the low signal-noise ratio of MRCP signals, mFBTRCA adds up the covariances of all the class and then solve the eigen-equation to get the spatial filter. In

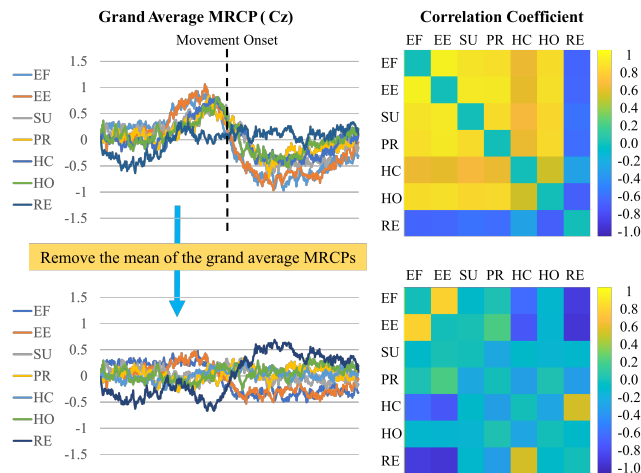


Fig. 9. The grand average MRCPs before and after removing the mean of the grand average MRCPs, and the correlation coefficients of the grand average MRCPs of the motion pairs. The time window of the EEG signals is located between one second before and after the movement onset.

EnsembleTRCA, the eigen-equations are solved for the covariances of each class respectively, and then the eigenvectors of all the classes are concatenated as the spatial filter.

(3) Optimized correlation features. In both mFBTRCA and EnsembleTRCA, the averaged signals across trials of each class ($\hat{\mathbf{X}}^k$ in Equation 6) are used as the templates to calculate the correlation. In EnsembleTRCA, a template is averaged of signals with a given frequency. The templates of EnsembleTRCA are of low similarity because they are located at different frequencies. However, in mFBTRCA, the templates are the same with the grand average MRCPs and are of high similarity. Therefore, mFBTRCA removes the mean of these templates before calculating the correlation.

The mFBTRCA method achieves an equivalent classification performance to the bFBTRCA method in the binary classification task, as shown in Table I. Because the binary classification between a movement state and the *resting* state (e.g., *elbow flexion* vs *resting*) has a higher classification accuracy, we assume that the classification performance between the motion pairs of the movement states will not fluctuate significantly. mFBTRCA is compared to state-of-the-art methods in classifying pairs of movement states and the *resting* state (3 classes). The three-class classification accuracies are close to those achieved in the binary classification, but with slight decreases. For instance, the classification accuracy achieved between *elbow flexion*, *pronation* and *resting* is close to the one achieved between *elbow flexion* and *pronation*.

The state-of-the-art methods used in this work include machine learning-based and deep learning-based methods. The compared machine learning-based methods are geared towards the multi-class classification of limb movements, such as those made by the left/right hand, the foot or the tongue. The classification of these motions is based on certain brain activity, such as motor imagery. The deep learning technique is a universal solution to classification. Although the deep learning methods are not specific to the classification of limb movements, they have a better performance than the machine learning methods. The proposed mFBTRCA method is also based on machine learning. However, mFBTRCA takes advantage of the differences of grand average MRCPs of motions. These differences are the reason why different limb movements can be classified. As a result, mFBTRCA performs better than the deep learning methods in the multi-class classification.

Besides the state-of-the-art methods, baseline methods were designed based on deep learning to be used for comparison purposes. In the design of these, we used the same idea as when designing the mFBTRCA method, namely using spatial filtering (spatial) and similarity measuring (temporal). The CNN layers were used to optimize the spatial characteristic of EEG signals. The RNN layers were then used to extract the temporal features. Finally, the CNN layer followed by a fully connected layer was the classifier used to predict the classes of EEG signals. We compared the mFBTRCA to the state-of-the-art and baseline methods in the multi-class classification task (more than three classes). As given in Table I, mFBTRCA shows an improved performance over the other compared methods, including the baseline methods.

Although the proposed mFBTRCA method improves on the performance of the state-of-the-art and baseline methods, it also has its bottleneck. The method classifies the EEG signals based on the differences in the grand average MRCPs of motions. Importantly, when the grand average MRCPs of motions are correlated or almost the same, mFBTRCA fails to classify these motions, such as *elbow flexion* and *elbow extension*. Furthermore, the movement onset localization is a problem for armless or paralyzed patients in the actual application of the brain-computer interface. Our future work will focus on fusing the proposed mFBTRCA method with the deep learning techniques and applying transfer learning to help with the localization of the movement onset.

Supplementary Materials

I. ACRONYMS

TABLE I
ACRONYMS AND THEIR CORRESPONDING FULL NAMES

Acronym	Full Name
Concept	
EEG	Electroencephalogram
MRCP	Movement-Related Component Analysis
Limb State	
EF	Elbow Flexion
EE	Elbow Extension
SU	Supination
PR	Pronation
HC	Hand Close
HO	Hand Open
RE	Resting
PG	Palmar Grasp
LG	Lateral Grasp
Machine Learning Method	
CSP	Common Spatial Pattern
mCSP	multi-class Common Spatial Pattern
FBCSP	Filter Bank Common Spatial Pattern
bSTRCA	binary Standard Task-Related Component Analysis
mSTRCA	multi-class Standard Task-Related Component Analysis
bFBTRCA	binary Filter Bank Task-Related Component Analysis
mFBTRCA	multi-class Filter Bank Task-Related Component Analysis
SPoC	Source Power Comodulation
MDM	Minimum Distance to Mean
LDA	Linear Discriminant Analysis
TSLDA	Tangent Space Linear Discriminant Analysis
Deep Learning Method	
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
GNN	Graph Neural Network
LSTM	Long Short-Term Memory
GRU	Gate Recurrent Unit
SCNN	Shallow Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
C-R-CNN	Convolutional-RNN-Convolutional Neural Network
C-L-CNN	Convolutional-LSTM-Convolutional Neural Network
C-G-CNN	Convolutional-GRU-Convolutional Neural Network
GC-G-CNN	Graph Convolutional-GRU-Convolutional Neural Network

II. RELATED WORKS

In this section, we will first introduce previous multi-class classification methods related to limb movements, including machine learning and deep learning methods. The binary classification methods based on MRCP signals are then presented, which are also related to limb movements.

Amongst brain activities that are related to limb movements, motor imagery is frequently used in brain-computer interfaces. The multi-class classification algorithms of limb movements have mostly been developed based on motor imagery in previous works. Motor imagery is related to the power change of EEG signals between 8 Hz and 30 Hz. When the left/right limb moves, the power of the EEG signals in some channels increases or decreases.

Common spatial pattern (CSP) + Linear Discriminant Analysis (LDA) is the basic binary classification algorithm used

for motor imagery [1]. It searches for the spatial filter CSP that maximizes and minimizes the self-covariance of EEG signals, and it then converts the filtered signals to the logarithm variances. The variances are the features used for classification by the LDA classifier. The filter bank CSP computes and combines multiple CSPs in various frequency bands, and selects features from these bands by feature selection methods based on mutual information. The filter bank CSP method has demonstrated its competitive performance in several competitions [2]. However, the original CSP+LDA and filter bank CSP algorithms can only be used in binary classification. The multi-class version of CSP+LDA puts the common spatial pattern in the framework of information theoretic feature extraction [3], or uses the one-versus-rest strategy in the spatial filter [4]. Therefore, the common spatial pattern method can classify multiple limb movements such as those made by the left/right hand, a foot or the tongue.

In the CSP+LDA algorithm, the spatial filtering and the logarithm covariance features can be regarded as the computation of a Riemann distance in the space of covariance matrices in the context of brain-computer interfaces [5]. The covariance matrices are used as EEG signal descriptors. In the Riemannian geometry, these matrices are classified directly using the topology of the manifold of symmetric and positive definite matrices. The computation on the Riemannian manifold facilitates to discern between the multiple classes of limb movements. The minimum distance to mean (MDM) is the straightforward algorithm relying on the Riemannian manifold and the Riemannian distance [6]. This algorithm regards the covariances of EEG signals as points on the manifolds. The center point of the points belonging to the same class is computed with the Riemannian mean of covariances. The classes of points are predicted by measuring the Riemannian distance between points and the center points. The tangent space linear discriminant analysis classifier is an optimized algorithm based on the Riemannian manifold [6]. It maps a set of covariance matrices to the Riemannian tangent space, and the dimensionally-reduced matrices are classified by a linear discriminant classifier.

Following the development of deep learning techniques, neural networks have proved to be very useful in EEG signal processing due to their competitive classification performances. During EEG acquisition, signals are sparsely sampled from several electrodes on the scalp, such that EEG signals have low-grade spatial characteristics. The convolutional neural network (CNN) has shown its efficiency in extracting and optimizing spatial features. The shallow CNN (SCNN) and the deep CNN (DCNN) are two CNN architectures used for the end-to-end EEG analysis [7]. The DCNN has been shown to perform at least as well in binary classification as the widely used filter bank CSP algorithm. Because both the SCNN and DCNN architectures are based on neural networks, and the

number of output neurons can be enlarged as is convenient, both can be used in multi-class classification. As EEG signals are multi-channel time series, the temporal dynamic processes will also be considered in signal processing. The recurrent neural network (RNN) can use the temporal characteristics of EEG signals, especially when equipped with long short-term memory units [8]. In some studies, CNN and RNN were used simultaneously in EEG signal processing to fuse the spatial and temporal features [9], [10].

Neural networks can capture the spatial and temporal characteristics of EEG signals. However, the transformation process going from EEG signals to features is obscure. Deep learning in EEG signal classification has achieved impressive performances. The main reasons are (1) the development of deep learning in pattern recognition, and (2) a deeper understanding of the traditional machine learning based algorithms. In motor imagery, CSP-based algorithms have three main steps carried out during feature extraction. Firstly, the spatial filter of CSP+LDA selects the most discriminant channels and optimizes the spatial characteristics. FBCSP then divides EEG signals into several filter banks and applies the CSP algorithm in each band, thus optimizing the features in different filter banks. Finally, multiple CSPs can also be applied to the different sliding time windows of EEG signals, and can then select CSP features of these windows [11]. The EEG signal processing procedure can thus be divided into spatial optimization, filter bank selection and time window selection steps. Different network architectures can be interpreted from the three points when applying deep learning techniques to EEG signals. For example, RNN uses the temporal characteristic [12]–[15] while the graph neural network (GNN) optimizes the spatial characteristic of the EEG signals [16]–[20].

However, the machine learning based algorithms cannot fully interpret the information in MRCP signals. Compared to the left/right limb movements in motor imagery, MRCP is related to the motions of the limb. The MRCP signals are located at the low-frequency bands of the EEG signals, namely 0.05~10 Hz. As a result, the noise or task-unrelated components in other bands have to be removed in the signal processing. The grand average MRCP is the approach used to visualize the MRCP signals. It removes the noises by taking the average of multiple EEG trials belonging to the same class.

Based on the grand average MRCP, Niazi *et al.* proposed a matched filter method to solve the binary classification task [21]. This method uses a spatial filter to maximize the MRCP energy and minimize the noise energy by optimizing the signal-to-noise ratio of the signal power. After the spatial filtering, it uses the likelihood ratio to match the relationship between the grand average MRCP and the signals before averaging, thus detecting the movement execution.

The manifold-learning method was introduced to MRCP processing by Xu *et al.* and showed improvements compared to the matched filter method [22]. This method projects the signals into the manifold by using locality-preserving projections. In the manifold space, these multi-channel signals are regarded as points. The grand average MRCP is located at the center of the trials belonging to the same class. These EEG signals are classified in the manifold space using linear discriminant anal-

ysis. Lin *et al.* optimized the manifold method proposed by Xu *et al.* by constructing the within-class graph and the between-class graph when projecting EEG signals onto the manifold [23]. A nearest-neighbour classifier is used to measure the distances between the grand average MRCP point and other points, thus predicting their labels. The manifold projection in the two methods and the spatial filtering are in fact used to find a matrix to reduce the dimension of the original EEG signals through matrix multiplication. The procedure of the traditional CSP-based method consists of spatial filtering and feature extraction. From this perspective, the manifold-based methods lack the step of feature extraction, because the signals are classified by comparing the distances after the manifold projection. Instead of measuring the differences with Riemannian distances, Chu *et al.* regarded the distances as the features and used the partial least squares regression to reduce the dimensions of the features [24].

The above methods analyze the signals from sensors. To optimize the spatial characteristics of signals, spatial filtering is used to optimize the spatial distributions. However, the sensor-based EEG signals have initial drawback of poor spatial resolution. Source imaging converts the sensor-based signals into the source-based signals so that the brain activity can be identified with high spatial resolution [25].

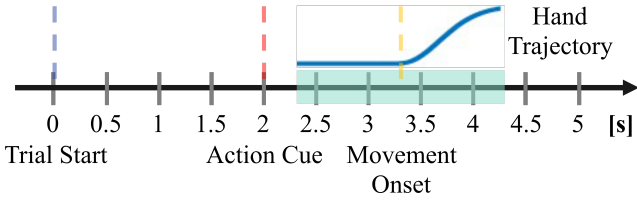
III. LOCALIZATION OF MOVEMENT ONSET IN DATASET I

The relationship between the hand trajectory, action cue and movement onset is given in Fig.1(a). The 1-order difference of the hand trajectory was filtered by a 1-order *Savitzky-Golay* finite impulse response smoothing filter. The time window length in the smoothing filter was set to 31. The hand trajectory was then normalized by dividing by the maximal absolute value. In *elbow flexion* and *elbow extension*, the hand trajectory has a higher amplitude when the limb moves. The location where the normalized trajectory equals the threshold of 0.05 was the movement onset. Trials were manually removed if the movement onset could not be located because of noise contamination. In the *resting* state, a fake movement onset was set to 0.5 s after the cue appeared on the screen. Trials in the *resting* state were rejected if the variances of normalized trajectory were larger than 0.02. In the other four motions, the function $f(x) = a * \exp(-(\frac{x-b}{c})^2) + d$ was used to fit the smoothed and normalized trajectory by tuning the parameters a, b, c, d . Trials were then rejected if the parameters of the tuned function met at least one of the following conditions: $a < 0.05$, $c > 100$ or $d > 10$. The movement onset was set to the time point whose absolute amplitude equalled 0.1. The original movement trajectories around the located movement onset are shown in Fig. 1(b). The average number of trials of each motion across subjects are given in Table II. The detailed processing steps are given in our code repository.

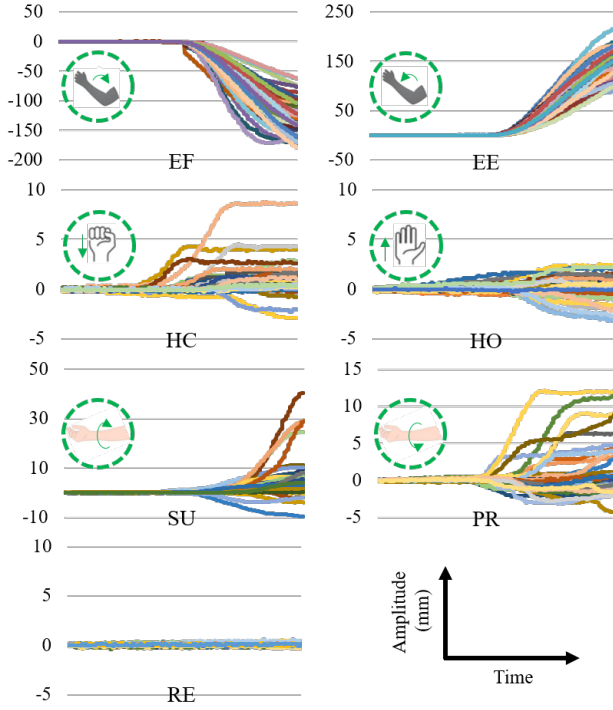
IV. THE BASELINE METHODS

The baseline method, convolutional-RNN-convolutional neural network (C-R-CNN), follows the steps in the previous analysis and includes three modules:

Module 1: CNN layers used to optimize spatial characteristics



(a) Relationship between action cue, movement onset and hand trajectory. The EEG signals for classification tasks are in a two-second time window.



(b) Motions and the corresponding hand trajectories in each trial, which is located at the two-second window in Fig.1(a).

Fig. 1. Localization of the movement onset and the corresponding movement trajectory around.

TABLE II
NUMBER OF TRIALS AFTER TRIAL REJECTION IN DATASET I

Motion	EF	EE	SU	PR	HC	HO	RE
Number	60	59	52	51	56	55	59

Module 2: RNN layers used to capture temporal characteristics
 Module 3: CNN layers followed by a fully connected classifier.
 The model structure is given in Table IV, where the batch size is B and the number of classes is K . The sequence of the output size is $[batch, channel, high, width]$.

The RNN layer in Table IV can also be replaced by either the gate recurrent unit (GRU) layer or the long short-term memory (LSTM) layer to achieve an improved performance. Therefore, we have three baseline methods: (1) C-R-CNN, (2) the convolutional-GRU-convolutional neural network (C-G-CNN) and (3) convolutional-LSTM neural network (C-L-CNN). The models of C-G-CNN and C-L-CNN correspond to the model that replaces the RNN layer of C-R-CNN with

TABLE III
THE MODEL STRUCTURE OF THE BASELINE METHOD

Module	Layer	Output Size
Input		$[B, 1, C, T]$
Module 1	ZeroPad2d	$[B, 1, C, T+31]$
	Conv2d	$[B, 8, C, T]$
	BatchNorm2d	$[B, 8, C, T]$
	LeakyReLU	$[B, 8, C, T]$
	Conv2d	$[B, 16, 1, T]$
	BatchNorm2d	$[B, 16, 1, T]$
	LeakyReLU	$[B, 16, 1, T]$
Module 2	Permutation	$[B, T, 16]$
	RNN	$[B, 1, T, 32]$
Module 3	Conv2d	$[B, 64, T//256, 1]$
	BatchNorm2d	$[B, 64, T//256, 1]$
	Flatten	$[B, 64*T//256]$
	Linear	$[B, 32]$
	LeakyReLU	$[B, 32]$
	Linear	$[B, 32]$

the GRU and LSTM layers, respectively. The three baseline methods share the same network architecture.

GNN is the other approach to optimizing the spatial characteristics of EEG signals. The channels of the signals are regarded as nodes of the graph. GNN optimizes the spatial quality of EEG signals by learning about the relation between these nodes with graph knowledge. The latent correlation layer in StemGNN is used to optimize the EEG signals in the C-G-CNN model [26]. The attention mechanism in the latent correlation layer is used to learn the latent correlations between the multi-channel time series. The learned attention graph is passed to the 4-order Chebyshev polynomial; thus, the output graph is of size $[4, C, C]$. The input EEG signals of size $[B, 1, C, T]$ are weighted with the learned graph. Therefore, the size of the latent correlation layer is $[B, 4, C, T]$. The EEG signals optimized with the latent correlation layer are then passed to the C-G-CNN model, with the *channel* changed to 4. This baseline model is referred to as Graph C-G-CNN (GC-G-CNN) in the following sections.

TABLE IV
THE MODEL STRUCTURE OF THE BASELINE METHOD

Module	Layer	Output Size
Input		$[B, 1, C, T]$
Module 1	ZeroPad2d	$[B, 1, C, T+31]$
	Conv2d	$[B, 8, C, T]$
	BatchNorm2d	$[B, 8, C, T]$
	LeakyReLU	$[B, 8, C, T]$
	Conv2d	$[B, 16, 1, T]$
	BatchNorm2d	$[B, 16, 1, T]$
	LeakyReLU	$[B, 16, 1, T]$
Module 2	Permutation	$[B, T, 16]$
	RNN	$[B, 1, T, 32]$
Module 3	Conv2d	$[B, 64, T//256, 1]$
	BatchNorm2d	$[B, 64, T//256, 1]$
	Flatten	$[B, 64*T//256]$
	Linear	$[B, 32]$
	LeakyReLU	$[B, 32]$
	Linear	$[B, 32]$

REFERENCES

- [1] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-robert Muller. Optimizing Spatial filters for Robust EEG

- Single-Trial Analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008.
- [2] Kai Keng Ang, Zhang Yang Chin, Haihong Zhang, and Cuntai Guan. Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 2390–2397. IEEE, 2008. event-place: Hong Kong, China.
- [3] Moritz Grosse-Wentrup and Martin Buss. Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction. *IEEE Transactions on Biomedical Engineering*, 55(8):1991–2000, 2008.
- [4] Weibo Yi, Shuang Qiu, Hongzhi Qi, Lixin Zhang, Baikun Wan, and Dong Ming. EEG feature comparison and classification of simple and compound limb motor imagery. *Journal of NeuroEngineering and Rehabilitation*, 10(1):106, 2013.
- [5] Alexandre Barachant, Stéphane Bon, Marco Congedo, and Christian Jutten. Common Spatial Pattern revisited by Riemannian geometry. In *2010 IEEE International Workshop on Multimedia Signal Processing*, pages 472–476. Saint-Malo, France, 2010. IEEE.
- [6] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten. Multiclass Brain-Computer Interface Classification by Riemannian Geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2012.
- [7] Robin Tibor Schirrmeyer, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for EEG decoding and visualization: Convolutional Neural Networks in EEG Analysis. *Human Brain Mapping*, 38(11):5391–5420, 2017.
- [8] Francisco Ordóñez and Daniel Roggen. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors*, 16(1):115, 2016.
- [9] Fengjie Wu, Weijian Mai, Yisheng Tang, Qingkun Liu, Jiangtao Chen, and Ziqian Guo. Learning Spatial-Spectral-Temporal EEG Representations with Deep Attentive-Recurrent-Convolutional Neural Networks for Pain Intensity Assessment. *Neuroscience*, 481:144–155, 2022.
- [10] Dalin Zhang, Lina Yao, Kaixuan Chen, Sen Wang, Xiaojun Chang, and Yunhao Liu. Making Sense of Spatio-Temporal Preserving Representations for EEG-Based Human Intention Recognition. *IEEE Transactions on Cybernetics*, 50(7):3033–3044, 2020.
- [11] Yu Zhang, Chang S. Nam, Guoxu Zhou, Jing Jin, Xingyu Wang, and Andrzej Cichocki. Temporally Constrained Sparse Group Spatial Patterns for Motor Imagery BCI. *IEEE Transactions on Cybernetics*, 49(9):3322–3332, 2019.
- [12] Qingguo Ma, Manlin Wang, Linfeng Hu, Linanzi Zhang, and Zhongling Hua. A Novel Recurrent Neural Network to Classify EEG Signals for Customers’ Decision-Making Behavior Prediction in Brand Extension Scenario. *Frontiers in Human Neuroscience*, 15:610890, 2021.
- [13] Gautam Krishna, Co Tran, Yan Han, Mason Carnahan, and Ahmed H Tewfik. Speech Synthesis Using EEG. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1235–1238. IEEE, 2020. event-place: Barcelona, Spain.
- [14] Gautam Krishna, Yan Han, Co Tran, Mason Carnahan, and Ahmed H. Tewfik. State-of-the-art Speech Recognition using EEG and Towards Decoding of Speech Spectrum From EEG. [arXiv:1908.05743 \[cs, eess\]](https://arxiv.org/abs/1908.05743), 2020. [arXiv: 1908.05743](https://arxiv.org/abs/1908.05743).
- [15] Gautam Krishna, Co Tran, Mason Carnahan, and Ahmed H. Tewfik. Advancing Speech Recognition With No Speech Or With Noisy Speech. [arXiv:1906.08871 \[cs, eess, stat\]](https://arxiv.org/abs/1906.08871), 2020. [arXiv: 1906.08871](https://arxiv.org/abs/1906.08871).
- [16] Priyanka Mathur and Vijay Kumar Chakka. Graph Signal Processing of EEG signals for Detection of Epilepsy. In *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*, pages 839–843. IEEE, 2020. event-place: Noida, India.
- [17] Shanzhi Xu, Hai Hu, Linhong Ji, and Peng Wang. An Adaptive Graph Spectral Analysis Method for Feature Extraction of an EEG Signal. *IEEE Sensors Journal*, 19(5):1884–1896, 2019.
- [18] Mirfarid Musavian Ghazani and Anh Huy Phan. Graph Convolutional Neural Networks for analysis of EEG signals, BCI application. [arXiv:2006.14540 \[eess\]](https://arxiv.org/abs/2006.14540), 2020. [arXiv: 2006.14540](https://arxiv.org/abs/2006.14540).
- [19] Yimin Hou, Shuyue Jia, Shu Zhang, Xiangmin Lun, Yan Shi, Yang Li, Hanrui Yang, Rui Zeng, and Jinglei Lv. Deep Feature Mining via Attention-based BiLSTM-GCN for Human Motor Imagery Recognition. [arXiv:2005.00777 \[cs, eess\]](https://arxiv.org/abs/2005.00777), 2020. [arXiv: 2005.00777](https://arxiv.org/abs/2005.00777).
- [20] D. Zhang, K. Chen, D. Jian, and L. Yao. Motor Imagery Classification via Temporal Attention Cues of Graph Embedded EEG Signals. *IEEE Journal of Biomedical and Health Informatics*, 24(9):2570–2579, 2020.
- [21] Imran Khan Niazi, Ning Jiang, Olivier Tiberghien, Jørgen Feldbæk Nielsen, Kim Dremstrup, and Dario Farina. Detection of Movement Intention from Single-Trial Movement-Related Cortical Potentials. *Journal of Neural Engineering*, 8(6):066009, 2011.
- [22] Ren Xu, Ning Jiang, Chuang Lin, Natalie Mrachacz-Kersting, Kim Dremstrup, and Dario Farina. Enhanced Low-Latency Detection of Motor Intention From EEG for Closed-Loop Brain-Computer Interface Applications. *IEEE Transactions on Biomedical Engineering*, 61(2):288–296, 2014.
- [23] Chuang Lin, Bing-Hui Wang, Ning Jiang, Ren Xu, Natalie Mrachacz-Kersting, and Dario Farina. Discriminative Manifold Learning Based Detection of Movement-Related Cortical Potentials. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(9):921–927, 2016.
- [24] Yaqi Chu, Xingang Zhao, Yijun Zou, Weiliang Xu, Guoli Song, Jianda Han, and Yiwen Zhao. Decoding multiclass motor imagery EEG from the same upper limb by combining Riemannian geometry features and partial least squares regression. *Journal of Neural Engineering*, 17(4):046029, 2020.
- [25] Bradley J. Edelman, Bryan Baxter, and Bin He. EEG Source Imaging Enhances the Decoding of Complex Right-Hand Motor Imagery Tasks. *IEEE Transactions on Biomedical Engineering*, 63(1):4–14, 2016.
- [26] Defu Cao, Jiachen Li, Hengbo Ma, and Masayoshi Tomizuka. Spectral Temporal Graph Neural Network for Trajectory Prediction. *CoRR*, abs/2106.02930, 2021.

3.5. Enabling Temporal-Spectral Decoding in Multi-class Single-side Upper Limb Classification

Enabling Temporal-Spectral Decoding in Multi-class Single-side Upper Limb Classification

Hao Jia^a, Shuning Han^a, Cesar F. Caiafa^b, Feng Duan^c, Yu Zhang^d, Zhe Sun^e, Jordi Solé-Casals^{a,f}

^a*Data and Signal Processing Research Group, Department of Engineering, University of Vic-Central University of Catalonia, Vic, 08500, Spain*

^b*Instituto Argentino de Radioastronomía, CONICET CCT La Plata/CIC-PBA/UNLP, V. Elisa, 1894, Argentina*

^c*Tianjin Key Laboratory of Brain Science and Intelligent Rehabilitation, College of Artificial Intelligence, Nankai University, Tianjin, 300110, China*

^d*Department of Bioengineering and the Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA 18015, USA*

^e*Faculty of Health Data Science, Juntendo University, UrayasuChiba, -, Japan*

^f*Department of Psychiatry, University of Cambridge, Cambridge, -, United Kingdom*

Abstract

This manuscript presents a novel approach for decoding pre-movement patterns from brain signals using a two-stage-training temporal-spectral neural network (TTSNet). The TTSNet employs a combination of filter bank task-related component analysis (FBTRCA) and convolutional neural network (CNN) techniques to enhance the classification of single-upper limb movements in non-invasive brain-computer interfaces.

In our previous work, we introduced the FBTRCA method which utilized filter banks and spatial filters to handle spectral and spatial information, respectively. However, we observed limitations in the temporal decoding phase, where correlation features failed to effectively utilize temporal information. To address this issue, our proposed method focuses on analyzing multi-channel signals in the temporal-spectral domain. The TTSNet first divides the signals into various filter banks, employing task-related component analysis to reduce dimensionality and eliminate noise. Subsequently, a CNN is employed to optimize the temporal characteristics of the signals and ex-

Email addresses: duanf@nankai.edu.cn (Feng Duan), z.sun.kc@juntendo.ac.jp (Zhe Sun), jordi.sole@uvic.cat (Jordi Solé-Casals)

tract class-related features. Finally, the class-related features from all filter banks are concatenated and classified using the fully connected layer.

To evaluate the effectiveness of our proposed method, we conducted experiments on two publicly available datasets. In binary classification tasks, the TTSNet achieved an improved accuracy of 0.7707 ± 0.1168 , surpassing the performance of EEGNet (accuracy: 0.7340 ± 0.1246) and FBTRCA (accuracy: 0.7487 ± 0.1250). In multi-class tasks, TTSNet achieved an accuracy of 0.4588 ± 0.0724 , exhibiting a 4.27% and 3.95% accuracy increase over EEGNet and FBTRCA, respectively.

The findings of this study suggest that the proposed TTSNet method holds promise for detecting limb movements and assisting in the rehabilitation of stroke patients.

Keywords: Brain-computer Interface, Electroencephalogram, Movement-related Cortical Potential, Single-side Upper Limb Movement, Pattern Recognition

1. Introduction

Non-invasive brain-computer interfaces bridge the gap between the human brain and external devices such as computers and robots [1, 2, 3]. They enable the conversion of brain activities into control commands by analyzing the electroencephalogram (EEG) information. The EEG signals are acquired non-invasively from the scalp using specialized acquisition devices.

Movement-related cortical potential (MRCP) is a brain activity associated with single-side upper limb movement [4, 5, 6, 7]. When a subject's limb moves, EEG signals acquired from the motor cortex exhibit an increase in amplitude in the low-frequency domain [5, 6]. Due to noise influences during signal acquisition, the EEG signals are averaged across multiple trials of repeated motions to reduce the noise impact. This averaged signal is referred to as the grand average MRCP [8]. Distinct differences in the grand average MRCPs are observed before and after movement onset when different upper limb motions, such as hand closure and elbow flexion, are executed [5]. These differences in grand average MRCPs serve as discriminative features for classifying the upper limb movements of subjects.

The previous studies on MRCP signals have two tasks: movement detection and movement classification [9]. Movement detection aims to distinguish between movement and resting states, representing a binary classification

task. In contrast, movement classification involves binary or multi-class classification of different movement states. During movement states, the grand average MRCP exhibits fluctuations characterized by an initial increase followed by a subsequent decrease around the movement onset. In contrast, the grand average MRCP during the resting state remains relatively stable in comparison to the movement state [5]. It is worth noting that movement detection, which classifies signals as fluctuating or steady, generally achieves higher classification accuracy compared to movement classification, as it represents a subset of the broader movement classification task.

The movement detection problem is addressed by the subject-dependent and section-wise spectral filtering method, which considers MRCP signals in two different temporal sections [10]. These sections consist of a two-second time window preceding the movement onset and a one-second time window following the movement onset. The signals within each section are averaged, and the mean amplitudes in both sections are used as features, which are then inputted into a regularized linear discriminant analysis classifier. This method utilizes the changes in amplitude of MRCP signals before and after the movement onset. However, the signal averaging approach overlooks the temporal dynamics of amplitude changes, resulting in potential information loss during signal processing.

EEG source imaging is another valuable contribution to the classification of single-side limb movement [11]. Instead of analyzing brain images directly, this technique initially identifies the region of interest. Subsequently, the EEG signals within the identified region of interest are analyzed in the time-frequency domain.

In our previous work, we proposed the standard task-related component analysis method (STRCA) to address the movement detection problem [12]. This method utilizes task-related component analysis as a spatial filter to remove noise and task-unrelated components from EEG signals. The spatial filtering process can also be viewed as a means to identify the regions of interest. After applying spatial filtering to both the unlabelled signals and the grand average MRCPs, canonical correlation coefficients are calculated to measure the similarity between the unlabelled signals and multiple grand average MRCPs. The extracted coefficients are then used as features for linear discriminant analysis, which classifies the signals and predicts their labels. To capture information about the amplitude changes in MRCP signals, STRCA compares the unlabelled signals with the grand average MRCPs and employs the calculated similarity as a feature.

Frequency domain analysis is a widely used approach for analyzing time series data. To leverage information across different frequencies, we proposed the filter bank task-related component analysis method (FBTRCA) [13]. The EEG signals are divided into multiple filter banks, with a consistent low cut-off frequency of 0.5 Hz. The high cut-off frequencies are arranged in an arithmetic sequence ranging from 1 Hz to 10 Hz. Task-related component analysis (TRCA) is employed to extract features from each filter bank, and the resulting features are concatenated. For feature selection, we utilize a method based on mutual information and minimum redundancy maximum relevance to select the most informative features. Finally, the selected features are classified using a support vector machine (SVM) classifier.

In STRCA, the spatial filter is derived by concatenating the task-related component analysis spatial filters of two classes. However, when the number of classes increases, the number of vectors along the concatenation axis exceeds the number of EEG channels. This limitation restricts the use of STRCA and FBTRCA in multi-class tasks. To address this limitation and enable the utilization of STRCA and FBTRCA in multi-class classification, we optimized the structure of the spatial filter and eliminated the common component from the grand average MRCPs. The common component is simply calculated by averaging the grand average MRCPs of multiple classes. This modification allows for the application of FBTRCA in both movement detection and classification tasks [14].

Analyzing EEG signals in the time domain provides valuable insights. For instance, in motor imagery tasks, the use of sliding time windows enhances the classification performance between left and right limbs. A deep learning method called EEGNet has proven to be an effective tool for processing EEG signals [15]. EEGNet is capable of leveraging the temporal information within EEG signals due to its shift-invariant properties in the convolutional layers.

Our previously proposed method, FBTRCA, leverages information from various filter banks. In FBTRCA, the temporal information is assessed using a simple correlation measure between unlabelled EEG signals and the grand average MRCPs. This correlation heavily relies on the grand average MRCPs, which can be susceptible to noise. Although the noise in the grand average MRCPs is mitigated through signal averaging within the same class, it is important to note that the grand average MRCPs used in FBTRCA may still be influenced by unknown noise sources. Furthermore, in scenarios where the precise localization of movement onset is challenging, there may be latency

differences between the MRCP signals of two trials [16].

In this work, our objective is to enhance the temporal decoding of FBTRCA and improve its performance. To achieve this, we propose a two-stage-training temporal-spectral neural network called TTSNet. TTSNet incorporates the shift-invariant properties of convolutional layers and integrates temporal information into the FBTRCA method. The TTSNet model consists of four steps: (1) dividing EEG signals into filter banks, (2) optimizing the EEG signals using the spatial filter, (3) extracting temporal features using EEGNet, and (4) concatenating the extracted features and performing classification using a fully connected layer.

The structure of this work is organized as follows. Section 2 introduces the EEG datasets used in this study and describes the preprocessing steps applied to these datasets. Additionally, it presents the detailed structure of the proposed method. Section 3 evaluates the classification performance of the proposed method on the EEG dataset and compares it against other state-of-the-art methods. Section 4 explains the design of the proposed method and how it utilizes information from the grand average MRCP. Finally, Section 5 provides the concluding remarks for this paper.

2. Material and Methods

2.1. Dataset Description

In this work, we evaluate the classification performance using two public datasets pertaining to upper limb movement, namely Dataset I and Dataset II. Dataset I comprises EEG data from 15 healthy subjects [5], while Dataset II contains EEG data from 10 subjects with cervical spinal cord injuries [6]. The preprocessing of the raw EEG signals in this study follows the same procedures as detailed in our previous works.

Each subject was seated in front of a computer with their arm supported by a table or exoskeleton to prevent muscle fatigue. The signal acquisition paradigm employed in this study followed a trial-based approach, with each trial lasting 5 seconds. At 0 seconds, the trial commenced with an auditory beep, accompanied by a cross displayed on the computer screen. Two seconds later, a cue was presented on the screen, indicating the required movement to be executed by the subject.

The dataset I consists of movements including *elbow flexion*, *elbow extension*, *supination*, *pronation*, *hand close*, *hand open*, and *resting*. In dataset II, the executed movements include *supination*, *pronation*, *hand open*, *palmar*

grasp, and *lateral grasp*. Dataset I comprises 60 trials per class, while dataset II consists of 72 trials per class.”

In dataset I, the movement trajectories of the hand were simultaneously recorded by the exoskeleton during the EEG signal acquisition, allowing for the accurate localization of the movement onset. However, in dataset II, the movement trajectories were not recorded, and thus, the movement onset for each trial is unknown. Additionally, the latency lag between the movement onset and the cue cannot be eliminated in dataset II. To investigate the influence of movement onset on the classification performance, we divide the experiments conducted on dataset I into two cases: dataset I(a) and dataset I(b). In dataset I(a), the recorded movement trajectories are used to locate the movement onset. The EEG signals recorded from one second before the movement onset to one second after the movement onset are used in the evaluation and classification tasks. In dataset I(b), the assumption is made that the movement trajectories were not recorded, making it impossible to reject contaminated trials or accurately locate the movement onset. Consequently, the EEG signals used in the classification task extend from the cue to two seconds after the cue. Similarly, in dataset II, the range of EEG signals used for classification consists of a two-second time window following the cue.

Because dataset I has the simultaneously acquired hand trajectories, but dataset II does not, the movement onset can be located with the trajectories in dataset I. In the localization of the movement onset in dataset I, the 1-order difference of the trajectory is first smoothed by the 1-order *Savitzky-Golay* finite impulse response smoothing filter with time window length 31. The filtered 1-order difference is then normalized by the maximum absolute value. For trials belonging to the *resting* state, a fake movement onset is set to 2.5 s after the trial starts. Trials in the *resting* state are rejected if the variances of normalized trajectory are less than 0.02. In *elbow flexion* and *elbow extension*, the amplitude of the hand trajectory is higher than the other four motions because the limb moves. The movement onset is set to the location where the normalized trajectory equals the threshold of 0.05. Trials are rejected manually when the movement onset is highly influenced by noise contamination. In the other four motions, the function $f(x) = a * \exp(-(\frac{x-b}{c})^2) + d$ is used to fit the smoothed and normalized trajectory by tuning the parameters a, b, c, d [12, 13]. Trials are rejected if the parameters of the tuned function fulfill $a < 0.05$, $c > 100$, and $d > 10$. The movement onset is set to the time point whose absolute amplitude equals 0.1.

The EEG signals for classification were acquired from the motor cortex of

the brain, namely $FC_z, C_3, C_z, C_4, CP_z, F_3, F_z, F_4, P_3, P_z$ and P_4 . The raw EEG signals are firstly downsampled to 256 Hz . The z-normalization is then used to normalize the EEG signals. Because the MRCP signals are located at the low-frequency bands of EEG signals, the normalized EEG signals are bandpassed to $0.5\sim 10\text{ Hz}$.

2.2. Two-stage-training Temporal-Spectral Network

The two-stage-training temporal-spectral network (TTSNet) is further developed based on the FBTRCA method and has four steps: (1) filter bank division, (2) spatial filtering, (3) temporal decoding, and (4) feature fusion and classification. The overall framework of both the FBTRCA and TTSNet is given in Fig. 1, which shows the relationship and the differences between the two methods.

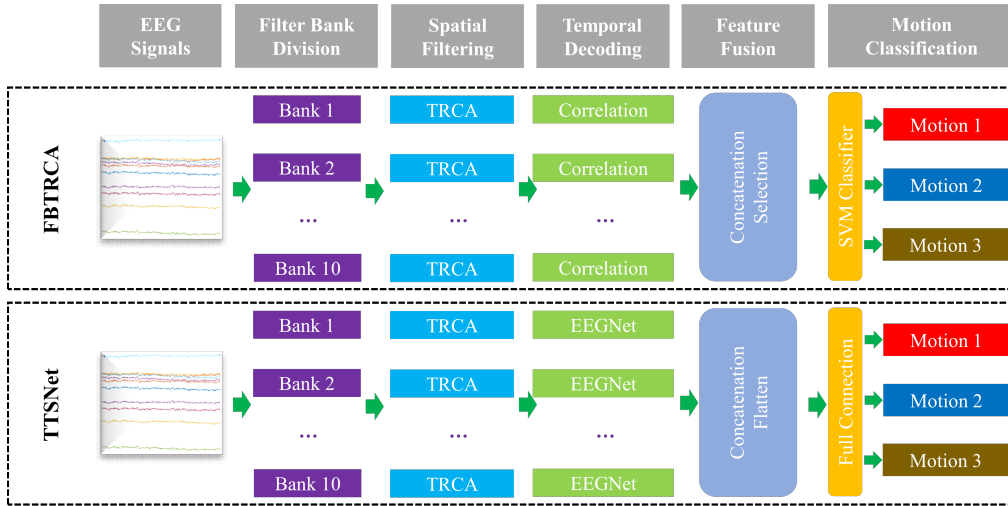


Figure 1: The overall framework of the multi-class FBTRCA method and the TTSNet method. In both methods, EEG signals are divided into filter banks and then optimized with the spatial filter TRCA. During decoding the temporal information from signals, FBTRCA uses the correlation between these signals and the grand average MRCPs as the features. TTSNet uses the EEGNet to capture the temporal information and use the shift-invariant of the convolution layers. After decoding the temporal information, output features from all the filter banks are concatenated. In FBTRCA, these features are selected and optimized by the minimum redundancy maximum relevance method and classified by the support vector machine (SVM) method. In TTSNet, these features are flattened and then classified by the fully connected layer.

2.2.1. Filter Bank Division

The MRCP signals are located at the low-frequency bands of EEG signals. The approximate range of the low-frequency bands is $0.5\sim 10$ Hz. To use the information in the frequency domain of MRCP signals, we proposed a filter bank division method for the low-frequency bands. The low cut-offs of these bands are fixed at 0.5 Hz. The high cut-offs of these bands are sorted as an arithmetic sequence from 1 Hz to 10 Hz with step 1 Hz. Therefore, the range of the low-frequency bands, $0.5\sim 10$ Hz, is divided into F filter banks, where $F = 10$.

2.2.2. Spatial Filtering

After the filter bank division, the MRCP signals are divided into various filter banks. In each filter bank, the multi-channel signals contain task-unrelated components because the filter bank division cannot remove the noises from the original signals. Here, we use the spatial filter to reject the noises and remove task-unrelated components from the original signals in each filter bank. A spatial filter is a matrix W with size $C \times P$, where C is the number of channels and P is an integer smaller than C . Spatial filtering is the operation that multiplies the spatial filter and the raw EEG signals. By the matrix multiplication between the original signals $X \in \mathbb{R}^{C \times T}$ and the W , the spatial-filtered signals $X^T W \in \mathbb{R}^{T \times P}$ has a decreased dimension, and the noises are removed. Here, task-related component analysis (TRCA) is used as the spatial filter. TRCA aims to find a W that maximizes the inter-trial covariances within a class. The training set belonging to the class k is $\mathcal{X}^k = \{X_1, X_2, \dots, X_N\}$, where N is the number of trials of class k and $X_N \in \mathbb{R}^{C \times T}$. The inter-trial covariance of class k is computed with the equation:

$$S^k = \sum_{i,j=1,i<j}^N X_i^T X_j + X_j^T X_i. \quad (1)$$

To normalize the inter-trial covariance, the self-covariance is introduced:

$$Q^k = \sum_{i=1}^N X_i^T X_i. \quad (2)$$

The spatial filter TRCA is obtained by solving the Eigen equation:

$$\max_{\omega} J = \frac{\omega^T S \omega}{\omega^T Q \omega} \quad (3)$$

where $\omega \in \mathcal{R}^{C \times 1}$ is the eigenvector. S and Q are the matrices that summarized the inter-covariances and self-covariances of K classes:

$$S = \sum_{k=1}^K S^k, Q = \sum_{k=1}^K Q^k. \quad (4)$$

The eigenvectors ω of maximum eigenvalues are concatenated into the spatial filter $W \in \mathbb{R}^{C \times P}$, where P is the number of selected eigenvectors.

2.2.3. Temporal Decoding

Temporal decoding serves as a feature extraction mechanism in both FBTRCA and TTSNet. In FBTRCA, the correlated coefficient is utilized as a feature to quantify the dissimilarities between the unlabelled EEG signals and the grand average MRCPs. However, the correlation coefficient can only capture stationary temporal characteristics and cannot effectively handle shifted temporal characteristics. Therefore, we propose TTSNet, which enhances the temporal decoding capability of FBTRCA by incorporating a convolutional neural network (CNN).

Correlation Coefficient. The correlation coefficient measures the similarity of two matrices. In FBTRCA, the two input matrices are the spatial-filtered unlabelled signals and the spatial-filtered grand average MRCPs. In Fig. 2, the structure of computing correlation coefficients in FBTRCA is given.

The grand average MRCP of class k is obtained from the training set \mathcal{X}^k by taking the average of all trials:

$$\hat{X}^k = \frac{1}{N} \sum_{i=1}^N X_i \quad (5)$$

The averaged signals of the grand average MRCPs of K classes is firstly removed from both the grand average MRCPs \hat{X}^k and the unlabelled signals $X \in \mathbb{R}^{C \times T}$:

$$\hat{X}_{\&}^k = \hat{X}^k - \frac{1}{K} \sum_{k=1}^K \hat{X}^k, X_{\&} = X - \frac{1}{K} \sum_{k=1}^K \hat{X}^k \quad (6)$$

The correlation coefficient is a normalized point-wise product of two input matrices. Given two input matrices $X, Y \in \mathbb{R}^{I \times J}$ that fulfil $mean(X) = 0$ and $mean(Y) = 0$, the correlation coefficient is computed using:

$$r = corr(X, Y) = \frac{X * Y}{\sqrt{(X * X) \times (Y * Y)}} \quad (7)$$

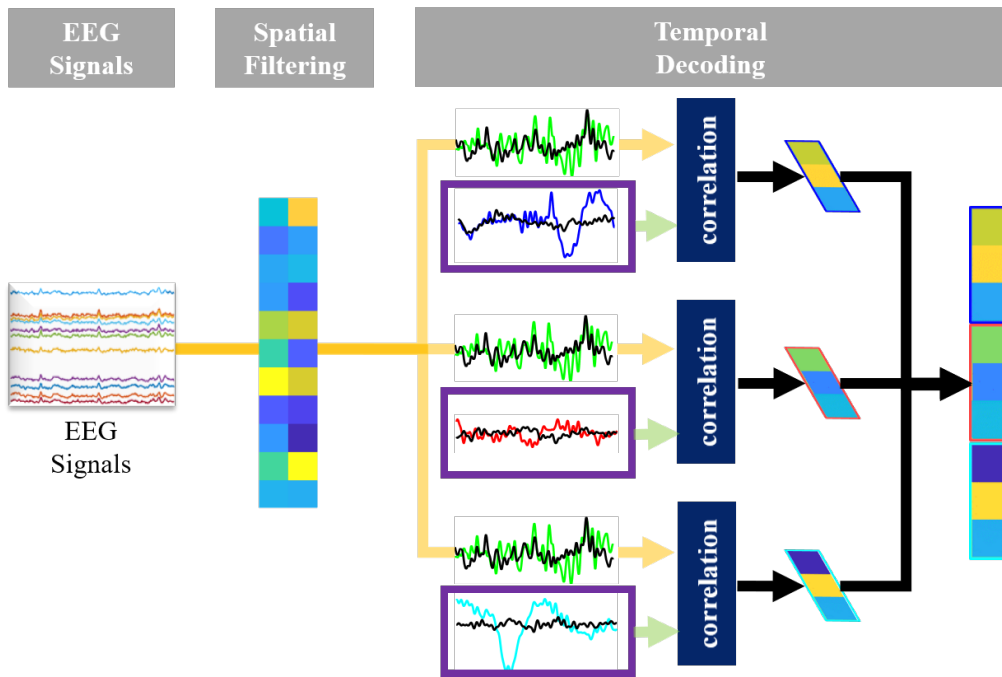


Figure 2: The structure of the temporal decoding component from one filter bank in the FBTRCA method. The EEG signals undergo initial filtering using the spatial filter TRCA to remove unrelated noises and reduce signal dimensions. To capture the amplitude changes of the filtered signals, the correlation between the filtered signals and the corresponding filtered grand average MRCPs (shown in the purple box) of their respective classes is computed as the features. These correlation features from all classes are concatenated to form the output features.

where $*$ denotes the summed-up point-wise products of two input matrices. The symbol \times multiplies two constants $X * X$ and $Y * Y$.

Three kinds of correlation coefficients are computed by taking $X_* = X_{\&}^T W$ and $X_k = \hat{X}_{\&}^k{}^T W$ as the inputs.

(1) Correlation Coefficient

$$\rho_{1,k} = \text{corr}(X_*, X_k); \quad (8)$$

(2) Canonical Correlation Coefficient

$$[A, B] = \text{cca}(X_*, X_k) \quad (9)$$

$$\rho_{2,k} = \text{corr}(X_*B, X_kB); \quad (10)$$

(3) Normalized Canonical Correlation Coefficient

$$[A, B] = \text{cca}(X_* - X_k, X_{-k} - X_k) \quad (11)$$

$$\rho_{3,k} = \text{corr}((X_* - X_k)A, (X_{-k} - X_k)A); \quad (12)$$

where X_{-k} is the mean of spatial-filtered grand average MRCPs of all classes except for class K :

$$X_{-k} = \frac{1}{K-1} \sum_{kk=1, kk \neq k}^K X_{kk}. \quad (13)$$

Therefore, there are $3 \times K \times F$ features in FBTRCA, where F is the number of filter banks.

Convolution Neural Network. The correlation coefficients used in FBTRCA measure the similarity between unlabelled trials and the grand average MRCPs, facilitating the temporal decoding of the signals. However, the temporal decoding aspect of FBTRCA requires improvement for two reasons:

(1) The grand average MRCP is obtained by averaging across trials of the same class. However, averaging alone does not guarantee that the center of trials in the training and testing sets will be accurately learned.

(2) In cases where the movement onset cannot be precisely located or is subject to bias, the MRCP signals may be shifted from the true onset. Correlation alone is unable to effectively handle the misalignment caused by the shifted onset. The issue of misaligned onsets will be discussed further in Section 4.

The correlation coefficient in Equation 7 has two inputs. One is the input MRCP signals, and the other is the grand average MRCP, which can be viewed as the pre-trained weights obtained by averaging EEG signals across trials in the training set. However, the grand average MRCP, serving as the weight applied to the input signals, can be further improved due to its simple derivation through trial averaging.

Replacing the role of correlation in temporal decoding, convolutional neural networks (CNNs) offer two advantages: (1) trainable weights and (2) shift-invariant properties. For this purpose, we employ the network architecture of EEGNet in temporal decoding due to its superior classification performance. The specific network architecture is provided in Table 2.

Table 1: The model structure of temporal decoding in TTSTNet

Layer	Output Size	Parameter
Input Layer	$[B, 1, C, T]$	
ZeroPad2d	$[B, 1, C, T+63]$	(31, 32, 0, 0)
Conv2d	$[B, 8, C, T]$	(1, 64)
BatchNorm2d	$[B, 8, C, T]$	
Conv2d	$[B, 16, 1, T]$	$(C, 1)$, <i>grouped</i>
BatchNorm2d	$[B, 16, 1, T]$	
ELU	$[B, 16, 1, T]$	
AvgPool2d	$[B, 16, 1, T//4]$	(1, 4)
Dropout	$[B, 16, 1, T//4]$	0.25
ZeroPad2d	$[B, 16, 1, T//4+15]$	(7, 8, 0, 0)
Conv2d	$[B, 16, 1, T//4]$	$(1, 15)$, <i>grouped</i>
Conv2d	$[B, 16, 1, T//4]$	(1, 1)
BatchNorm2d	$[B, 16, 1, T//4]$	
ELU	$[B, 16, 1, T//4]$	
AvgPool2d	$[B, 16, 1, T//32]$	(1, 8)
Dropout	$[B, 16, 1, T//32]$	0.25
Flatten	$[B, 16*T//32]$	
Linear	$[B, K]$	<i>bias = False</i>

2.2.4. Classification

In FBTRCA, the features of all filter banks are sorted and selected by the minimum redundancy maximum relevance method and then classified by the support vector machine classifier. The proposed TTSNet uses a fully connected layer to optimize these features and classify the features. whose architecture is given in Table 2.

Table 2: The model structure of classification in TTSNet

Layer	Output Size	Parameter
Input Layer	$[B, K, F]$	
Flatten	$[B, K * F]$	
Linear	$[B, K * F * 2]$	$bias = False$
Relu	$[B, K * F * 2]$	
Linear	$[B, K * F // 2]$	$bias = False$
Relu	$[B, K * F // 2]$	
Linear	$[B, K]$	$bias = False$

2.2.5. Two-stage Training

The TTSNet has two modules that consist of neural networks, the CNN for temporal decoding and the fully connected layer for classification. During training the TTSNet, the two modules are trained separately, *i.e.*, in a two-stage approach. In the first stage, a CNN will be trained for each of the filter banks, and thus the number of trained CNNs is F . The output of F CNNs is concatenated and flattened. In the second stage, the fully connected layer is trained with the flattened features from CNNs. As given in Table 1, the output layer is a linear layer with output size K , where K is the number of classes. Therefore, the network can be trained with the losses between the train label and the outputs. The train label is used twice in the two-stage training process, the temporal decoding in Table 1 and the classification in Table 2. In the first stage, the weights in the fully connected layer are fixed, and the CNNs are trained with train labels. In the second stage, the weights in the CNNs are fixed, and the fully connected layer is trained with train labels.

2.3. Compared Methods

In our previous work [14], the FBTRCA method was compared against both machine learning methods and deep learning methods, such as multi-class common spatial pattern [17], minimum distance to mean [18], WaveNet [19], and Deep CNN [20]. The methods based on common spatial pattern have a worse performance than FBTRCA in the decoding of single-side upper limb movement. The methods for comparison in this work have two state-of-the-art methods (FBTRCA, EEGNet) and three baseline methods.

Baseline I: TEGNet Compared to the EEGNet, The TTSNet fuses the temporal features from all filter banks. To bridge the gap between TTSNet and EEGNet in the comparison, we additionally introduce the Tasked-related EEGNet (TEGNet). The TEGNet has the spatial filtering and temporal decoding part of TTSNet and processes only signals from one filter bank.

Baseline II, OTSNet The TTSNet is trained in two steps. To show the necessity of the two-stage training, the performance of TTSNet is compared to the one-step-training temporal-spectral neural network (OTSNet), which is training in one step without training the temporal decoding CNN module additionally.

Baseline III, TTSSVM In TTSNet, two modifications have been made to FBTRCA, the temporal decoding with EEGNet and the classification with the fully connected layer. To observe the effects of the two modifications, the fully connected layer is replaced with the support vector machine classifier, which is used in FBTRCA.

3. Result Analysis

3.1. Parameter Setting

The classification tasks in this experiment include (1) the binary classification between two motions, *e.g.*, *elbow flexion* and *elbow extension*, and (2) the multi-class classification among all the motions in each dataset. In both classification tasks, the two datasets are split into the training set and the testing set by ten-fold cross-validation. The mean and standard deviation of the accuracy of ten folds are used to evaluate the performance of the mentioned classification methods.

The training process of EEGNet, TEGNet, OTSNet, and TTSNet all has the parameters: learning rate (0.001), batch size (50), optimizer (Adam), and loss function (cross-entropy). In TTSNet, the network is trained in two

steps. The CNN module is firstly trained for 300 epochs. The classification accuracy has converged within the number of training epochs. The model weight of the CNN module is then not updated in the following training process of the fully connected layer. In the training process of the second module, the Adam optimizer is additionally equipped with the weight decay (L2 penalty) of 0.1 to ensure the early stop of the training process.

The number of selected maximum eigenvalues P in the spatial filter is a hyper-parameter. This hyper-parameter is determined by a nested cross-validation. In the nested cross validation, the classification performance is evaluated by the STRCA method for each of the filter banks.

3.2. Result

In this section, five methods are compared to the proposed TTSNet method, including FBTRCA, EEGNet, TEGNet, OTSNet and TTSSVM. EEG signals from Dataset I and Dataset II are used in three cases: (1) Dataset I(a)-Dataset I with the movement onset located, (2) Dataset I(b)-Dataset I without the located movement onset and (3) Dataset II without the located movement onset. In this section, we will use the dataset I(a) to analyze (1) the simplified network architecture in temporal decoding, (2) the influence of the temporal decoding and classifier, and (3) the necessity of the two-stage training process. The overall performance of these methods in the three cases is compared finally.

Before the statistic analysis in the three cases, we first compare the EEG signals before and after spatial filtering (with task-related component analysis) in the channel space and the source space, which is given in Fig 3. The figures for the channel space are plotted with EEGLab [21]. The figures for the source space are plotted with Brainstorm [22]. During plotting the source space figures, the brain anatomy used is BCI-DNI BrainSuite [23]. The surface is segmented by the boundary element method with Brainstorm [24]. The source spaces figures are finally given by the standardized low-resolution brain electromagnetic tomography method [25].

3.2.1. Simplified Neural Network in Temporal Decoding

In TEGNet, the multi-channel EEG signals are first passed by a spatial filter, the filtered signals are then devoted into the EEGNet. Compared to EEGNet, the TEGNet reduces the dimension of EEG signals with a linear transform of size $C \times P$. This linear transform can be easily obtained by solving the Eigen equation in Equation 3 with the training set. With the

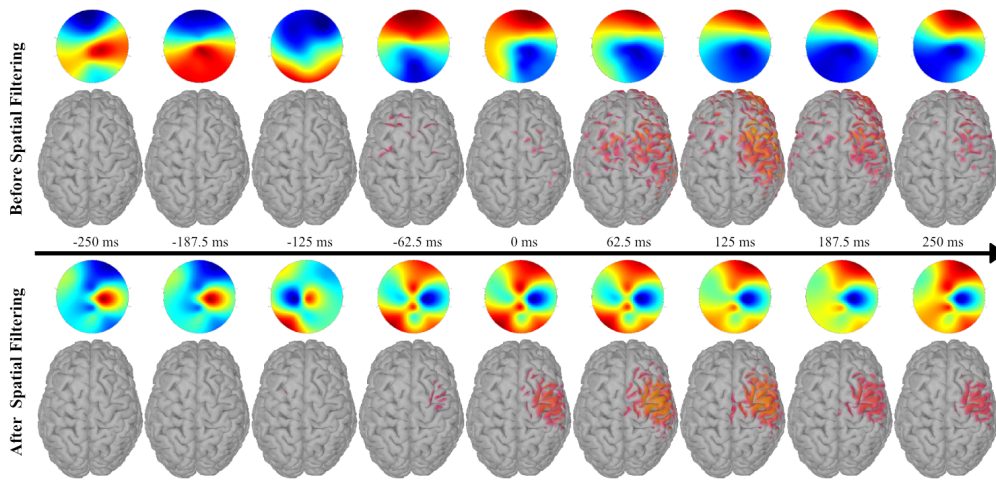


Figure 3: Comparison of the grand average MRCP signals ($0.5\sim 10\text{ Hz}$) before and after the spatial filtering. This figure visualizes the grand average MRCP signals of elbow flexion of subject 1 from dataset I in both the channel space and the source space. The 0 ms in the figure denotes the movement onset. When visualizing the signals after spatial filtering, the problem is that the number of channels of filtered signals is decreased. To solve this problem, the filtered signals $X^T W$ are multiplied by the inverse spatial filter W^T , and we get $X^T W W^T$. $X^T W W^T$ reserves the filtered signals to the original channels that are the same as the channels before spatial filtering.

spatial filtering, the input of the neural network part in TEGNet has fewer number of channels. To show the equivalence of EEGNet and TEGNet, Figure 4 compares the multi-class classification performance in the dataset I(a). The average classification accuracy is 0.4161 ± 0.0748 and 0.4205 ± 0.0776 on average for EEGNet and TEGNet, respectively.

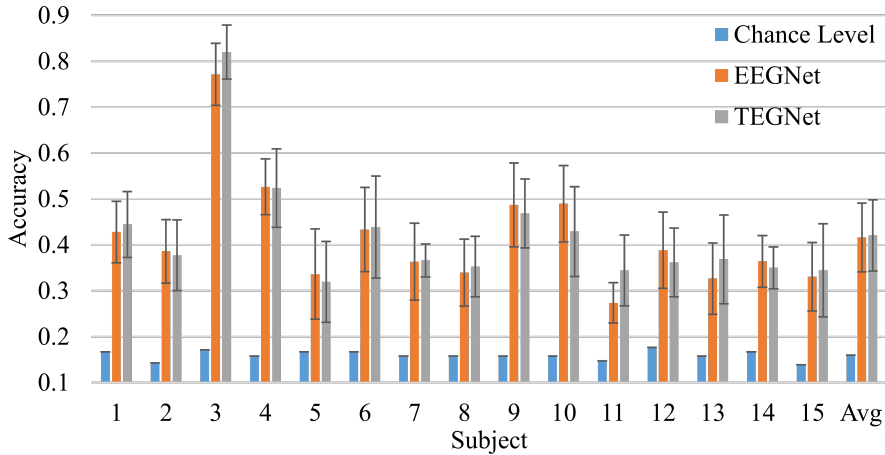


Figure 4: Multi-class classification performance comparison between EEGNet and TEGNet, validated in dataset I(a). EEGNet and TEGNet have an equivalent performance in average.

3.2.2. Neural Networks in Temporal Decoding and Classifier

The TTSNet has two modifications on the FBTRCA method, including the EEGNet in the temporal decoding and the fully connected layer in the classification. Both modifications may lead changes to the classification performance. To check the influence of modified temporal decoding on the classification performance, the TTSSVM is compared to the proposed TTSNet. TTSSVM replaces only the temporal decoding part. In the classification, the same feature selection and classification are used as in the FBTRCA. Figure 5 illustrates the classification comparison of FBTRCA, TTSSVM and TTSNet. It shows that the modifications in the temporal decoding and the classifier can increase the average classification accuracy with 1.92% and 2.03%, respectively. In total, TTSNet has an increase of 3.95% on the FBTRCA method.

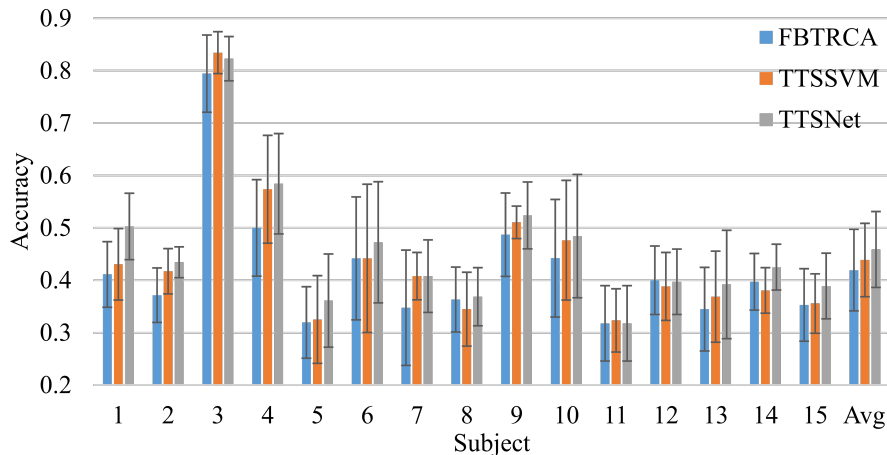


Figure 5: Multi-class classification performance comparison among FBTRCA, TTSSVM and TTSNet.

3.2.3. Two-stage-training process

The training processing of parameters in the TTSNet adopts a two-stage process. In the first stage, the spatial filter is learned by solving the Eigen equation, and the parameters in EEGNets are learned with the filtered signals as inputs and the true labels as outputs. In the second stage, the learning of parameters in the EEGNet is turned off and the fully connected layer is trained. OTSNet holds the same structure as TTSNet but parameters in both EEGNet and fully connected layer are trained simultaneously. Figure 6 compares the performance differences between OTSNet and TTSNet. The reason why TTSNet has a better performance than OTSNet is further explored in the discussion.

3.3. Performance Summary

In Table 3, the overall classification accuracies are listed, including binary and multi-class results in three dataset cases. Table 4 lists the p -values between TTSNet and the other methods in dataset I(a). In the dataset I(b) and the dataset II, the movement onset cannot be exactly located. The misalignment of the movement onset has a negative influence on the multi-class classification. The reason of the negative influence of the misalignment will be analyzed in the discussion. The TTSNet helps to improve the multi-class classification in dataset I(b). However, the classification performance in dataset I(b) still has a distance from the performance of dataset I(a). In

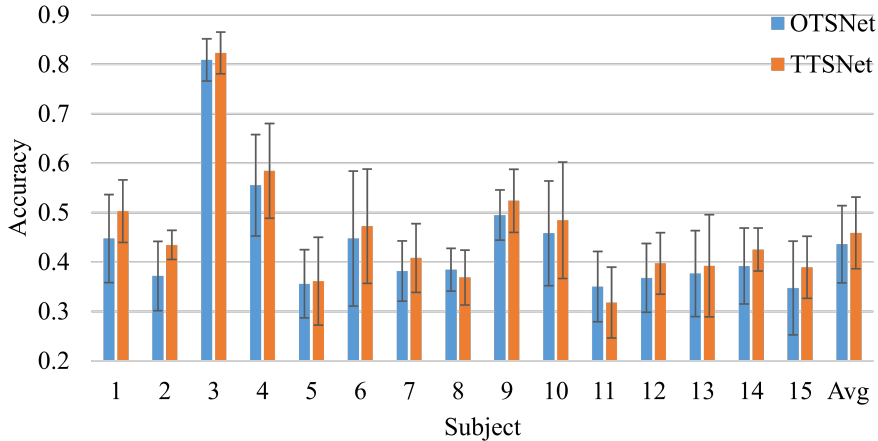


Figure 6: Performance difference between the one-stage-training process and the two-stage-training process in the multi-class classification task.

summary, TTSNet contributes to the classification performance of FBTRCA by enhancing the temporal decoding process.

Table 3: Overall average accuracy comparison

Dataset	Dataset I(a)		Dataset I(b)		Dataset II	
	Binary	Multi-class	Binary	Multi-class	Binary	Multi-class
Chance level	0.5301±0.0000	0.1594±0.0000	0.5000±0.0000	0.1429±0.0000	0.5000±0.0000	0.2000±0.0000
FBTRCA	0.7487±0.1250	0.4193±0.0780	0.7414±0.1258	0.3878±0.0664	0.6647±0.1190	0.3975±0.0675
EEGNet	0.7340±0.1246	0.4161±0.0748	0.7246±0.1161	0.3889±0.0729	0.6718±0.1160	0.4057±0.0689
TEGNet	0.7176±0.1252	0.4205±0.0776	0.7126±0.1189	0.3843±0.0724	0.6627±0.1148	0.3905±0.0847
OTSNet	0.7521±0.1211	0.4358±0.0780	0.7375±0.1171	0.3946±0.0633	0.6857±0.1135	0.4079±0.0743
TTSSVM	0.7659±0.1182	0.4385±0.0702	0.7494±0.1126	0.4017±0.0633	0.7021±0.1155	0.4121±0.0736
TTSNet	0.7707±0.1168	0.4588±0.0724	0.7526±0.1122	0.4141±0.0679	0.7075±0.1159	0.4311±0.0700

4. Discussion

In this work, our focus lies on the multi-class classification of single-side upper limb movements. We enhance the performance of our previously proposed FBTRCA method by introducing temporal decoding using a convolutional neural network. To underscore the contributions of this work, it is essential to discuss three key points: (1) the relationship between movements of the double-side and single-side of the limb, (2) the rationale for replacing correlation with a convolutional neural network, and (3) the novelty of the two-stage training process.

Table 4: The p -value between TTSNet and other methods in dataset I(a)

Subject	1	2	3	4	5	6	7	8
FBTRCA	0.0044	0.0038	0.3005	0.0604	0.2550	0.5650	0.1586	0.8434
EEGNet	0.0193	0.0556	0.0556	0.1243	0.5583	0.4159	0.2103	0.3306
TEGNet	0.0699	0.0421	0.9020	0.1536	0.3061	0.5191	0.1057	0.5693
OTSNNet	0.1260	0.0175	0.4613	0.5229	0.8776	0.6636	0.3798	0.4872
	9	10	11	12	13	14	15	
	0.2687	0.4238	1.0000	0.9191	0.2666	0.2218	0.2380	
	0.3097	0.9095	0.1147	0.7911	0.1250	0.0141	0.0747	
	0.0934	0.2685	0.4377	0.2666	0.6035	0.0014	0.2546	
	0.2766	0.6061	0.3253	0.3332	0.7159	0.2482	0.2612	

4.1. Single-side and Double-side Upper Limb Movement

The multi-class classification of limb movements involves two distinct tasks: (1) classification of multiple limbs and (2) classification of multiple movements on the same limb.

In the first task, classification methods are developed based on the brain phenomenon known as event-related desynchronization/synchronization (ERD/ERS). This phenomenon involves power changes on the brain scalp, with a decrease in power on the same side of the brain as the executing limb and an increase on the contra-side half of the brain. When different limbs execute movements, the activated brain regions vary accordingly. The spatial filter known as common spatial pattern (CSP) is employed in this task to distinguish the activated brain regions [26]. The classification methods for this task focus on power changes in multiple brain regions within the frequency range of 8 to 40 Hz.

In the second task, an intuitive idea based on ERD/ERS is that power changes evoked by movements of the same limb are localized to similar regions. Therefore, the classification of multiple movements on the same limb relies on the similarity of activated brain regions. In this task, the differences between classes are reflected in the grand average MRCPs of each class. After averaging the signals across trials, the signals are smoothed and limited to the frequency range of 0.01 to 10 Hz.

The FBTRCA method optimizes performance robustness by fusing features from multiple filter banks. The role of the spatial filter, task-related component analysis, can be interpreted as the localization of regions of interest, as depicted in Figure 3. In the temporal decoding part, simple correlation is used to extract temporal information. However, it may not be sufficient to fully decode the temporal information.

4.2. From Correlation to Convolutional Neural Network

In our previously proposed STRCA method, both the spatial filter and the grand average MRCP were obtained from the training set. The grand average MRCP was computed by averaging EEG trials belonging to the same class in the training set. In the correlation features of an EEG trial, both the grand average MRCP and the EEG trial were first optimized with the spatial filter, and then the correlation between the EEG trial and the grand average MRCP was computed. The correlation value represented the sum of the weighted signals in the EEG trials, where the weights were determined by the grand average MRCP. Consequently, the grand average MRCP played a crucial role in extracting the correlation features. Averaging the signals in the grand average MRCP aimed to reduce the influence of noise in the original EEG trials. As illustrated in Figure 7, the noise in the relatively high-frequency bands was effectively eliminated in the grand average MRCP.

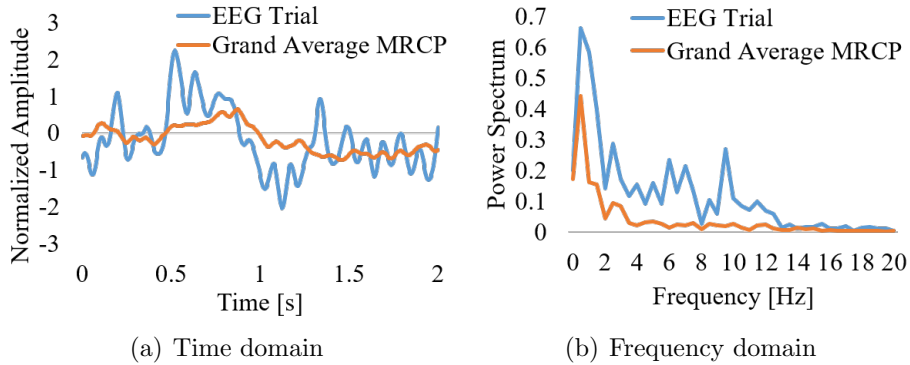
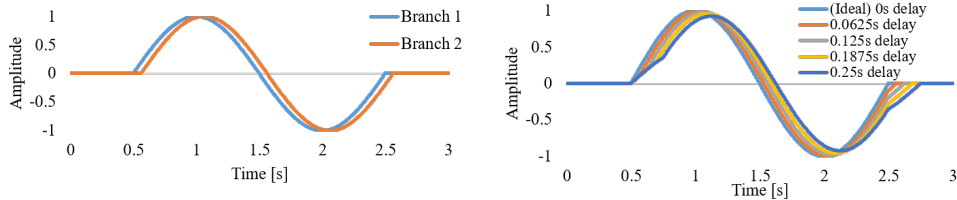


Figure 7: The difference between the grand average MRCP and an EEG trial. This figure is an example whose data is extracted from the channel C_z of the *elbow flexion*.

However, the averaging operation can introduce inaccurate grand average MRCPs due to misalignment of movement onsets. To simplify the explanation of the misalignment problem, we made two assumptions. Firstly, we simplified the ideal curve of MRCP signals (an increase followed by a decrease) as a sine function. Secondly, we assumed that the noises in the EEG signals are uniformly distributed along the timeline and are fully eliminated by averaging.

In the training set of a class, two branches were assumed, each with an equal number of trials, and the noises were removed by averaging the trials

within each branch. The movement onsets of trials within each branch were strongly aligned. However, the movement onsets between the two branches were not the same, and this was the only difference between them. After averaging the trials within each branch, Figure 8(a) illustrates the latency lag between the two branches. Considering that both branches belong to the training set, the final grand average MRCP is obtained by averaging the MRCPs from the two branches. However, the latency lag between the two branches can influence the shape of the grand average MRCP, as shown in Figure 8(b). The latency lag distorts the shape of the grand average MRCP in two ways: by reducing the highest amplitude/power and by introducing bias in the timing.



(a) The grand average MRCP of two branches. Noises are fully removed after training set, the grand average MRCP of averaging. There is a lag off between the two branches. (b) In the grand average MRCP of the two branches are averaged. The lag off between two branches leads to the distortion of the grand average MRCP.

Figure 8: Influences of misaligned movement onset on the grand average MRCP. The ideal grand average MRCP is simplified into a sine wave and noises are diminished to zeros after averaging across trials. The training set consists of two branches and two branches have the same number of trials.

In the computation of the real grand average MRCP, the situation becomes more complex as the movement onsets of all trials may not be the same, unlike the two distinct movement onsets illustrated in Figure 8. In cases where the movement trajectory is available, such as in dataset I(a), the movement onsets can be localized, and the EEG signals before and after the movement onset can be sliced for the classification task. By individually locating the movement onsets for all trials, the latency lag depicted in Figure 8(a) can be minimized. However, in most cases, the movement trajectories are not recorded simultaneously, and the EEG signals are typically sliced from the two-second time window after the cue to execute the movement. Due to variations in the reflection time from the brain to the limb, the dis-

tance between the movement onset and the cue to execute the movement differs across trials, and this cannot be precisely measured. Consequently, this leads to distortions in the ideal grand average MRCP, as shown in Figure 8(b).

Although the movement onset in dataset I(a) can be localized with the movement trajectory, we acknowledge the possibility of misalignment caused by the inaccurate localization of the onset and the unmeasurable reflection time from the brain to the limb. To mitigate the influence of the distorted grand average MRCP on feature extraction, we propose replacing this weight (the grand average MRCP) with a more trainable weight (EEGNet).

4.3. Two-stage-training Process

In the proposed TTSNet method, two modifications are applied to the previous FBTRCA method: the utilization of EEGNet in the temporal decoding and the incorporation of fully connected layers in the classification. During the training of TTSNet, the parameters of the EEGNet and the fully connected layer are trained in two separate steps. This two-stage training process is implemented to fulfill the requirements of the filter bank technique in MRCP signal processing.

The FBTRCA method was developed by incorporating the filter bank technique into the STRCA method, as described in our previous work [13]. In contrast to the filter bank common spatial pattern method used for decoding double-side limb movements [26], the filter bank technique aims to enhance the robustness of the classification performance in single-side limb movements. In double-side limb movements, common spatial patterns are extracted from multiple frequency bands and sliding windows. The resulting features indicate differences ('1') or no differences ('0'), representing a sparse '0' and '1' problem. Feature selection in double-side limb movements focuses on selecting the '1' features against the '0' features.

However, in the FBTRCA method, the features in each filter bank are extracted using STRCA, and the classification performance among the filter banks is generally similar. The filter bank selection in FBTRCA is employed to choose the features with the best performance from the features that demonstrate acceptable performance. For instance, the classification accuracy of STRCA ranges from 0.85 to 0.95 across these filter banks, and the feature selection method selects the features with an accuracy of 0.95.

In the decoding of single-side upper limb movements, the features within each filter bank are expected to capture class-related information. To select

the most informative features, either feature selection or a fully connected layer is employed for feature fusion across all filter banks. In the TTSNet method, spatial filtering and temporal decoding are utilized to extract features from each filter bank. The resulting features from the EEGNet component should contain the necessary information for predicting the class labels. Therefore, the spatial filter and EEGNet are initially trained in the first stage. Subsequently, the fully connected layer acts as both a feature selector and classifier to enhance the robustness of the extracted features.

5. Conclusion

In this work, we propose TTSNet, a method that leverages both temporal and spectral information to decode patterns from MRCP signals. TTSNet integrates the temporal decoding capability of EEGNet into the FBTRCA method, enabling the extraction of distinct features that capture both temporal and spectral characteristics. This method is applicable to both binary and multi-class classification tasks for upper limb movements. Our results demonstrate that TTSNet outperforms both FBTRCA and EEGNet in decoding pre-movement patterns. The findings of this study have implications for the rehabilitation of individuals with disabled or weak upper limbs. The code repository for this work can be accessed via the following link: <https://github.com/plustar/Movement-Related-Cortical-Potential.git>.

6. Acknowledgement

This work is carried out as part of doctoral programme in University of Vic-Central University of Catalunya. C.F.C work was partially supported by grants PICT 2020-SERIEA-00457 and PIP 112202101 00284CO (Argentina).

Statement: During the preparation of this work the author(s) used Pytorch in order to generate neural network models. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

- [1] Alexander E. Hramov, Vladimir A. Maksimenko, and Alexander N. Pisarchik. Physical Principles of Brain–computer Interfaces and Their Applications for Rehabilitation, Robotics and Control of Human Brain States. *Physics Reports*, 918:1–133, June 2021.

- [2] Simanto Saha, Khondaker A. Mamun, Khawza Ahmed, Raqibul Mostafa, Ganesh R. Naik, Sam Darvishi, Ahsan H. Khandoker, and Mathias Baumert. Progress in Brain Computer Interface: Challenges and Opportunities. *Frontiers in Systems Neuroscience*, 15:578875, 2021.
- [3] Aleksandra Kawala-Sterniuk, Natalia Browarska, Amir Al-Bakri, Mariusz Pelc, Jaroslaw Zygarlicki, Michaela Sidikova, Radek Martinek, and Edward Jacek Gorzelanczyk. Summary of over Fifty Years with Brain-Computer Interfaces—A Review. *Brain Sciences*, 11(1):43, 2021.
- [4] Sharon Olsen, Nada Signal, Imran Khan Niazi, Thomas Christensen, Mads Jochumsen, and Denise Taylor. Paired Associative Stimulation Delivered by Pairing Movement-Related Cortical Potentials With Peripheral Electrical Stimulation: An Investigation of the Duration of Neuromodulatory Effects: PAS VIA PAIRING MRCPs WITH ELECTRICAL STIMULATION. *Neuromodulation: Technology at the Neural Interface*, 21(4):362–367, 2018.
- [5] Patrick Ofner, Andreas Schwarz, Joana Pereira, and Gernot R. Müller-Putz. Upper Limb Movements can be Decoded from the Time-domain of Low-frequency EEG. *PLOS ONE*, 12(8):e0182578, 2017.
- [6] Patrick Ofner, Andreas Schwarz, Joana Pereira, Daniela Wyss, Renate Wildburger, and Gernot R. Müller-Putz. Attempted Arm and Hand Movements can be Decoded from Low-Frequency EEG from Persons with Spinal Cord Injury. *Scientific Reports*, 9(1):7134, 2019.
- [7] Dario Farina, Winnie Jensen, and Metin Akay, editors. *Introduction to neural engineering for motor rehabilitation*. John Wiley & Sons, Hoboken, N.J, 2013.
- [8] Marta Borràs, Sergio Romero, Joan F Alonso, Alejandro Bachiller, Leidy Y Serna, Carolina Migliorelli, and Miguel A Mañanas. Influence of the Number of Trials on Evoked Motor Cortical Activity in EEG Recordings. *Journal of Neural Engineering*, 19(4):046050, August 2022.
- [9] Rasmus L. Kaeseler, Tim Warburg Johansson, Lotte N. S. Andreasen Struijk, and Mads Jochumsen. Feature and Classification Analysis for Detection and Classification of Tongue Movements From Single-Trial

- Pre-Movement EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:678–687, 2022.
- [10] Ji-Hoon Jeong, No-Sang Kwak, Cuntai Guan, and Seong-Whan Lee. Decoding Movement-Related Cortical Potentials Based on Subject-Dependent and Section-Wise Spectral Filtering. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(3):687–698, 2020.
- [11] Bradley J. Edelman, Bryan Baxter, and Bin He. EEG Source Imaging Enhances the Decoding of Complex Right-Hand Motor Imagery Tasks. *IEEE Transactions on Biomedical Engineering*, 63(1):4–14, 2016.
- [12] Feng Duan, Hao Jia, Zhe Sun, Kai Zhang, Yangyang Dai, and Yu Zhang. Decoding Premovement Patterns with Task-Related Component Analysis. *Cognitive Computation*, 13(5):1389–1405, 2021.
- [13] Hao Jia, Zhe Sun, Feng Duan, Yu Zhang, Cesar F Caiafa, and Jordi Solé-Casals. Improving pre-movement pattern detection with filter bank selection. *Journal of Neural Engineering*, 2022.
- [14] Hao Jia, Fan Feng, Cesar F. Caiafa, Feng Duan, Yu Zhang, Zhe Sun, and Jordi Solé-Casals. Multi-class Classification of Upper Limb Movements with Filter Bank Task-related Component Analysis. *IEEE Journal of Biomedical and Health Informatics*, pages 1–11, 2023.
- [15] Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance. EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces. *Journal of Neural Engineering*, 15(5):056013, 2018.
- [16] Andreea Ioana Sburlea, Luis Montesano, and Javier Minguez. Continuous detection of the self-initiated walking pre-movement state from EEG correlates without session-to-session recalibration. *Journal of Neural Engineering*, 12(3):036007, 2015.
- [17] Moritz Grosse-Wentrup and Martin Buss. Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction. *IEEE Transactions on Biomedical Engineering*, 55(8):1991–2000, August 2008.

- [18] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten. Multiclass Brain–Computer Interface Classification by Riemannian Geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2012.
- [19] Punyawish Thuwajit, Phurin Rangpong, Phattarapong Sawangjai, Phairot Autthasan, Rattanaphon Chaisaen, Nannapas Banluesombatkul, Puttaranun Boonchit, Nattasate Tatsaringkansakul, Thapanun Sudhawiyangkul, and Theerawat Wilaiprasitporn. EEGWaveNet: Multiscale CNN-Based Spatiotemporal Feature Extraction for EEG Seizure Detection. *IEEE Transactions on Industrial Informatics*, 18(8):5547–5557, 2022.
- [20] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggenesperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for EEG decoding and visualization: Convolutional Neural Networks in EEG Analysis. *Human Brain Mapping*, 38(11):5391–5420, 2017.
- [21] Arnaud Delorme and Scott Makeig. EEGLAB: An Open Source Toolbox for Analysis of Single-trial EEG Dynamics including Independent Component Analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004.
- [22] François Tadel, Sylvain Baillet, John C. Mosher, Dimitrios Pantazis, and Richard M. Leahy. Brainstorm: A User-Friendly Application for MEG/EEG Analysis. *Computational Intelligence and Neuroscience*, 2011:1–13, 2011.
- [23] Anand A. Joshi, Soyoung Choi, Yijun Liu, Minqi Chong, Gaurav Sonkar, Jorge Gonzalez-Martinez, Dileep Nair, Jessica L. Wisnowski, Justin P. Haldar, David W. Shattuck, Hanna Damasio, and Richard M. Leahy. A Hybrid High-resolution Anatomical MRI Atlas with Sub-parcellation of Cortical Gyri using Resting fMRI. *Journal of Neuroscience Methods*, 374:109566, 2022.
- [24] J.H.M. Frijns, S.L. de Snoo, and R. Schoonhoven. Improving the Accuracy of the Boundary Element Method by the Use of Second-order Interpolation Functions [EEG Modeling Application]. *IEEE Transactions on Biomedical Engineering*, 47(10):1336–1346, 2000.

- [25] Roberto D Pascual-Marqui. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Clinical Pharmacology*, page 16, 2002.
- [26] Yu Zhang, Chang S. Nam, Guoxu Zhou, Jing Jin, Xingyu Wang, and Andrzej Cichocki. Temporally Constrained Sparse Group Spatial Patterns for Motor Imagery BCI. *IEEE Transactions on Cybernetics*, 49(9):3322–3332, 2019.

4. SUMMARY OF THE WORK CARRIED OUT WITHIN THE THESIS

4.1. Data augmentation on the small-size dataset

Deep learning has gained significant popularity in signal processing and analysis due to its impressive performance in tasks such as classification and regression. However, achieving optimal performance in deep learning models requires a large number of parameters, which in turn necessitates a substantial amount of labeled data for training and fine-tuning. Unfortunately, in the case of small-size Alzheimer's disease (AD) datasets, fulfilling this requirement becomes challenging.

4.1.1. Pre-requisition

In this subsection, several fundamental concepts are introduced. These concepts include the notion of training and testing data, the notion of functional connectivity, and the notion of data augmentation in the small-size AD datasets.

Concept of training and testing data

The evaluation of the classification models is performed using the acquired electroencephalogram (EEG) signals. As depicted in Figure 4.1, the EEG signals are divided into a training set and a testing set.

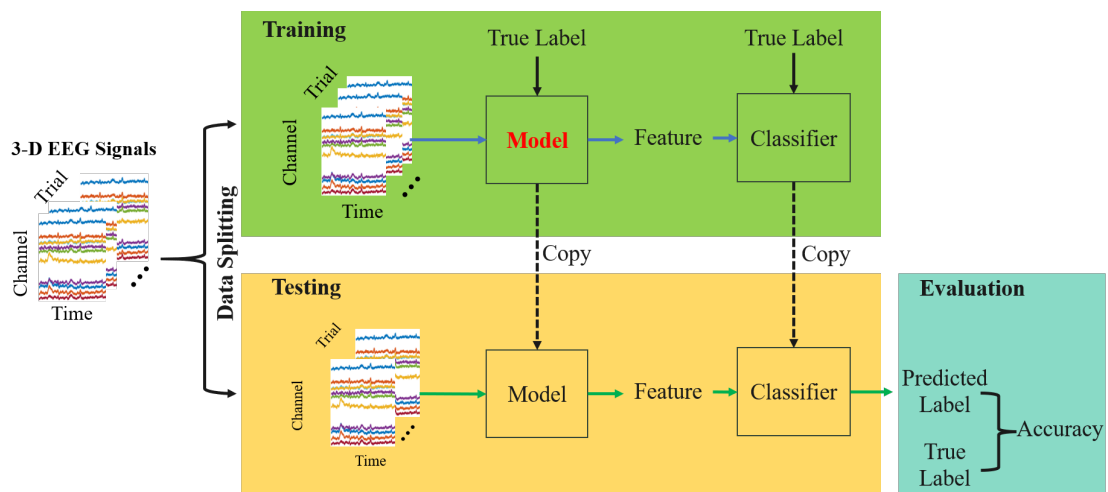


Figure 4.1: Evaluation of classification performance using the training set and testing set.

The classification model is separated into two modules: the model for feature extraction and the classifier for classification. In the training set, the label of each trial is known,

allowing the model and the classifier to be fitted with the corresponding data and labels. In the testing set, the true labels are assumed to be known. The model trained on the training set is employed to generate data features for the testing set. The classifier trained on the training set predicts the labels of these testing features, resulting in predicted labels. The classification accuracy is evaluated by comparing the predicted labels with the true labels of the testing set. To assess the robust performance of the classification model, ten-fold cross-validation is utilized.

Concept of functional connectivity

In the detection of AD, the key differences between classes lie in the spatial distribution or the relationships among channels. Therefore, in order to capture the inter-channel relationships, the EEG signals are transformed into functional connectivity (FC) measures. The FC matrices are computed using normalized coherence. For two temporal signals, denoted as $x(t)$ and $y(t)$, the coherence between them can be defined as:

$$C_{xy} = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)}, \quad (4.1)$$

where G_{xy} represents the cross-spectral density between x and y , while G_{xx} and G_{yy} correspond to the power spectral densities of x and y respectively. The resulting FC matrix calculates the coherence between each pair of EEG signals and has a size of $C \times C$, where C denotes the number of channels.

Concept of data augmentation

EEG signals, being non-invasive and harmless to individuals, provide a convenient means to analyze brain activity. However, when it comes to the classification of patients with AD and healthy individuals, the available trials in the training set are often scarce. The limited amount of training data poses the risk of overfitting the deep learning model. Due to the constraints imposed by the number of available AD patients, it becomes increasingly difficult to acquire a substantial amount of data. Therefore, enlarging the size of the training set becomes crucial in order to adequately fit the parameters of deep learning models, as illustrated in Figure 4.2.

4.1.2. Data augmentation with a decomposition and recombination strategy

Data augmentation is employed on the training set to expand the number of available trials. It generates artificial signals by applying operations such as rotation and reversal to the original signals in the training set. In this work, a decomposition and recombination (DR) strategy is utilized to generate artificial signals. The DR strategy consists of two steps:

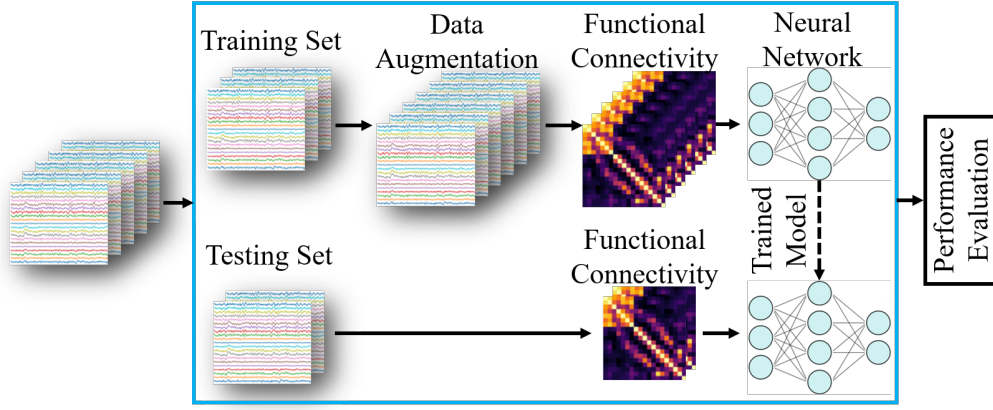


Figure 4.2: Data augmentation is a technique used to increase the size of the training set, allowing deep learning models to learn more effectively.

- Decompose the original signals into multiple non-overlapped components;
- Recombine these components to generate artificial signals.

To formalize the DR strategy, we consider two domains: the fusion domain \mathcal{U} and the intrinsic domain \mathcal{V} . The original signal $\mathbf{x}^i \in \mathbb{R}^{T \times 1}$ and the artificial signal $\mathbf{x}' \in \mathbb{R}^{T \times 1}$ reside in the fusion domain \mathcal{U} , where $i = 1, 2, \dots, I$; I is the number of original signals; T is the number of sample points in a time series. The set of decomposed components $Y = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_R] \in \mathbb{R}^{T \times R}$ is in the intrinsic domain \mathcal{V} , where R is the number of decomposed components. The decomposition of fusion signals is a forward process from the original domain to the intrinsic domain,

$$\mathcal{U} \rightarrow \mathcal{V} : Y^i = f(\mathbf{x}^i). \quad (4.2)$$

The recombination of these components is an inverse process from the intrinsic domain to the fusion domain,

$$\mathcal{V} \rightarrow \mathcal{U} : \mathbf{x}^i = f'(Y^i) = f'([\mathbf{y}_1^i, \mathbf{y}_2^i, \dots, \mathbf{y}_R^i]). \quad (4.3)$$

The artificial signal \mathbf{x}' is obtained by a random combination of components in the intrinsic domain \mathcal{V} ,

$$\mathbf{x}' = f'([\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_R]), \quad (4.4)$$

where \mathbf{y}'_r is randomly selected from the set $\{\mathbf{y}_r^1, \mathbf{y}_r^2, \dots, \mathbf{y}_r^I\}$, $r = 1, 2, \dots, R$.

Signal decomposition

In signal decomposition, the empirical mode decomposition (EMD) is utilized to decompose the time series into multiple intrinsic mode functions (IMFs). This method offers two advantages when employed as the decomposition technique in the DR strategy:

- The IMFs exhibit minimal overlapping regions in the frequency domain.
- The original signal can be easily reconstructed by summing all the IMFs.

This work mainly discusses four types of EMD, namely classical EMD (CEMD), serial EMD (SEMD), and multivariate EMD (MEMD) and graph EMD (GEMD). SEMD and MEMD are both based on CEMD, and the proposed SEMD-MEMD method combines both SEMD and MEMD.

CEMD CEMD represents the initial EMD method with a simple architecture. It decomposes a single-channel time series into multiple IMFs, as depicted in Figure 4.3(a). In the frequency domain, the spectrograms of these IMFs do not overlap, as illustrated in Figure 4.3(b). The procedure for CEMD is outlined in Algorithm 1.

The sifting processing of CEMD is the loop process in Algorithm 1, which searches for the IMFs iteratively. CEMD has two key points in the sifting process, including the signal interpolation and the stop criterion. In Algorithm 1, the signals between extremes are interpolated with cubic spline interpolation. The stop criterion is but not limited to the number of extremes in the residual IMF. The other possible stop criteria are the number of IMFs, the number of iterations in the sifting process, and so on.

Algorithm 1 Classical empirical mode decomposition

Require: $T \in \mathbb{Z}$

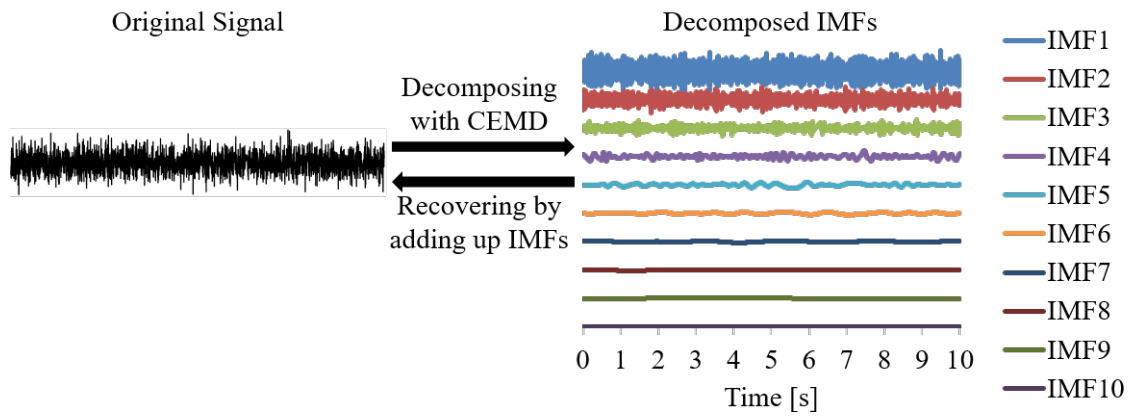
Ensure: $\mathbf{x} \in \mathbb{R}^{T \times 1}$

```

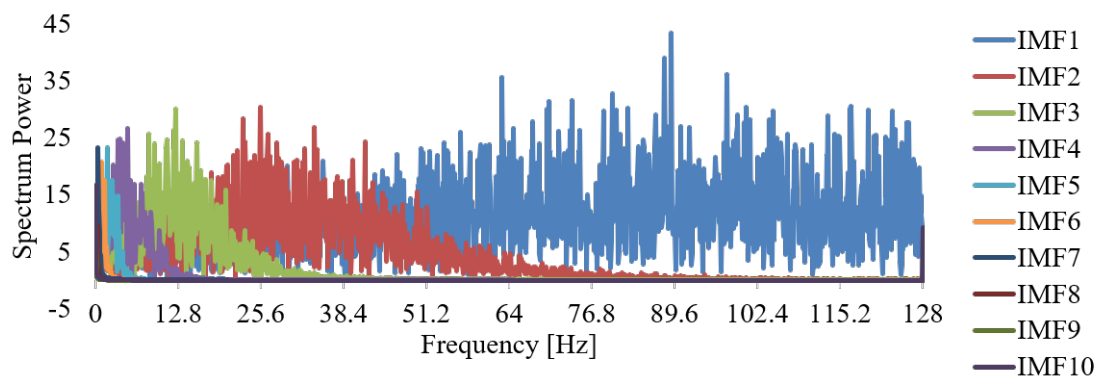
while number of extreme values in  $\mathbf{x}$  is greater than 3 do
    interpolate the maxima extremes (minima extremes) of  $\mathbf{x}$ 
    obtain the envelop of extremes  $e_{max}$  ( $e_{min}$ )
     $\mathbf{m} = (e_{max} + e_{min})/2$ 
     $\mathbf{h} = \mathbf{x} - \mathbf{m}$ 
    if  $\mathbf{h}$  is an IMF then
        save  $\mathbf{h}$  as an IMF
         $\mathbf{x} = \mathbf{x} - \mathbf{r}$ 
    else
         $\mathbf{x} = \mathbf{h}$ 
    end if
end while
save  $\mathbf{x}$  as an IMF

```

SEMD CEMD is employed for decomposing single-channel time series. To extend this decomposition technique to multi-channel time series, SEMD converts the multi-channel time series into a single-channel time series by concatenating the individual time series along the time axis. The flowchart of SEMD is presented in Figure 4.4.



(a) IMFs decomposed by CEMD



(b) The spectrum of the IMFs in the frequency domain

Figure 4.3: The IMFs decomposed by CEMD in the time domain (a) and the frequency domain (b)

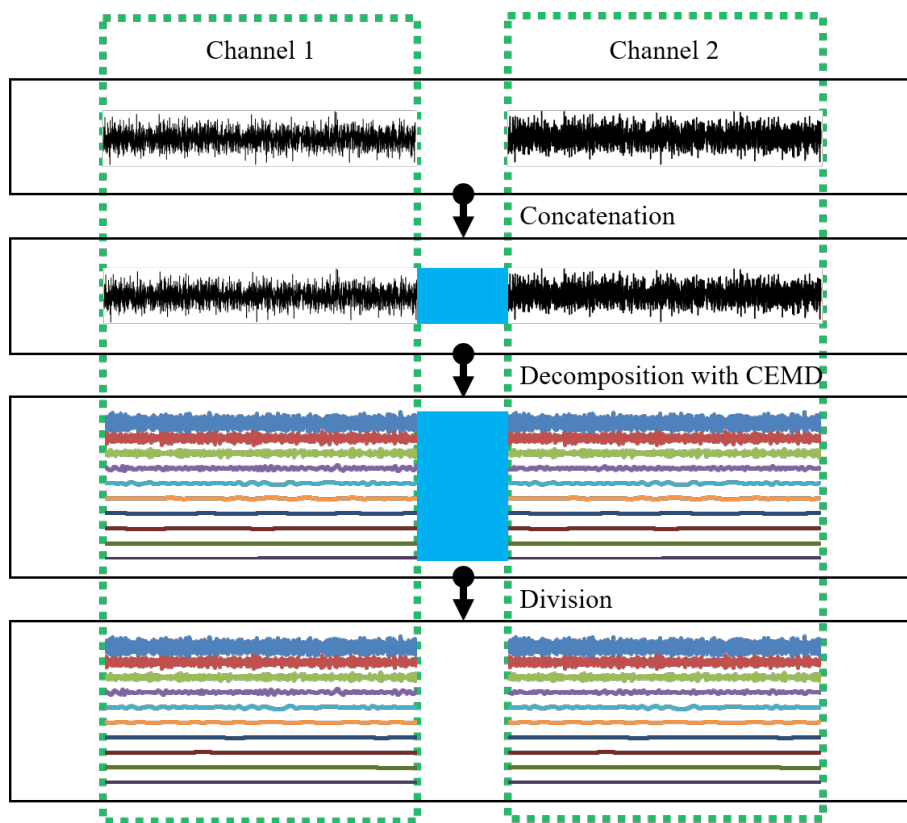


Figure 4.4: Flowchart of SEMD. An example is shown using a two-channel time series. The blue block represents white noise or empty signals containing no data. After signal decomposition, the signals in the blue block are removed from the IMFs.

MEMD MEMD simultaneously decomposes multi-channel time series, as illustrated in Figure 4.5. MEMD projects the multi-channel time series onto a tangent space using the direction vector. The procedure for MEMD is outlined in Algorithm 2, where C denotes the number of channels in the multi-channel time series. In MEMD, the original signals can also be reconstructed by summing the decomposed IMFs.

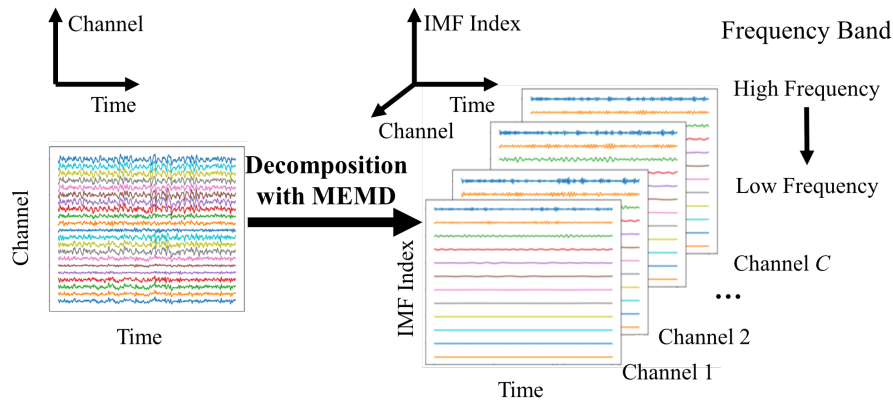


Figure 4.5: Data decomposition with MEMD. MEMD decomposes multi-channel signals into IMFs. The IMFs are located in different frequency bands, and within each decomposed channel, the k -th IMF corresponds to the same frequency band. In this figure, the IMFs are sorted in descending order in the frequency domain.

SEMD-MEMD In SEMD, multi-channel time series are concatenated so that the frequency bands of the IMFs are aligned across all channels. In MEMD, the frequency alignment of the frequency bands is ensured by the tangent space. When decomposing the multi-channel EEG signals in the training set, the frequency bands of all the trials are expected to be aligned. However, in MEMD, the maximum number of channels is limited to sixteen. Therefore, the concept of concatenation in SEMD is incorporated into MEMD and the SEMD-MEMD is proposed, whose flowchart is given in Figure 4.6.

GEMD In the above mentioned methods in the EMD family, the inputs of the decomposition methods are the time series. In the brain signal analysis, the signals are not only the time series but also the structured signals. For instance, the magnetic resonance imaging (MRI) data are morphometric features in multiple brain regions. The morphometric features are the measured properties of a brain region, such as gray matter volume, cortical thickness, surface area, intrinsic curvature, meancurvature, curvature index, and fold index. In the MRI analysis, Figure 4.2 has an alternative form of the inputs. The input signals are the morphometric features and a graph describing the distances between these brain regions. The input signals are then converted into the structural connectivity by Pearson correlation, which has the same format as the functional connectivity. In this case, the SEMD-MEMD can fulfill the requirements of decomposing such a data structure. The GEMD are proposed to be used in the signal decomposition in the MRI data, as

Algorithm 2 Multi-variate empirical mode decomposition

Require: $T \in \mathbb{Z}$

Ensure: $\mathbf{x} \in \mathbb{R}^{T \times C}$

obtain N_{dir} uniformly-distributed sequences ($\theta = \theta_1, \theta_2, \dots, \theta_P$) from Hammersley sequence, $N_{dir} > P$;

cast θ into tangent space and obtain direction vector \mathbf{v} .

while number of extreme values in \mathbf{x} is greater than 3 **do**

for $don = 1 : N_{dir}$

 cast X into tangent space with inner product ($X \cdot \mathbf{v}$)

 locate the positions of extremes in $X \cdot \mathbf{v}$, which are assumed to be the positions of extremes in X

 interpolate the maxima extremes (minima extremes) of \mathbf{X}

 obtain the envelop of extremes e_{max}^n (e_{min}^n) of the n -th uniform-distributed sequence.

end for

$e_{max} = \text{mean}(e_{max}^n)$ for $n = 1 : N_{dir}$

$e_{min} = \text{mean}(e_{min}^n)$ for $n = 1 : N_{dir}$

$\mathbf{m} = (e_{max} + e_{min})/2$

$\mathbf{h} = \mathbf{x} - \mathbf{m}$

if \mathbf{h} is an IMF **then**

 save \mathbf{h} as an IMF

$\mathbf{x} = \mathbf{x} - \mathbf{r}$

else

$\mathbf{x} = \mathbf{h}$

end if

end while

save \mathbf{x} as an IMF

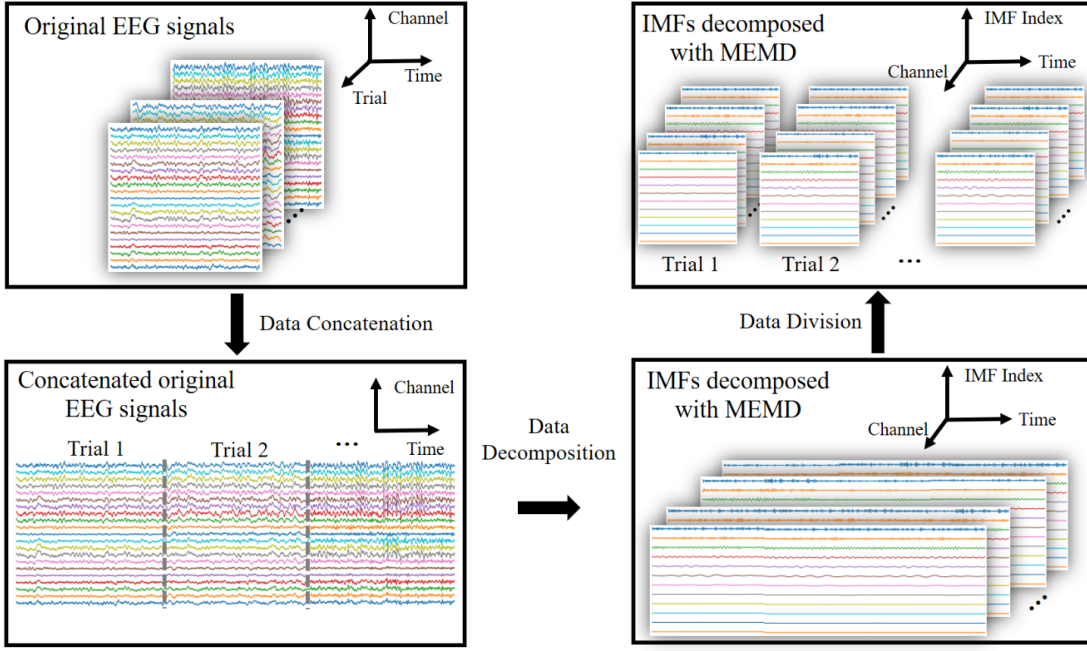


Figure 4.6: Decomposition of multi-channel time series with SEMD-MEMD. Multi-channel EEG signals of multiple trials are first concatenated along the time axis. The concatenated EEG signals are then decomposed with MEMD.

shown in Figure 4.7.

A graph \mathcal{G} consists of the nodes V and the edges E , which is denoted as $\mathcal{G} = (V, E)$. When converting MRI data to a graph, the nodes are brain regions and the edges measure the distances between the brain regions. A graph is expressed as an adjacent matrix A with the given nodes and edges. L is a diagonal matrix whose diagonal elements sum up the elements in each row of A . The Laplacian matrix of a graph \mathcal{G} is computed with $L - A$. The morphometric features are the signals \mathbf{x} stored in the nodes.

GEMD has the same sifting process as the CEMD but the modified local extreme and interpolation. The local extreme is defined on the graph. For the signal \mathbf{x} on V , a node n is a local maxima when $\mathbf{x}(n) > \mathbf{x}(k)$ for all its neighbors k on the graph \mathcal{G} . In the local minima, $\mathbf{x}(n) < \mathbf{x}(k)$ should be fulfilled. In the interpolation, the signal \mathbf{x} consists of the known signal \mathbf{x}_P and the unknown signal \mathbf{x}_Q , which can be denoted as $\mathbf{x} = [\mathbf{x}_P^T, \mathbf{x}_Q^T]^T$. By reordering the node sequence of the Laplacian matrix as in \mathbf{x} , L is presented by partitioned matrix

$$L = \begin{pmatrix} L_P & R \\ R^T & L_Q \end{pmatrix}. \quad (4.5)$$

The interpolation of the unknown signal \mathbf{x}_Q is solved by minimizing the equation

$$\mathbf{x}^T L \mathbf{x} = [\mathbf{x}_P^T, \mathbf{x}_Q^T] \begin{pmatrix} L_P & R \\ R^T & L_Q \end{pmatrix} [\mathbf{x}_P^T, \mathbf{x}_Q^T]^T. \quad (4.6)$$

The solution to this equation is given by $L_Q \mathbf{x}_Q = -R \mathbf{x}_P$.

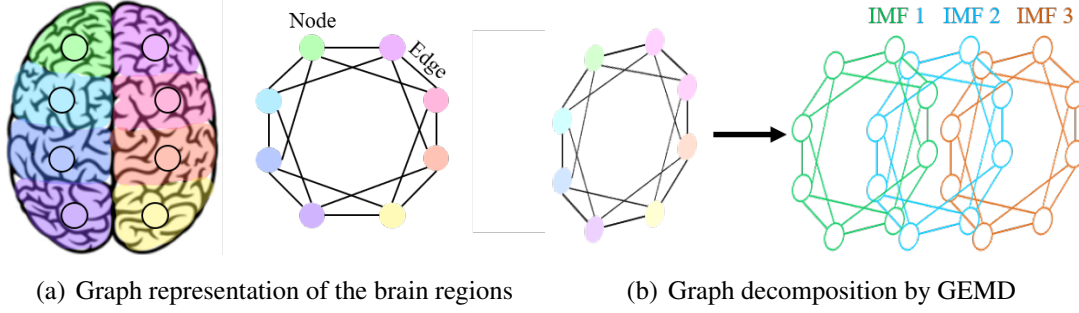


Figure 4.7: Signal decomposition of MRI data. In MRI data, the signals are represented as a graph. The graph edges measure the distances of the brain regions and the graph nodes denote the brain regions. The features of the brain regions are the signals in the nodes. After converting MRI data into a graph, the MRI data are decomposed in a graph format by GEMD.

Signal recombination

In these EMD-based decomposition methods, the original signals can be easily reconstructed by summing the decomposed IMFs. The inverse process is expressed as

$$f'(Y) = f'([y_1, y_2, \dots, y_R]) = \sum_{r=1}^R y_r \quad (4.7)$$

The artificial signal \mathbf{x}' is presented by

$$\mathbf{x}' = f'([y'_1, y'_2, \dots, y'_R]) = \sum_{r=1}^R y'_r \quad (4.8)$$

where y'_r is randomly selected from $\{y_r^1, y_r^2, \dots, y_r^I\}$, $r = 1, 2, \dots, R$, R is the number of IMFs. The flowchart illustrating the generation of an artificial signal is presented in Figure 4.8.

4.1.3. Deep-learning classifiers

Following data augmentation, the training set now contains both original and artificial trials, represented as multi-channel time series. These time series are then transformed into FC matrices using Equation 4.1. To extract features and predict labels from these FC matrices, two neural networks, namely ResNet [57] and BrainNet CNN [58], are utilized.

BrainNet CNN

The BrainNet CNN consists of three blocks: edge-to-edge (E2E), edge-to-node (E2N), and node-to-graph (N2G). These blocks are convolution layers with distinct convolution kernels.

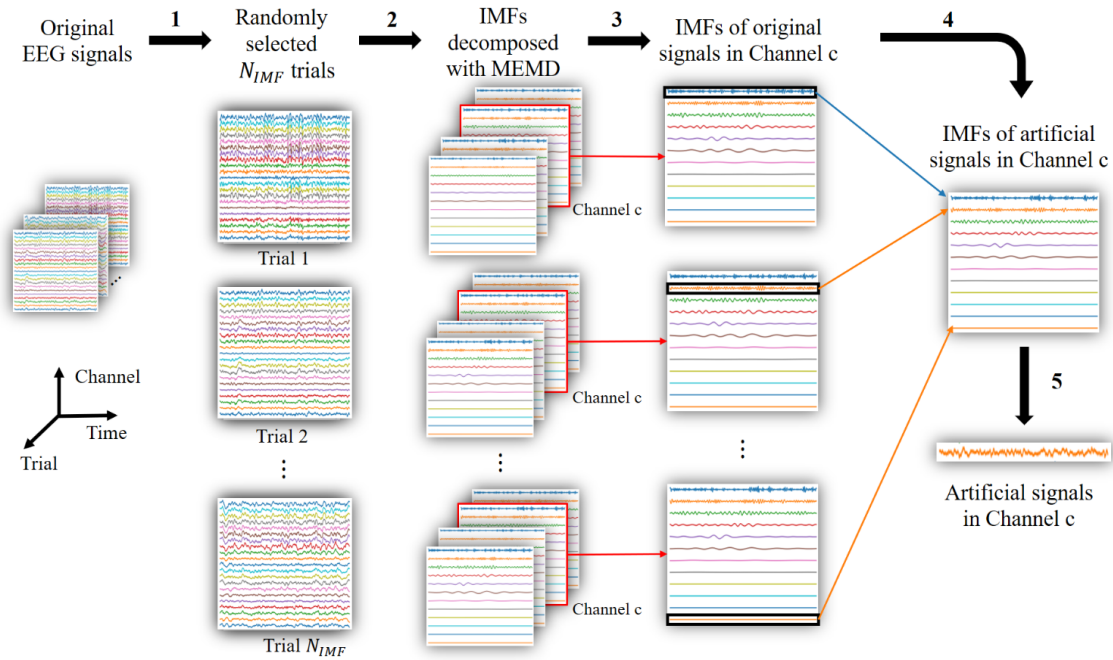


Figure 4.8: Flowchart depicting the recombination of IMFs to generate artificial signals in multi-channel EEG signals using MEMD.

In a graph, an edge represents the connection between two nodes, reflecting the relationship between them. In the context of FC matrices, where the channels represent nodes and the values denote the edges between nodes, the BrainNet CNN processes the FC matrices within a graph framework.

The E2E block learns the connections between edges. It transforms the input FC matrix into a new representation, as illustrated in Figure 4.9. This block utilizes two kernels, $\tau_1 \in \mathbb{R}^{1 \times C}$ and $\tau_2 \in \mathbb{R}^{C \times 1}$, which are applied to the input separately. The outputs of the two convolutions are element-wise summed to produce the output of the E2E block. The E2N block learns the projection from edges to nodes and is equipped with a kernel $\tau \in \mathbb{R}^{1 \times C}$. Finally, in the N2G block, the input is the node information derived from the E2N block. This block converts the node information to the graph information through a convolution with the kernel $\tau \in \mathbb{R}^{C \times 1}$. The structure of the BrainNet CNN is summarized in Table 4.1.

ResNet

Compared to deep CNNs, ResNet is distinguished by its inclusion of skip connections. In deep CNNs, as the number of layers increases, the training process encounters the issue of gradient vanishing during backpropagation. However, skip connections can alleviate the impact of gradient vanishing in deep CNNs. A layer that incorporates a skip connection is referred to as a residual module within the ResNet framework. Figure 4.10 depicts the concept of a skip connection.

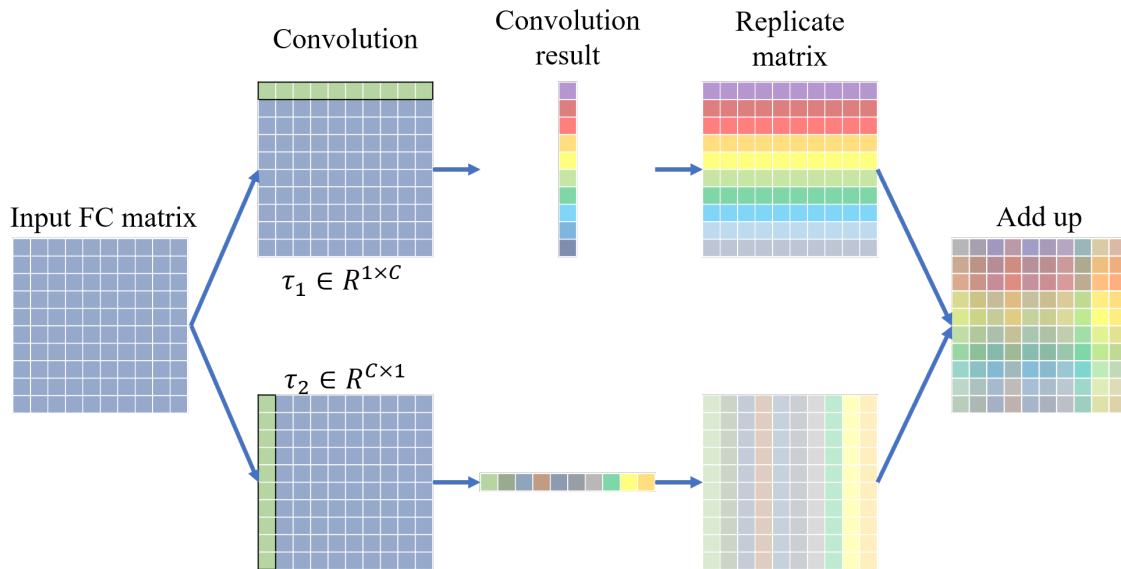


Figure 4.9: The E2E block in BrainNet CNN.

Table 4.1: Model structure of BrainNet CNN

Layer	Output Size	Parameter
Input Layer	$[B, 4, C, C]$	
BatchNorm	$[B, 4, C, C]$	$(C, 1)$
ReLU	$[B, 4, C, C]$	
E2E	$[B, 16, C, C]$	
BatchNorm	$[B, 16, C, C]$	
ReLU	$[B, 16, C, C]$	
E2E	$[B, 32, C, C]$	
ReLU	$[B, 32, C, C]$	
E2N	$[B, 64, C, 1]$	$(1, C)$
N2G	$[B, 512, 1, 1]$	$(C, 1)$
Flatten	$[B, 512]$	
Linear & Softmax	$[B, 2]$	

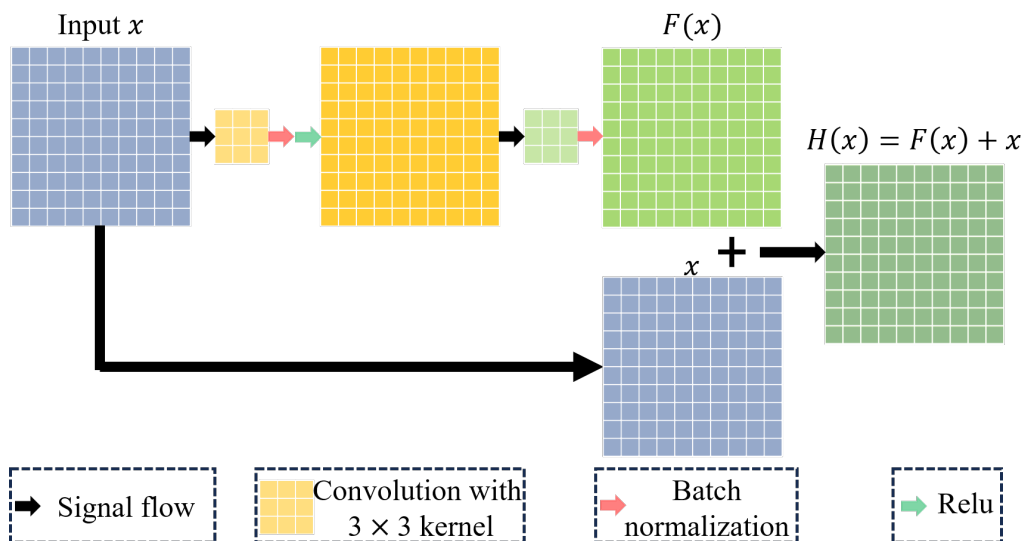


Figure 4.10: The residual module with a skip connection in the ResNet.

In deep CNNs, x and $F(x)$ represent the input and output of a convolutional layer, respectively. By adding the input x to the output $F(x)$, the skip connection generates the output of the residual module as $H(x) = F(x) + x$. This approach facilitates faster training of the deep CNN model, enhances training performance, and effectively addresses the gradient vanishing problem. For evaluating the performance of data augmentation, the ResNet-18 architecture is employed.

4.2. Single-side upper limb movement classification

The EEG signals consist of multi-channel time series characterized by low signal-to-noise ratio, high temporal resolution, and low spatial resolution. The movement-related cortical potential (MRCP) signals, found within the low-frequency bands of EEG signals on the motor cortex, can be extracted through Fourier transformation of the raw EEG signals. Similar to machine-learning classification models for EEG signal processing, the classification model developed for MRCP signal processing also involves three main steps: spatial filtering, feature extraction, and filter bank analysis. This work follows these steps in the development of the MRCP classification model.

4.2.1. Pre-requisition

In this subsection, several fundamental concepts are introduced. These concepts include data preparation techniques, the concept of nested cross validation, and the concept of the grand average MRCP.

Dataset preparation

The dataset used for model evaluation is a publicly available dataset ¹. The EEG signals were downsampled to 256 Hz, and a notch filter at 50 Hz was applied to mitigate the influence of power line interference.

The data acquisition paradigm for this dataset is illustrated in Figure 4.11. The subjects were seated on a chair, facing a screen. EEG signals were captured from scalp electrodes positioned at channels FC_z , C_3 , C_z , C_4 , CP_z , F_3 , F_z , F_4 , P_3 , P_z , and P_4 . At the start of each trial, a cross was displayed on the screen. Two seconds later, a cue indicating the intended motion appeared on the screen. The executed motions included *elbow flexion* (EF), *elbow extension* (EE), *supination* (SU), *pronation* (PR), *hand open* (HO), *hand close* (HC), and the *resting* (RE) state. Simultaneously, the hand trajectory was recorded alongside the EEG signals. The movement onset of each executed motion could be determined using the hand trajectory. The time window utilized for classification purposes

¹<http://bnci-horizon-2020.eu/database/data-sets>, index 25, Upper limb movement decoding from EEG (001-2017)

encompassed one second before the onset to one second after the onset. With repeated movements, a three-dimensional EEG tensor was acquired for each motion.

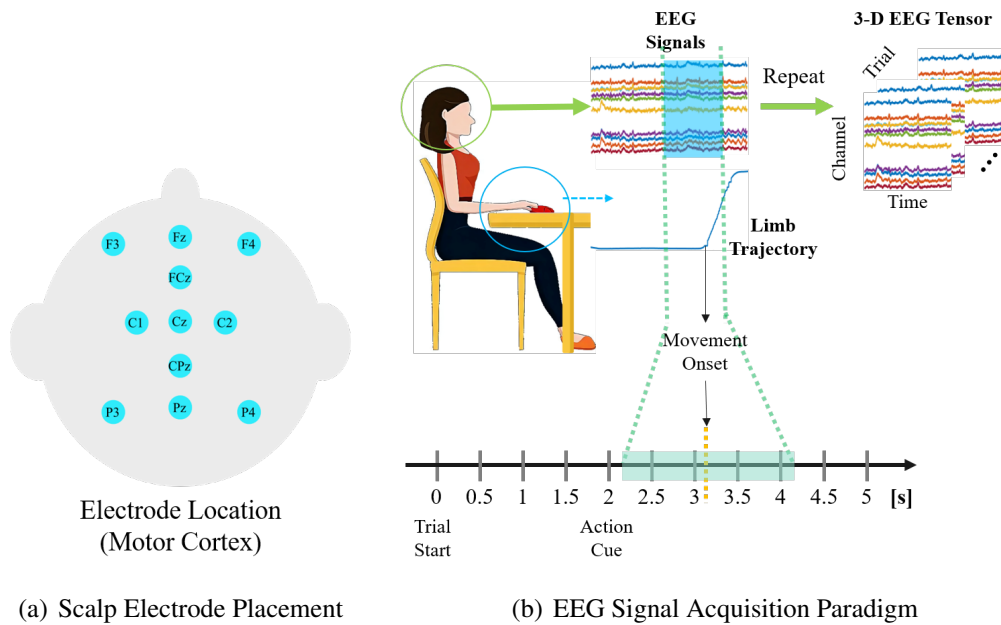


Figure 4.11: EEG signal processing: From raw signals to three-dimensional EEG tensors. The EEG signals are acquired using 11 electrodes. The raw EEG signals around the movement onset are sliced and concatenated to form EEG tensors.

Concept of nested cross-validation

The classification model may involve hyperparameters that cannot be determined during the training process. To address this, nested cross-validation is employed, as shown in Figure 4.12.

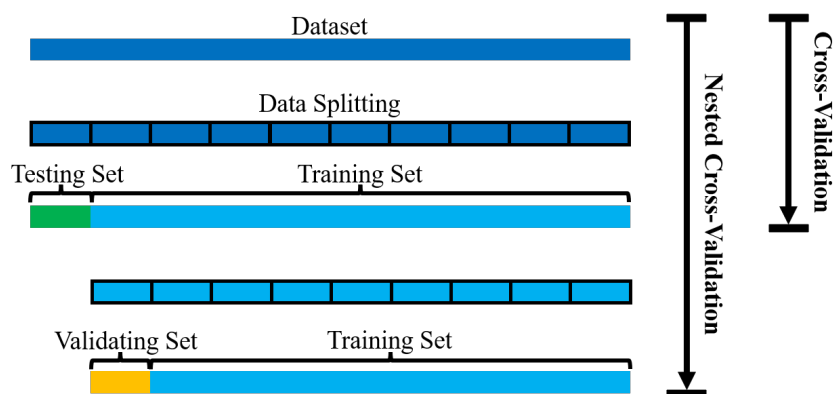


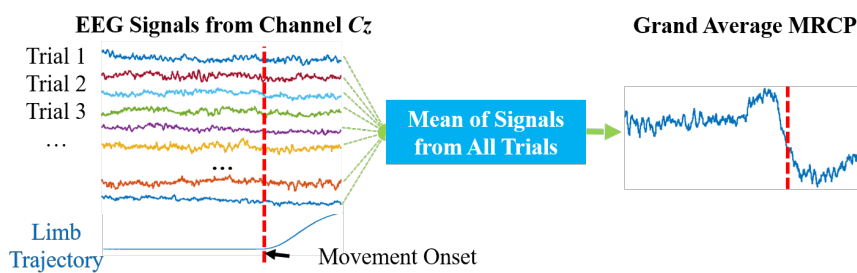
Figure 4.12: Determination of hyperparameters of classification models using nested cross-validation.

The nested cross-validation tackles the problem of hyperparameter determination by splitting the training set into a validation set and a smaller sub-training set. In both the

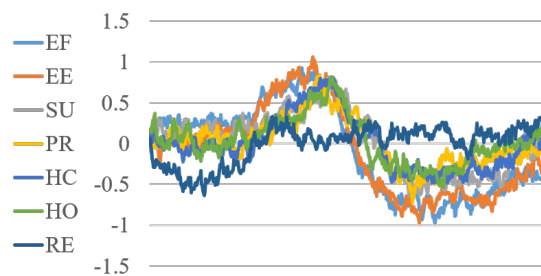
validation set and the new sub-training set, the true labels of all trials are known, allowing for the immediate measurement of classification performance. The hyperparameters associated with the best classification performance are assumed to be the optimal ones. When predicting labels for the testing set, the best hyperparameters will be employed.

Concept of grand average MRCP

To classify motions using MRCP signals, it is essential to understand the differences among these motions in EEG signals. The grand average MRCP represents the averaged EEG signals within the same class, as illustrated in Figure 4.13(a). In different motions, the grand average MRCPs exhibit variations despite having a similar overall trend 4.13(b).



(a) Computation of the grand average MRCP.



(b) Comparison of the grand average MRCP across motions.

Figure 4.13: Concept of grand average MRCP.

Figure 4.14 demonstrates that the grand average MRCP in the low-frequency band (0 Hz-10 Hz) reflects the amplitude increase and decrease of the signals, while the high-frequency band contains noise rather than relevant signals. When computing the grand average MRCP, signal averaging is employed to remove noise and task-unrelated components from each trial. Additionally, it is crucial to establish the relationship between the grand average MRCP and the signal of each trial. Based on this, three key points are summarized for developing the classification models:

- Restrict the signals to the low-frequency band.
- Eliminate task-unrelated components from the trial signals.
- Measure the similarity between the grand average MRCP and the trial signals.

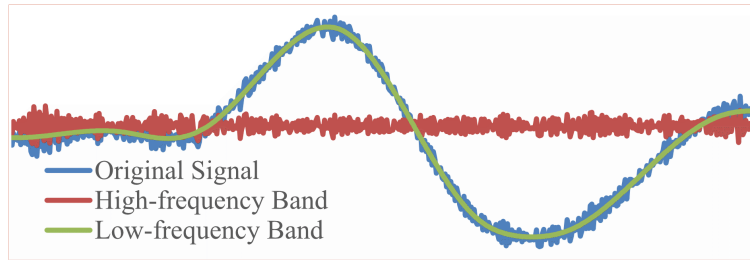


Figure 4.14: Comparison between signals in the low-frequency band and the high-frequency band.

This thesis proposes two classification models: the Filter Bank Task-Related Component Analysis (FBTRCA) and the Two-Step-Training Temporal Spectral Neural Network (TTSNet). The frameworks of these models are illustrated in Figure 4.15. Each model comprises three modules and a dedicated classifier for label prediction. The modules are listed in Table 4.2.

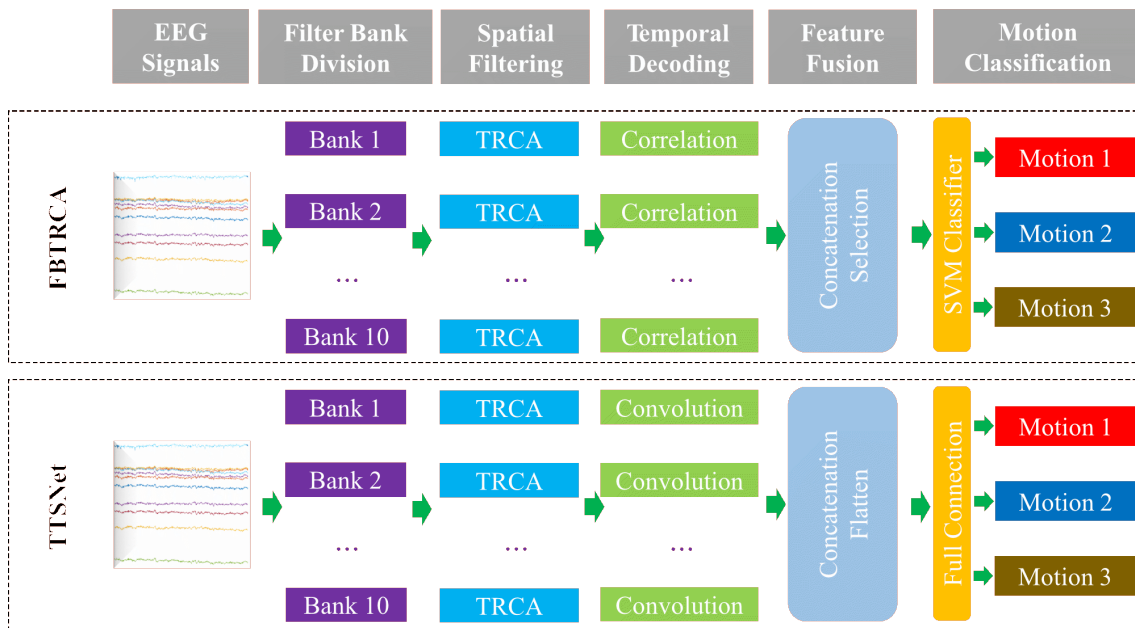


Figure 4.15: The flowchart of FBTRCA and TTSNet in the multi-class classification task.

In both methods, EEG signals are first divided into filter banks and then optimized using the spatial filter to remove task-unrelated components. For decoding the temporal information from the signals, FBTRCA utilizes the correlation between the signals and the grand average MRCs as features. TTSNet employs EEGNet to capture the temporal information and exploits the shift-invariant property of the convolution layers. After decoding the temporal information, the output features from all the filter banks are concatenated.

In FBTRCA, these features are selected and optimized using the Minimum Redundancy Maximum Relevance (MRMR) method and classified using the Support Vector Machine (SVM) method. In TTSNet, the features are flattened and then classified using fully-connected layers.

Table 4.2: An overall view on the modules of the proposed two methods

Module	Method 1 (FBTRCA)	Method 2 (TTSNet)	Function
Spatial filter	Task-related component analysis		Remove task-unrelated component
Correlation feature	Summed-up weight	Shift-invariant weight	Measure the similarity
Filter bank technique	Filter band division		Improve the robustness in the multiple bands
	Feature selection	Feature flatten	
Classifier	Support vector machine	Fully-connected layers	Predict labels

4.2.2. Spatial filter

This section introduces the basic concept of spatial filtering and explains how to calculate the spatial filter for binary and multi-class classification of MRCP signals.

Linear Transform

Spatial filtering applies a linear channel-wise transform to the raw EEG signals $\mathbf{X} \in \mathbb{R}^{C \times T}$, where C is the number of channels and T is the number of sample points. The linear transform operator is the spatial filter, which is denoted as $\mathbf{W} \in \mathbb{R}^{C \times P}$, $P \leq C$. The process of spatial filtering can be denoted as $\mathbf{W}^T \mathbf{X} \in \mathbb{R}^{P \times T}$.

In the signal processing of different brain activities, the spatial filter \mathbf{W} is obtained by optimizing certain relationships within the EEG signals. For example, in motor imagery, which involves left or right limb movement, the signal acquired from the scalp on the opposite side shows a power decrease when the limb moves or imagines the movement. In this case, \mathbf{W} is obtained by maximizing the spatial differences between the two motor imagery classes. On the other hand, steady-state visual evoked potentials are evoked by external visual stimuli with different frequencies, where the power in the visual cortex increases when the stimulus occurs. In visual evoked potentials, the spatial differences are relatively subtle compared to motor imagery. Hence, \mathbf{W} is obtained by maximizing the signal-to-noise ratio of the signals in each class.

Spatial filtering in MRCP classification is more similar to visual evoked potentials. MRCP signals are used to classify movements on a single-side limb. In contrast to motor imagery, which deals with movements on both sides, distinguishing spatial differences in MRCP signals is more challenging. Therefore, the spatial filter in this work focuses on optimizing the signal-to-noise ratio.

Task-related Component Analysis

Task-related component analysis (TRCA) aims to find a linear transform $\omega \in \mathbb{R}^{C \times P}$ that maximizes the intra-trial covariances of trials belonging to the same class. In the k th class, the linear transform is denoted as $\omega_k \in \mathbb{R}^{C \times P_k}$, where k is the index of classes, $k = 1, 2, \dots, K$ and K is the number of classes. The spatial filter W for the classification of MRCP signals is a linear combination of ω_k . The EEG signals of class k in the training set is given as $\mathcal{X}_k = \{\mathbf{X}_1^k, \mathbf{X}_2^k, \dots, \mathbf{X}_{I_k}^k\}$. I_k represents the number of trials of class k . \mathbf{X} are multi-channel EEG signals of size $C \times T$.

The intra-trial covariance is the covariance of the same trials, which can be given by

$$\mathbf{C}_i^k = \mathbf{X}_i^k (\mathbf{X}_i^k)^T, \quad (4.9)$$

The inter-trial covariance is the covariance of two different trials, which can be given by

$$\mathbf{C}_{i,j}^k = \mathbf{X}_i^k (\mathbf{X}_j^k)^T + \mathbf{X}_j^k (\mathbf{X}_i^k)^T. \quad (4.10)$$

TRCA aims to find ω which maximizes the inter-trial covariance

$$\max_{\omega_k} J^k = \frac{\omega_k^T \mathbf{S}^k \omega_k}{\omega_k^T \mathbf{Q}^k \omega_k}. \quad (4.11)$$

\mathbf{S}^k is the sum of inter-trial covariances of class k

$$\mathbf{S}^k = \sum_{i,j=1, i < j}^{I_k} \mathbf{C}_{i,j}^k, \quad (4.12)$$

and \mathbf{Q}^k is the sum of the intra-trial covariances of class k

$$\mathbf{Q}^k = \sum_{i=1}^{I_k} \mathbf{C}_i^k. \quad (4.13)$$

The eigenequation $\max_{\omega} J^k$ can be solved with the generalized Schur decomposition as the generalized eigenvalue problem.

Architecture in the binary classes In the binary classification, we have $\omega_1 \in \mathbb{R}^{C \times P_1}$ and $\omega_2 \in \mathbb{R}^{C \times P_2}$. In most cases, it fulfills the assumption that $P_1 + P_2 \leq C$. Therefore, the spatial filter W can be obtained by concatenating ω_1 and ω_2 ,

$$\mathbf{W} = [\omega_1, \omega_2] \in \mathbb{R}^{C \times (P_1 + P_2)}. \quad (4.14)$$

Architecture in the multi-class classes In the multi-class classification, we have $\omega_k \in \mathbb{R}^{C \times P_k}$, $k = 1, 2, \dots, K$. As the number of classes increases, we cannot always ensure that $P_1 + P_2 + \dots + P_K \leq C$. Considering that the spatial distributions of signals on the scalp are similar when the movements are executed on the single-side limb, we assume that all

the classes have similar linear transforms ω_k . Therefore, the spatial filter $\mathbf{W} \in \mathbb{R}^{C \times P}$ can be obtained by solving the summed-up Equation 4.11,

$$\max_{\mathbf{W}} J^k = \frac{\mathbf{W}^T \mathbf{S} \mathbf{W}}{\mathbf{W}^T \mathbf{Q} \mathbf{W}}. \quad (4.15)$$

where $\mathbf{S} = \sum_{k=1}^K \mathbf{S}^k$ and $\mathbf{Q} = \sum_{k=1}^K \mathbf{Q}^k$. The value of P is determined by the nested cross-validation in the experiments.

4.2.3. Correlation feature

The correlation features can be utilized to quantify the similarity between the grand average MRCPs and the signals of individual trials. A high degree of similarity between the grand average MRCP and the trial signal indicates the true label of the trial. In the proposed methods, this similarity is transformed into correlation features and employed in the classifiers.

Summed-up weighted features

The grand average MRCP is the mean of MRCP signals across trials. The grand average MRCP of \mathcal{X}_k is denote as

$$\hat{\mathbf{X}}^k = \sum_{i=1}^{I_k} \mathbf{X}_i^k / I_k. \quad (4.16)$$

The correlation feature measures the correlation between the EEG signals \mathbf{X} and the grand average MRCP $\hat{\mathbf{X}}^k$. Because the classification of single-side limb movements highly relies on the differences of $\hat{\mathbf{X}}^k$. To reduce the similarity of the grand average MRCPs, the mean of $\hat{\mathbf{X}}^k$ is removed from \mathbf{X} and $\hat{\mathbf{X}}^k$.

$$\mathbf{X} \rightarrow \mathbf{X} - \frac{1}{K} \sum_{k=1}^K \hat{\mathbf{X}}^k; \hat{\mathbf{X}}^k \rightarrow \hat{\mathbf{X}}^k - \frac{1}{K} \sum_{k=1}^K \hat{\mathbf{X}}^k. \quad (4.17)$$

After reducing the similarity, three correlations are calculated between \mathbf{X} and $\hat{\mathbf{X}}^k$.

(1) Correlation between \mathbf{X} and $\hat{\mathbf{X}}^k$:

$$\mathbf{X}_* = \mathbf{X}; \mathbf{X}_k = \hat{\mathbf{X}}^k; \quad (4.18)$$

$$\rho_{1,k} = \text{corr}(\mathbf{X}_*^T \mathbf{W}, \mathbf{X}_k^T \mathbf{W}); \quad (4.19)$$

(2) Correlation between \mathbf{X} and $\hat{\mathbf{X}}^k$ after canonical correlation analysis:

$$\mathbf{X}_* = \mathbf{X}; \mathbf{X}_k = \hat{\mathbf{X}}^k; \quad (4.20)$$

$$[\mathbf{A}_k, \mathbf{B}_k] = \text{cca}(\mathbf{X}_*^T \mathbf{W}, \mathbf{X}_k^T \mathbf{W}); \quad (4.21)$$

$$\rho_{2,k} = \text{corr}(\mathbf{X}_*^T \mathbf{W} \mathbf{B}_k, \mathbf{X}_k^T \mathbf{W} \mathbf{B}_k); \quad (4.22)$$

(3) Correlation between $X - \hat{X}^k$ and $\hat{X}^{3-k} - \hat{X}^k$ after canonical correlation analysis:

$$X_* = X - \hat{X}^k; X_k = \frac{1}{K-1} \sum_{kk=1, kk \neq k}^K \hat{X}^{kk} - \hat{X}^k; \quad (4.23)$$

$$[A_k, B_k] = cca(X_*^T W, X_k^T W); \quad (4.24)$$

$$\rho_{3,k} = corr(X_*^T W A_k, X_k^T W A_k); \quad (4.25)$$

In the above equations, the symbol *corr* represents the two-dimensional Pearson correlation coefficient, while the function *cca* computes the canonical coefficients for the two input data matrices.

The Pearson correlation coefficient requires that the two input signals be normalized using z-normalization. Hence, the raw EEG signals undergo z-normalization before undergoing low-frequency bandpass filtering and spatial filtering.

Given z-normalized matrices *A* and *B* with *N* observations, the Pearson correlation coefficient can be calculated as follows:

$$corr(A, B) = \frac{1}{N-1} \sum_{i=1}^N A_i * B_i, \quad (4.26)$$

where A_i and B_i represent the *i*-th elements of matrices *A* and *B*, respectively.

In the context of our correlation features, matrix *A* pertains to the EEG signals, while matrix *B* corresponds to the grand average MRCPs. Matrix *B* is a matrix of pre-trained constants. Consequently, the three types of correlations can be interpreted as weighted features, where the grand average MRCPs serve as the weights.

Shift-invariant weighted features

Compared to using grand average MRCPs as weights, the shifted-invariant weight is more versatile as MRCP signals may exhibit temporal shifts. With the advancements in deep learning techniques, convolutional neural networks (CNNs) have become widely employed in signal processing. The shift-invariant weight in CNN layers can effectively accommodate the processing requirements of shifted signals.

In CNNs, shift-invariance refers to the property of being insensitive to small displacements in input signals. This shift-invariance is achieved through shared weights, also known as convolution filters. If $x(t) \in \mathbb{R}^T$ represents the input and $w(t) \in \mathbb{R}^L$ represents the shared weight, the convolution operation is denoted as:

$$y(t) = x(t) * w(t) = \sum_{s=0}^t x(t-s)w(s). \quad (4.27)$$

Here, the weight $w(t)$ is shared for $t = 1, 2, \dots, T$. The small displacement in the input $x(t)$ has minimal impact on the shared weight because the shared weight learns a more generalizable detector. However, a learned weight $w(t)$ cannot completely replace the role of the grand average MRCP in correlation features. Therefore, multiple convolution filters are employed in the CNN model.

Table 4.3 presents an EEGNet model that can be used as a replacement for the correlation coefficients, thereby enhancing the robustness of temporal decoding [59]. In this table, B denotes the batch size.

Table 4.3: Model structure of the adopted CNN layers

Layer	Output Size	Parameter
Input Layer	$[B, 1, C, T]$	
ZeroPad2d	$[B, 1, C, T+63]$	(31, 32, 0, 0)
Conv2d	$[B, 8, C, T]$	(1, 64)
BatchNorm2d	$[B, 8, C, T]$	
Conv2d	$[B, 16, 1, T]$	$(C, 1)$, <i>grouped</i>
BatchNorm2d	$[B, 16, 1, T]$	
ELU	$[B, 16, 1, T]$	
AvgPool2d	$[B, 16, 1, T//4]$	(1, 4)
Dropout	$[B, 16, 1, T//4]$	0.25
ZeroPad2d	$[B, 16, 1, T//4+15]$	(7, 8, 0, 0)
Conv2d	$[B, 16, 1, T//4]$	(1, 15), <i>grouped</i>
Conv2d	$[B, 16, 1, T//4]$	(1, 1)
BatchNorm2d	$[B, 16, 1, T//4]$	
ELU	$[B, 16, 1, T//4]$	
AvgPool2d	$[B, 16, 1, T//32]$	(1, 8)
Dropout	$[B, 16, 1, T//32]$	0.25
Flatten	$[B, 16*T//32]$	
Linear	$[B, K]$	<i>bias = False</i>

4.2.4. Filter bank technique

The correlation features obtained after spatial filtering may not exhibit robustness in the frequency domain. The classification performance can vary across different frequency bands, and it is challenging to identify the frequency band that yields the best performance. Furthermore, the optimal frequency band can vary among individuals due to individual differences. To address these issues, a filter bank technique is employed to enhance the robustness of the proposed method.

The filter bank technique involves extracting features from multiple frequency bands, and these features are subsequently concatenated for classification using classifiers. There are three key considerations in the filter bank technique: (1) determining the appropriate multiple frequency bands to be utilized, and (2) handling the increased number of features

prior to classification.

Filter Bank Division

The initial step of the filter bank technique involves dividing the signals into multiple frequency bands. Figure 4.16 illustrates three potential frequency range settings for these bands. In setting M1, the frequency range consists of equal-length bands arranged in an arithmetic sequence. Setting M2 also employs arithmetic sequences, but the low cut-off frequency and high cut-off frequency are derived from different common differences. In setting M3, the low cut-off frequency remains constant while the high cut-off frequency follows an arithmetic sequence.

As depicted in Figure 4.14, the grand average MRCP is partitioned into the low-frequency band and the high-frequency band. When separating the high-frequency band from the low-frequency band, the overall trend is eliminated. However, the variations among the grand average MRCPs of different movements are reflected in the overall trend. Consequently, it is essential to preserve the signal in the frequency band that closely resembles zero. Therefore, the frequency range M3 is employed for dividing the EEG signals into multiple bands.

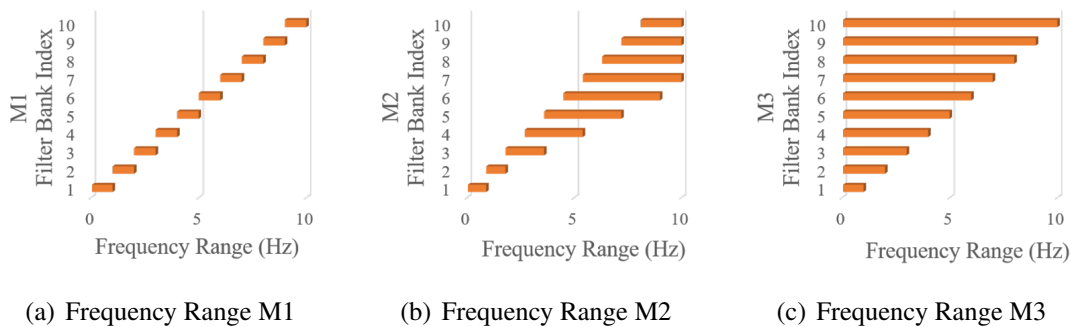


Figure 4.16: Three frequency range settings in filter bank division. Because the differences of the grand average MRCPs are reflected in the low-frequency band, as shown in Figure 4.13(a) and Figure 4.14, the frequency range M3 is used in the filter bank division.

Feature Selection and Classification

In each filter bank, features are extracted using the spatial filter and correlation. These features from multiple bands are then concatenated into a feature vector, resulting in a larger size compared to features extracted from a single band. However, not all features in this concatenated feature vector significantly contribute to the classification performance. Moreover, the large number of features can introduce computational overhead for the classifiers.

In the FBTRCA method, feature selection is employed based on mutual information, in conjunction with a support vector machine (SVM) classifier. A feature selec-

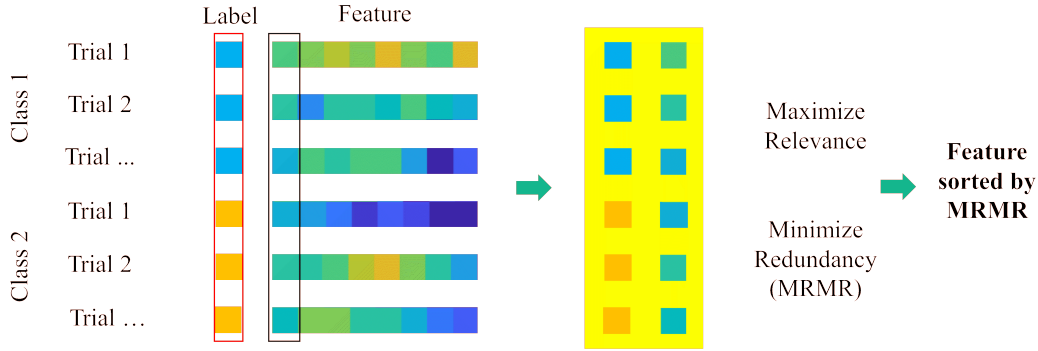


Figure 4.17: Feature selection using mutual information. Relevance refers to the mutual information between a single feature and the corresponding label. Redundancy represents the mutual information between two features.

tion method called minimum redundancy-maximum relevance (MRMR) is utilized, as depicted in Figure 4.17 [60]. MRMR aims to rank the features by maximizing relevance (measured through mutual information between a feature and the label) and minimizing redundancy (measured through mutual information between two features). The best feature according to MRMR is assumed to possess the maximum relevance and minimum redundancy. After identifying the best feature, the remaining features are considered as a new set, and MRMR continues to find the new best feature within this set. This iterative process allows the features to be sorted in sequence. The selected features by MRMR are subsequently employed for classification using the SVM classifier. In this classifier, features are mapped into a high-dimensional feature space utilizing kernel tricks, enabling more complex discrimination between non-convex sets compared to the original space.

In the TTSNet method, fully-connected layers are utilized for feature reduction and classification simultaneously. These layers can effectively decrease the feature dimensionality while performing classification. The model structure of the adopted fully-connected layers is presented in Table 4.4, where F denotes the number of filter banks.

Table 4.4: Model structure of fully-connected layers in TTSNet

Layer	Output Size	Parameter
Input Layer	$[B, K, F]$	
Flatten	$[B, K * F]$	
Linear	$[B, K * F * 2]$	$bias = False$
Relu	$[B, K * F * 2]$	
Linear	$[B, K * F / 2]$	$bias = False$
Relu	$[B, K * F / 2]$	
Linear	$[B, K]$	$bias = False$

4.3. Conclusion

This thesis presents four methods for the application of EEG signals. As non-invasive signals, EEG signals have two main applications: (1) brain disease analysis and (2) human-robot interaction. In brain disease analysis, EEG signals can be used to identify brain differences between patients and healthy individuals. When applying deep learning methods to disease detection, one of the challenges is the insufficient training data, leading to overfitting of the trained deep learning models. This thesis proposes a data augmentation method based on a decomposition and recombination strategy to increase the size of the training set in small-size AD datasets. In human-robot interaction, EEG signals are utilized to analyze the brain activities of healthy individuals. By classifying the states of the brain, the brain-computer interface generates control commands for external devices. Three methods proposed in this thesis focus on the classification of brain states evoked by single-side upper limb movements.

Data augmentation task in small-size AD datasets EEG signals can be utilized for the detection of AD, involving a binary classification between AD patients and healthy individuals. To facilitate this classification, the EEG signals are transformed into FC matrices. Subsequently, neural networks such as ResNet and BrainNet CNN are employed to extract pertinent features from these FC matrices. However, due to limited data availability from patients, the trained neural networks tend to be overly fitted and encounter the issue of overfitting. To address this problem, a proposed data augmentation method is introduced to expand the training set. This method is founded on a decomposition and recombination strategy. In the decomposition phase, the original signals within the training set are broken down into multiple IMF components. During the recombination process, these IMF components are randomly selected and combined to generate artificial signals. In the training of neural networks, both the original signals and the artificially generated signals are employed to optimize the weights within the neural networks.

Classification task of single-side upper limb movement The brain activity differences evoked by motions involving single-side upper limb movement are reflected in the grand average MRCPs. The grand average MRCP represents the averaged EEG signals across trials, effectively eliminating noise in the EEG signals through signal averaging. Importantly, the signals in the low-frequency bands are preserved even after averaging. Moreover, the grand average MRCPs exhibit variations across different motions. Three key points can be derived from the characteristics of the grand average MRCP: the importance of low-frequency bands, the removal of task-unrelated signals in the raw EEG signals, and the measurement of similarity between the grand average MRCPs and the EEG trials.

The first method employed in this study is a binary classification approach. To enhance the robustness of classification performance within the low-frequency bands, a filter bank technique is proposed. Initially, the EEG signals are divided into multiple

low-frequency bands. The low cut-off frequencies for these bands are set to a constant value close to zero, while the high cut-off frequencies are arranged in an arithmetic sequence. Within each band, the spatial filter task-related component analysis is applied to eliminate task-unrelated components and noise from the raw signals. Additionally, correlation features are employed to gauge the similarity between the EEG trials and the grand average MRCPs. These correlation features from all bands are concatenated, and a minimum redundancy-maximum relevance approach is utilized to select the essential features from the concatenated set. Finally, the essential features are classified using a support vector machine classifier.

In the second method, the architecture of the first method is slightly modified to accommodate the requirements of a multi-class classification task. Like the first method, the second method consists of three modules: filter bank technique, spatial filter, and correlation feature. However, in the second method, both the spatial filter and correlation features are enhanced to suit the multi-class classification scenario. In the first method, the spatial filter was obtained by concatenating the eigenvectors of two classes. However, with an increased number of classes, using concatenation in the calculation would lead to an increase in the dimensionality of the spatial filter. Moreover, it was observed that the active channels exhibited significant similarity when different movements were executed on the single-side limb. As a result, it was assumed that these motions shared a similar spatial distribution, allowing for the sharing of spatial filters across multiple classes. Based on this assumption, the spatial filter is obtained by solving the eigenequation given in Equation 4.15. The classification of single-side upper limb movements heavily relies on the differences observed in the grand average MRCPs. To reduce the influence of common components in the grand average MRCPs, the mean of all grand average MRCPs is subtracted from the grand average MRCPs of all classes within the correlation features.

In the correlation features, the grand average MRCPs can be considered as pre-trained weights, as they contribute to the weighted summation of the EEG signals. However, in the third method, the weights used in the correlation are replaced with shift-invariant weights. This shift-invariant weighting in the convolutional layers addresses the issue of signal shift along the time axis. In situations where the movement onsets of limb movements cannot be accurately located, such as when the movement trajectory cannot be obtained, biases may exist between two EEG trials along the time axis. Consequently, the grand average MRCPs may not adequately represent the weights within the correlation features. By introducing shift-invariant weights, to some extent, the influences caused by such biases can be mitigated. The temporal decoding architecture employed in this method is EEGNet, which is a type of Convolutional Neural Network (CNN).

4.4. Future work

The following future research directions can be pursued based on the findings and limitations identified in this dissertation:

- *Data Augmentation Effects on FC Matrices* In the context of small-size AD datasets, this study explored data augmentation methods to enhance EEG signals in the training set. However, the impact of data augmentation on the resulting FC matrices, which play a crucial role in AD detection due to their spatial features, remains unexplored. Future research should examine the effects of data augmentation on FC matrices to assess its potential for improving AD detection accuracy.
- *Spatial Differences in MRCP Signals* The multi-class classification methods FBTRCA and TTSNet assume a similar spatial distribution of brain activities for all motions on the single limb. However, it is evident that the spatial distributions are not entirely uniform. Future work should focus on developing classification methods that can effectively exploit the potential spatial differences in MRCP signals to enhance classification accuracy.
- *Overcoming Similarities in Grand Average MRCPs* The performance of the FBTRCA classification method heavily relies on differences in the grand average MRCPs. However, certain motions, such as elbow flexion and elbow extension, exhibit high similarities in their grand average MRCPs, leading to poor classification performance between these motions. Future research should explore strategies to address this issue and improve the classification accuracy for similar MRCP patterns.
- *Generalization to Electromyography (EMG) Classification* Beyond the realm of brain-computer interfaces, FBTRCA has the potential to be applied in the classification of motions using electromyography (EMG) signals. Similar to FBTRCA, EMG classification requires consideration of low-frequency bands and high correlation between trials. This promising application of FBTRCA in EMG classification has not been explored yet and warrants further investigation.
- *Exploration on Continuous Movements* FBTRCA is designed to classify the motions on the same limb. These motions are independent to each other. When decoding the continuous movement on the same limb, however, the influences are unclear on the brain activities between the coupled dependent motions. This pattern remains to be explored based on further datasets and experiments.

It is crucial to consider the feasibility of these future research directions in terms of available resources, ethical considerations, and practical constraints. Researchers should conduct thorough evaluations of potential challenges and devise appropriate strategies to address them effectively.

BIBLIOGRAPHY

- [1] M. Ciotti, M. Ciccozzi, A. Terrinoni, W.-C. Jiang, C.-B. Wang, and S. Bernardini, “The COVID-19 pandemic,” *Critical Reviews in Clinical Laboratory Sciences*, vol. 57, no. 6, pp. 365–388, 2020, Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/10408363.2020.1783198>. doi: [10.1080/10408363.2020.1783198](https://doi.org/10.1080/10408363.2020.1783198). [Online]. Available: <https://doi.org/10.1080/10408363.2020.1783198>.
- [2] T. Abel and D. McQueen, “The COVID-19 pandemic calls for spatial distancing and social closeness: Not for social distancing!” en, *International Journal of Public Health*, vol. 65, no. 3, pp. 231–231, 2020. doi: [10.1007/s00038-020-01366-7](https://doi.org/10.1007/s00038-020-01366-7). [Online]. Available: <http://link.springer.com/10.1007/s00038-020-01366-7>.
- [3] S. N. Williams, C. J. Armitage, T. Tampe, and K. Dienes, “Public perceptions and experiences of social distancing and social isolation during the COVID-19 pandemic: A UK-based focus group study,” en, *BMJ Open*, vol. 10, no. 7, e039334, 2020. doi: [10.1136/bmjopen-2020-039334](https://doi.org/10.1136/bmjopen-2020-039334). [Online]. Available: <https://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2020-039334>.
- [4] T. L. D. Huynh, “Does culture matter social distancing under the COVID-19 pandemic?” en, *Safety Science*, vol. 130, p. 104 872, 2020. doi: [10.1016/j.ssci.2020.104872](https://doi.org/10.1016/j.ssci.2020.104872). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0925753520302691>.
- [5] X. Liu and S. Zhang, “COVID-19: Face masks and human-to-human transmission,” en, *Influenza and Other Respiratory Viruses*, vol. 14, no. 4, pp. 472–473, 2020. doi: [10.1111/irv.12740](https://doi.org/10.1111/irv.12740). [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1111/irv.12740>.
- [6] E. Bontempi, “The europe second wave of COVID-19 infection and the Italy “strange” situation,” en, *Environmental Research*, vol. 193, p. 110 476, 2021. doi: [10.1016/j.envres.2020.110476](https://doi.org/10.1016/j.envres.2020.110476). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0013935120313736>.
- [7] M. Javaid, A. Haleem, A. Vaish, R. Vaishya, and K. P. Iyengar, “Robotics Applications in COVID-19: A Review,” *Journal of Industrial Integration and Management*, vol. 05, no. 04, pp. 441–451, 2020, _eprint: <https://doi.org/10.1142/S2424862220300033>. doi: [10.1142/S2424862220300033](https://doi.org/10.1142/S2424862220300033). [Online]. Available: <https://doi.org/10.1142/S2424862220300033>.

- [8] X. V. Wang and L. Wang, “A literature survey of the robotic technologies during the COVID-19 pandemic,” en, *Journal of Manufacturing Systems*, vol. 60, pp. 823–836, 2021. doi: [10.1016/j.jmsy.2021.02.005](https://doi.org/10.1016/j.jmsy.2021.02.005). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0278612521000339>.
- [9] J. Miseikis *et al.*, “Lio-A Personal Robot Assistant for Human-Robot Interaction and Care Applications,” en, *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5339–5346, 2020. doi: [10.1109/LRA.2020.3007462](https://doi.org/10.1109/LRA.2020.3007462). [Online]. Available: <https://ieeexplore.ieee.org/document/9134708/>.
- [10] H. Kim, K. K. F. So, and J. Wirtz, “Service robots: Applying social exchange theory to better understand human–robot interactions,” en, *Tourism Management*, vol. 92, p. 104537, 2022. doi: [10.1016/j.tourman.2022.104537](https://doi.org/10.1016/j.tourman.2022.104537). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0261517722000504>.
- [11] G. Michalos *et al.*, “Seamless human robot collaborative assembly – An automotive case study,” en, *Mechatronics*, vol. 55, pp. 194–211, 2018. doi: [10.1016/j.mechatronics.2018.08.006](https://doi.org/10.1016/j.mechatronics.2018.08.006). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0957415818301326>.
- [12] R. De Kervenoael, R. Hasan, A. Schwob, and E. Goh, “Leveraging human-robot interaction in hospitality services: Incorporating the role of perceived value, empathy, and information sharing into visitors’ intentions to use social robots,” en, *Tourism Management*, vol. 78, p. 104042, 2020. doi: [10.1016/j.tourman.2019.104042](https://doi.org/10.1016/j.tourman.2019.104042). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0261517719302407>.
- [13] C. Esterwood and L. Robert, “Robots and COVID-19: Re-imagining Human–Robot Collaborative Work in Terms of Reducing Risks to Essential Workers,” en, *SSRN Electronic Journal*, 2021. doi: [10.2139/ssrn.3767609](https://doi.org/10.2139/ssrn.3767609). [Online]. Available: <https://www.ssrn.com/abstract=3767609>.
- [14] P. Polygerinos *et al.*, “Soft Robotics: Review of Fluid-Driven Intrinsically Soft Devices; Manufacturing, Sensing, Control, and Applications in Human-Robot Interaction: Review of Fluid-Driven Intrinsically Soft Robots,” en, *Advanced Engineering Materials*, vol. 19, no. 12, p. 1700016, 2017. doi: [10.1002/adem.201700016](https://doi.org/10.1002/adem.201700016). [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/adem.201700016>.
- [15] A. Mohebbi, “Human-Robot Interaction in Rehabilitation and Assistance: A Review,” en, *Current Robotics Reports*, vol. 1, no. 3, pp. 131–144, 2020. doi: [10.1007/s43154-020-00015-4](https://doi.org/10.1007/s43154-020-00015-4). [Online]. Available: <https://link.springer.com/10.1007/s43154-020-00015-4>.
- [16] M. Alimardani and K. Hiraki, “Passive Brain-Computer Interfaces for Enhanced Human-Robot Interaction,” en, *Frontiers in Robotics and AI*, vol. 7, p. 125, 2020. doi: [10.3389/frobt.2020.00125](https://doi.org/10.3389/frobt.2020.00125). [Online]. Available: <https://www.frontiersin.org/article/10.3389/frobt.2020.00125/full>.

- [17] A. Cangelosi and S. Invitto, “Human-Robot Interaction and Neuroprosthetics: A review of new technologies,” en, *IEEE Consumer Electronics Magazine*, vol. 6, no. 3, pp. 24–33, 2017. doi: [10.1109/MCE.2016.2614423](https://doi.org/10.1109/MCE.2016.2614423). [Online]. Available: <https://ieeexplore.ieee.org/document/7948880/>.
- [18] K. Gui, H. Liu, and D. Zhang, “Toward Multimodal Human–Robot Interaction to Enhance Active Participation of Users in Gait Rehabilitation,” en, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 2054–2066, 2017. doi: [10.1109/TNSRE.2017.2703586](https://doi.org/10.1109/TNSRE.2017.2703586). [Online]. Available: <https://ieeexplore.ieee.org/document/7926461/>.
- [19] A. E. Hramov, V. A. Maksimenko, and A. N. Pisarchik, “Physical principles of brain–computer interfaces and their applications for rehabilitation, robotics and control of human brain states,” en, *Physics Reports*, vol. 918, pp. 1–133, 2021. doi: [10.1016/j.physrep.2021.03.002](https://doi.org/10.1016/j.physrep.2021.03.002). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0370157321001095>.
- [20] X. Gao, Y. Wang, X. Chen, and S. Gao, “Interface, interaction, and intelligence in generalized brain–computer interfaces,” en, *Trends in Cognitive Sciences*, vol. 25, no. 8, pp. 671–684, 2021. doi: [10.1016/j.tics.2021.04.003](https://doi.org/10.1016/j.tics.2021.04.003). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1364661321000966>.
- [21] S. N. Flesher *et al.*, “A brain-computer interface that evokes tactile sensations improves robotic arm control,” en, *Science*, vol. 372, no. 6544, pp. 831–836, 2021. doi: [10.1126/science.abd0380](https://doi.org/10.1126/science.abd0380). [Online]. Available: <https://www.science.org/doi/10.1126/science.abd0380>.
- [22] A. Kawala-Sterniuk *et al.*, “Summary of over Fifty Years with Brain-Computer Interfaces—A Review,” en, *Brain Sciences*, vol. 11, no. 1, p. 43, 2021. doi: [10.3390/brainsci11010043](https://doi.org/10.3390/brainsci11010043). [Online]. Available: <https://www.mdpi.com/2076-3425/11/1/43>.
- [23] A. N. Belkacem, N. Jamil, J. A. Palmer, S. Ouhbi, and C. Chen, “Brain Computer Interfaces for Improving the Quality of Life of Older Adults and Elderly Patients,” en, *Frontiers in Neuroscience*, vol. 14, p. 692, 2020. doi: [10.3389/fnins.2020.00692](https://doi.org/10.3389/fnins.2020.00692). [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnins.2020.00692/full>.
- [24] S. Aggarwal and N. Chugh, “Review of Machine Learning Techniques for EEG Based Brain Computer Interface,” en, *Archives of Computational Methods in Engineering*, vol. 29, no. 5, pp. 3001–3020, 2022. doi: [10.1007/s11831-021-09684-6](https://doi.org/10.1007/s11831-021-09684-6). [Online]. Available: <https://link.springer.com/10.1007/s11831-021-09684-6>.
- [25] K. Värbu, N. Muhammad, and Y. Muhammad, “Past, Present, and Future of EEG-Based BCI Applications,” en, *Sensors*, vol. 22, no. 9, p. 3331, 2022. doi: [10.3390/s22093331](https://doi.org/10.3390/s22093331). [Online]. Available: <https://www.mdpi.com/1424-8220/22/9/3331>.

- [26] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, “A comprehensive review of EEG-based brain–computer interface paradigms,” en, *Journal of Neural Engineering*, vol. 16, no. 1, p. 011 001, 2019. doi: [10.1088/1741-2552/aaf12e](https://doi.org/10.1088/1741-2552/aaf12e). [Online]. Available: <https://iopscience.iop.org/article/10.1088/1741-2552/aaf12e>.
- [27] K. M. Hossain, M. A. Islam, S. Hossain, A. Nijholt, and M. A. R. Ahad, “Status of deep learning for EEG-based brain–computer interface applications,” en, *Frontiers in Computational Neuroscience*, vol. 16, p. 1 006 763, 2023. doi: [10.3389/fncom.2022.1006763](https://doi.org/10.3389/fncom.2022.1006763). [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fncom.2022.1006763/full>.
- [28] R. Alazrai, H. Alwanni, and M. I. Daoud, “EEG-based BCI system for decoding finger movements within the same hand,” en, *Neuroscience Letters*, vol. 698, pp. 113–120, 2019. doi: [10.1016/j.neulet.2018.12.045](https://doi.org/10.1016/j.neulet.2018.12.045). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0304394018309029>.
- [29] E. P. Torres, E. A. Torres, M. Hernández-Álvarez, and S. G. Yoo, “EEG-Based BCI Emotion Recognition: A Survey,” en, *Sensors*, vol. 20, no. 18, p. 5083, 2020. doi: [10.3390/s20185083](https://doi.org/10.3390/s20185083). [Online]. Available: <https://www.mdpi.com/1424-8220/20/18/5083>.
- [30] L. Xu, M. Xu, T.-P. Jung, and D. Ming, “Review of brain encoding and decoding mechanisms for EEG-based brain–computer interface,” en, *Cognitive Neurodynamics*, vol. 15, no. 4, pp. 569–584, 2021. doi: [10.1007/s11571-021-09676-z](https://doi.org/10.1007/s11571-021-09676-z). [Online]. Available: <https://link.springer.com/10.1007/s11571-021-09676-z>.
- [31] Z. Liu, J. Shore, M. Wang, F. Yuan, A. Buss, and X. Zhao, “A systematic review on hybrid EEG/fNIRS in brain-computer interface,” en, *Biomedical Signal Processing and Control*, vol. 68, p. 102 595, 2021. doi: [10.1016/j.bspc.2021.102595](https://doi.org/10.1016/j.bspc.2021.102595). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1746809421001920>.
- [32] M. A. Bakhshali, M. Khademi, A. Ebrahimi-Moghadam, and S. Moghimi, “EEG signal classification of imagined speech based on Riemannian distance of correntropy spectral density,” en, *Biomedical Signal Processing and Control*, vol. 59, p. 101 899, 2020. doi: [10.1016/j.bspc.2020.101899](https://doi.org/10.1016/j.bspc.2020.101899). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1746809420300550>.
- [33] Y. Jiao, Y. Zhang, Y. Wang, B. Wang, J. Jin, and X. Wang, “A Novel Multilayer Correlation Maximization Model for Improving CCA-Based Frequency Recognition in SSVEP brain-computer interface,” *International journal of neural systems*, vol. 28, no. 04, p. 1 750 039, 2018. doi: [10.1142/S0129065717500393](https://doi.org/10.1142/S0129065717500393). [Online]. Available: <https://www.worldscientific.com/doi/abs/10.1142/S0129065717500393>.

- [34] M. Nakanishi, Y. Wang, X. Chen, Y. Wang, X. Gao, and T. Jung, “Enhancing Detection of SSVEPs for a High-Speed Brain Speller Using Task-Related Component Analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 1, pp. 104–112, 2018. doi: [10.1109/TBME.2017.2694818](https://doi.org/10.1109/TBME.2017.2694818). [Online]. Available: <https://ieeexplore.ieee.org/document/7904641>.
- [35] B. Liu, X. Chen, N. Shi, Y. Wang, S. Gao, and X. Gao, “Improving the Performance of Individually Calibrated SSVEP-BCI by Task-Discriminant Component Analysis,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1998–2007, 2021. doi: [10.1109/TNSRE.2021.3114340](https://doi.org/10.1109/TNSRE.2021.3114340). [Online]. Available: <https://ieeexplore.ieee.org/document/9541393/>.
- [36] N. Guo *et al.*, “SSVEP-Based Brain Computer Interface Controlled Soft Robotic Glove for Post-Stroke Hand Function Rehabilitation,” en, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1737–1744, 2022. doi: [10.1109/TNSRE.2022.3185262](https://doi.org/10.1109/TNSRE.2022.3185262). [Online]. Available: <https://ieeexplore.ieee.org/document/9803244/>.
- [37] Y. Chen, C. Yang, X. Ye, X. Chen, Y. Wang, and X. Gao, “Implementing a calibration-free SSVEP-based BCI system with 160 targets,” *Journal of Neural Engineering*, vol. 18, no. 4, p. 046094, 2021, Publisher: IOP Publishing. doi: [10.1088/1741-2552/ac0bfa](https://doi.org/10.1088/1741-2552/ac0bfa). [Online]. Available: <https://dx.doi.org/10.1088/1741-2552/ac0bfa>.
- [38] C. M. Wong, B. Wang, Z. Wang, K. F. Lao, A. Rosa, and F. Wan, “Spatial Filtering in SSVEP-Based BCIs: Unified Framework and New Improvements,” en, *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 11, pp. 3057–3072, 2020. doi: [10.1109/TBME.2020.2975552](https://doi.org/10.1109/TBME.2020.2975552). [Online]. Available: <https://ieeexplore.ieee.org/document/9006809/>.
- [39] J. Liu, F. Ye, and H. Xiong, “Multi-class Motor Imagery EEG Classification Method with High Accuracy and Low Individual Differences based on Hybrid Neural Network,” *Journal of Neural Engineering*, vol. 18, no. 4, 0460f1, 2021. doi: [10.1088/1741-2552/ac1ed0](https://doi.org/10.1088/1741-2552/ac1ed0). [Online]. Available: <https://doi.org/10.1088/1741-2552/ac1ed0>.
- [40] F. Mattioli, C. Porcaro, and G. Baldassarre, “A 1D CNN for high accuracy classification and transfer learning in motor imagery EEG-based brain-computer interface,” en, *Journal of Neural Engineering*, vol. 18, no. 6, p. 066053, 2021. doi: [10.1088/1741-2552/ac4430](https://doi.org/10.1088/1741-2552/ac4430). [Online]. Available: <https://iopscience.iop.org/article/10.1088/1741-2552/ac4430>.
- [41] D. Zhang, K. Chen, D. Jian, and L. Yao, “Motor Imagery Classification via Temporal Attention Cues of Graph Embedded EEG Signals,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 9, pp. 2570–2579, 2020. doi: [10.1109/JBHI.2020.2967128](https://doi.org/10.1109/JBHI.2020.2967128).

- [42] W. Yi, S. Qiu, H. Qi, L. Zhang, B. Wan, and D. Ming, "EEG feature comparison and classification of simple and compound limb motor imagery," en, *Journal of NeuroEngineering and Rehabilitation*, vol. 10, no. 1, p. 106, 2013. doi: [10.1186/1743-0003-10-106](https://doi.org/10.1186/1743-0003-10-106). [Online]. Available: <https://jneuroengrehab.biomedcentral.com/articles/10.1186/1743-0003-10-106>.
- [43] D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, and Y. Liu, "Making Sense of Spatio-Temporal Preserving Representations for EEG-Based Human Intention Recognition," en, *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 3033–3044, 2020. doi: [10.1109/TCYB.2019.2905157](https://doi.org/10.1109/TCYB.2019.2905157). [Online]. Available: <https://ieeexplore.ieee.org/document/8698218/>.
- [44] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally Constrained Sparse Group Spatial Patterns for Motor Imagery BCI," *IEEE Transactions on Cybernetics*, vol. 49, no. 9, pp. 3322–3332, 2019. doi: [10.1109/TCYB.2018.2841847](https://doi.org/10.1109/TCYB.2018.2841847). [Online]. Available: <https://ieeexplore.ieee.org/document/8386437/>.
- [45] J.-H. Jeong, N.-S. Kwak, C. Guan, and S.-W. Lee, "Decoding Movement-Related Cortical Potentials Based on Subject-Dependent and Section-Wise Spectral Filtering," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 3, pp. 687–698, 2020. doi: [10.1109/TNSRE.2020.2966826](https://doi.org/10.1109/TNSRE.2020.2966826). [Online]. Available: <https://ieeexplore.ieee.org/document/8960436/>.
- [46] F. Karimi, J. Kofman, N. Mrachacz-Kersting, D. Farina, and N. Jiang, "Detection of Movement Related Cortical Potentials from EEG Using Constrained ICA for Brain-Computer Interface Applications," *Frontiers in Neuroscience*, vol. 11, p. 356, 2017. doi: [10.3389/fnins.2017.00356](https://doi.org/10.3389/fnins.2017.00356). [Online]. Available: <http://journal.frontiersin.org/article/10.3389/fnins.2017.00356/full>.
- [47] F. Guo, T. Zhang, N. J. Hanson, and R. Zhang, "Brain source imaging based on movement-related cortical potentials induced by fatigue during self-paced handgrip contractions," *NeuroReport*, vol. 31, no. 4, pp. 300–304, 2020. doi: [10.1097/WNR.0000000000001395](https://doi.org/10.1097/WNR.0000000000001395). [Online]. Available: <https://journals.lww.com/10.1097/WNR.0000000000001395>.
- [48] F. Duan, H. Jia, Z. Sun, K. Zhang, Y. Dai, and Y. Zhang, "Decoding Premovement Patterns with Task-Related Component Analysis," *Cognitive Computation*, 2021. doi: [10.1007/s12559-021-09941-7](https://doi.org/10.1007/s12559-021-09941-7). [Online]. Available: <https://doi.org/10.1007/s12559-021-09941-7>.
- [49] C. Lin, B.-H. Wang, N. Jiang, R. Xu, N. Mrachacz-Kersting, and D. Farina, "Discriminative Manifold Learning Based Detection of Movement-Related Cortical Potentials," en, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 9, pp. 921–927, 2016. doi: [10.1109/TNSRE.2016.2531118](https://doi.org/10.1109/TNSRE.2016.2531118). [Online]. Available: <https://ieeexplore.ieee.org/document/7423780/>.

- [50] P. Ofner, A. Schwarz, J. Pereira, and G. R. Müller-Putz, “Upper Limb Movements can be Decoded from the Time-domain of Low-frequency EEG,” *PLOS ONE*, vol. 12, no. 8, D. Zhang, Ed., e0182578, 2017. doi: [10.1371/journal.pone.0182578](https://doi.org/10.1371/journal.pone.0182578). [Online]. Available: <https://dx.plos.org/10.1371/journal.pone.0182578>.
- [51] P. Ofner, A. Schwarz, J. Pereira, D. Wyss, R. Wildburger, and G. R. Müller-Putz, “Attempted Arm and Hand Movements can be Decoded from Low-Frequency EEG from Persons with Spinal Cord Injury,” *Scientific Reports*, vol. 9, no. 1, p. 7134, 2019. doi: [10.1038/s41598-019-43594-9](https://doi.org/10.1038/s41598-019-43594-9). [Online]. Available: <http://www.nature.com/articles/s41598-019-43594-9>.
- [52] O. Hansson *et al.*, “The Alzheimer’s Association appropriate use recommendations for blood biomarkers in Alzheimer’s disease,” en, *Alzheimer’s & Dementia*, vol. 18, no. 12, pp. 2669–2686, 2022. doi: [10.1002/alz.12756](https://doi.org/10.1002/alz.12756). [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/alz.12756>.
- [53] “2023 Alzheimer’s disease facts and figures,” en, *Alzheimer’s & Dementia*, vol. 19, no. 4, pp. 1598–1695, 2023. doi: [10.1002/alz.13016](https://doi.org/10.1002/alz.13016). [Online]. Available: <https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13016>.
- [54] M. Venkatesh, J. Jaja, and L. Pessoa, “Comparing functional connectivity matrices: A geometry-aware approach applied to participant identification,” *NeuroImage*, vol. 207, p. 116398, 2020. doi: [10.1016/j.neuroimage.2019.116398](https://doi.org/10.1016/j.neuroimage.2019.116398). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1053811919309899>.
- [55] F. Duan *et al.*, “Topological network analysis of early Alzheimer’s disease based on resting-state EEG,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2020.
- [56] L. Alzubaidi *et al.*, “A survey on deep learning tools dealing with data scarcity: Definitions, challenges, solutions, tips, and applications,” en, *Journal of Big Data*, vol. 10, no. 1, p. 46, 2023. doi: [10.1186/s40537-023-00727-2](https://doi.org/10.1186/s40537-023-00727-2). [Online]. Available: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-023-00727-2>.
- [57] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” en, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, 2016, pp. 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90). [Online]. Available: <http://ieeexplore.ieee.org/document/7780459/>.
- [58] J. Kawahara *et al.*, “BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment,” en, *NeuroImage*, vol. 146, pp. 1038–1049, 2017. doi: [10.1016/j.neuroimage.2016.09.046](https://doi.org/10.1016/j.neuroimage.2016.09.046). [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1053811916305237>.

- [59] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018. doi: [10.1088/1741-2552/aace8c](https://doi.org/10.1088/1741-2552/aace8c). [Online]. Available: <https://iopscience.iop.org/article/10.1088/1741-2552/aace8c>.
- [60] H. Peng, F. Long, and C. Ding, "Feature Selection Based on Mutual Information Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005. doi: [10.1109/TPAMI.2005.159](https://doi.org/10.1109/TPAMI.2005.159). [Online]. Available: <https://ieeexplore.ieee.org/document/1453511>.