A Study on Deep Learning Algorithms for Electroencephalography (EEG) and Functional Magnetic Resonance Imaging (fMRI) Signal Processing

by

Zhen ZHANG

Student ID Number: 1248002

A dissertation submitted to the
Engineering Course, Department of Engineering,
Graduate School of Engineering,
Kochi University of Technology,
Kochi, Japan

in partial fulfillment of the requirements for the degree of
Doctor of Engineering

Assessment Committee:
Supervisor: Prof. Makoto IWATA, School of Informatics, Kochi University of Technology

Co-Supervisor: Prof. Kiminori MATSUZAKI, School of Informatics, Kochi University of Technology

Co-Supervisor: Prof. Masaki TAKEDA, Research Center for Brain Communication, Kochi University of Technology

Committee member: Prof. Shinichi YOSHIDA, School of Informatics, Kochi University of Technology

Committee member: Prof. Xiaoyan YU, Department of Physics and Electronic Engineering, Harbin Normal University

September 2023

Abstract

A Study on Deep Learning Algorithms for Electroencephalography (EEG) and Functional Magnetic Resonance Imaging (fMRI) Signal Processing

Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) are neuroimaging techniques used to study brain function. fMRI provides high spatial resolution, allowing for detailed visualization of local brain activities. EEG offers high temporal resolution, allowing for the study of rapid brain processes. These techniques are crucial for understanding various cognitive processes, such as perception, attention, memory, and language. Deep learning (DL) algorithms as data-driven analysis tools for pattern recognition and regression problems have been applied to fMRI and EEG data. DL-based techniques can learn complex and meaningful representations from the raw data, which can extract high-level features from the voxel-level fMRI volumes and be applied to tasks such as brain state classification or decoding cognitive processes. For EEG analysis, DL-based techniques can learn temporal patterns and relationships in the electrical signals, facilitating tasks like emotion recognition or seizure detection. Therefore, this thesis conducted research based on deep learning algorithms from three perspectives, including P300 detection (a type of EEG signal), EEG denoising, and fMRI classification, and proposed a spatial-temporal neural network (STNN), a multi-module neural network (MMNN), and a multi-pooling 3D convolutional neural network (MP3DCNN) to resolve these challenges. Specifically, the first chapter provided a brief overview of EEG, fMRI, and deep learning algorithms, where the definition, characteristics, and application of EEG and fMRI signals are given. Then the principle, key concepts, and real applications of deep learning algorithms are introduced. Finally, the potential and risks of EEG and fMRI signal processing based on deep learning are listed. From the second to the fourth chapter, the background, materials, methods, experiments, and discussion about the proposed algorithms will be described, respectively. The fifth chapter summarizes this thesis. The main works are brief described as follows:

1) Spatial-temporal neural networks (STNN) for P300 detection: The P300 signal, also known as P3 or P300 component, is an electrophysiological response observed in the event-related potential (ERP) waveform recorded from the scalp using EEG tools. It typically occurs around 250-500 milliseconds after the presentation of a rare or unexpected stimulus and is associated with attention, cognitive processes, and decision-making. P300 spellers are common brain-computer interface (BCI) systems designed to transfer information between human brains and computers. In most P300 detections, the P300 signals are collected by averaging multiple electroencephalographic (EEG) changes to the same target stimuli, so the participants are obliged to endure multiple repeated stimuli. Therefore, an STNN is proposed to detect P300 signals, which can detect signals by combining the outputs from a temporal unit and a spatial unit. The temporal unit is a flexible framework consisting of several temporal modules designed for analyzing brain potential changes in the time domain. The spatial unit combines one-dimensional convolutions (Conv1Ds) and linear layers to generalize P300 features from the space domain, and it can decode EEG signals recorded using different numbers of electrodes. We demonstrate the effectiveness of our model using three public databases: P300 speller with ALS patients, covert and overt ERP-based BCI, and BCI Competition III-dataset II. In the dataset of P300 speller with ALS patients, EEG signals from eight ALS patients were recorded, and every participant in the study went through 35 trials, with 10 rounds of repeated stimuli in each trial. Every round of stimuli contained two target stimuli and 10 nontarget stimuli. In the dataset of covert and overt ERP-based BCI, 10 healthy subjects took part in the experiment. The EEG data are recorded on Farwell and Donchin's paradigm and the Geometric Speller. Each experiment included three sessions, with six trials in each session. Each trial contained eight rounds of repeated stimuli with 12 stimuli within each round of stimuli. In the dataset of BCI Competition III-dataset II, both EEG signals from two subjects (A and B) were divided into a training set (85 trials) and a testing set (100 trials). Every trial contained 15 rounds of repeated stimuli. We implemented a within-subject P300 detection and a cross-subject P300 detection, respectively using the dataset of P300 speller with ALS patients as well as covert and overt ERP-based BCI.

The results showed that both amyotrophic lateral sclerosis (ALS) patients and healthy subjects can benefit from this study. In the within-subject P300 detection and the cross-subject P300 detection, the proposed STNN gained higher performance with fewer repeated stimuli than other comparative approaches. Furthermore, we applied the proposed STNN in the P300 detection challenge of BCI Competition III. The accuracy score was 89% in the fifth round of repeated stimuli, outperforming the best result in the literature (accuracy = 80%) to the best of our knowledge. The results demonstrate that the proposed STNN performs well with limited stimuli and is robust enough for various P300 detections. The main reasons are as follows: 1) the temporal unit, as a flexible DL-based network dedicated to time-domain modeling, can capture the temporal dependencies from brain potential changes by constructing an end-to-end multi-level sequential mapping, so it is more sensitive than the previously mentioned approaches when detecting P300 signals; 2) the spatial unit can constantly generalize and compress P300 features in the space domain, which hedges complex noise interference to a certain extent; 3) a joint decision-making mechanism is built into the network by connecting the temporal unit and the spatial unit concurrently, which can utilize the above advantages of the two units, thus achieving both better performance and stronger robustness. In the future, the proposed STNN is predicted to reach a high information transfer rate (ITR) when

implementing online P300 detection. Moreover, we consider that this network has potential for applications in EEG-BCI systems and some other areas of signal processing, such as Electrocardiogram (ECG) classification, seeing that it is designed with a flexible structure and can be fast training and testing with limited data.

2) Multi-module neural networks (MMNN) for EEG denoising: This study proposed an MMNN to remove ocular artifacts (OAs) and myogenic artifacts (MAs) from noisy single-channel electroencephalogram (EEG) signals. This network consists of multiple denoising modules connected in parallel. Each denoising module is built using one-dimensional convolutions (Conv1Ds) and fully connected (FC) layers, and it estimates not only clean EEG signals but also artifacts. The proposed MMNN has two main advantages. First, the multiple denoising modules can purify noisy input EEG signals by continuously removing artifacts in the forward propagation. Second, the parallel architecture allows the parameters of each denoising module to be updated concurrently in the backpropagation, thereby improving the learning capacity of neural networks. We tested the network denoising performance using a recent public database, namely, EEGdenoiseNet. This database provides large-scale clean EEG and artifact epochs, involving 4514 clean EEG epochs, 3400 EOG epochs, and 5598 EMG epochs. These epochs were used to synthesize the training and testing data. Among them, we implemented the model evaluation using 3000 pairs and 400 pairs of training and testing epochs for the OA removal, as well as 5000 pairs and 598 pairs of training and testing epochs for the MA removal. These epochs are synthesized at ten different noise levels, aiming to simulate the real applications. And we performed a 10-fold cross-validation on the training samples for hyperparameter tuning.

The testing results revealed that the proposed network reduced the temporal relative root mean square error (T-RRMSE) and spectral relative root mean square error (S-RRMSE) by at least 6% and enhanced the correlation coefficient (CC) by at least 3% over the state-of-the-art approaches. Observing the deviation distribution between the denoised and clean signals confirmed these significant performance improvements. Furthermore, the proposed network achieved a similar performance efficiency with only 60% of the training data compared to the existing DL models. Finally, the proposed model was compared with the non-deep learning techniques. According to the ANOVA results with Holm-Bonferroni correction, the performance improvement is significant (all p-values < 0.001) in both the OA and MA removals. In the future, there are some challenges worth exploring using the proposed model. For example, OAs and MAs are entangled with motion artifacts in a real EEG epoch, however, there is no available public database to evaluate the model performance for the mixed signals. Given that the proposed model offers significant advantages over the conventional and DL models in this study, the related research is within the scope of further work.

3) Multi-pooling 3D Convolutional Neural Networks (MP3DCNN) for fMRI Classification: Neural decoding of visual object classification via functional magnetic resonance imaging (fMRI) data is challenging and is vital to understand underlying brain mechanisms. The previous study performed categorical (face vs. object), face sub-categorical (male face vs. female face), and object sub-categorical (natural object vs. artificial object) classifications via a classic three-layer 3DCNN, revealing that the human visual system recognizes objects following the principle of going from categories into sub-categories. However, the previous classification model did not significantly present a high accuracy even with 9-fold fMRI data averaging, especially for sub-categorical classification tasks. Therefore, a novel multi-pooling 3D convolutional neural network (MP3DCNN) is proposed, which is expected to reach a higher accuracy than the previous model and play a valuable role in decoding brain mechanisms. The proposed MP3DCNN included a feature extraction, a feature combination, and a classifier, where the feature extraction has a mainchain and two branches. The mainchain is a three-layer 3DCNN, where each 3D convolution is combined with a batch normalization, an average 3D pooling layer, and a rectified linear unit (ReLU). The first and second 3D convolutions each have a branch connection, where an average 3D pooling layer and a linear layer are used to generalize further and connect the extracted features. In the end, through the feature combination and classifier, the model can provide a sophisticated decision using multi-level features in fMRI classification.

In the work, the fMRI dataset is from the previous study, where 53 healthy subjects participated in the visual stimulus task, and each one experienced an average of 9.96 ± 2.88 rounds of visual stimuli, where the subject clicked the button corresponding to a random visual stimulus (an image of a male face, female face, natural object, or artificial object) within 0.5s. During this period, a Siemens 3T MRI scanner recorded the subject's brain states as T1-weighted 3D fMRI volumes. Through SPM12, the fMRI volumes were realigned, co-registered, normalized to the standard Montreal Neurological Institute (MNI) template, and resampled to 2-mm isotropic voxels. After data cleaning, there are 17306 fMRI volumes from 50 subjects available, including 4453, 4399, 4214, and 4240 volumes corresponding to the visual stimuli of male face, female face, natural object, and artificial object, respectively. To suppress the background noise and the irrelevant neural activities, the fMRI dataset of each subject was multi-fold averaged as an option to improve the data quality (for example 9-fold fMRI data averaging). We performed 9-fold cross-validation with 25 training iterations, where the batch size was 64, the learning rate was 0.00001, and the loss function was binary cross entropy (BCE). Within the 25 iterations, the model parameters with the highest accuracy score on each validation dataset were used for model testing. Finally, we used a majority voting scheme to ensemble the nine results in determining the classification accuracy.

The results showed that this model can improve the classification accuracy for categorical (Face vs. Object), face sub-categorical (Male face vs. Female face), and object sub-categorical (Natural object vs. Artificial object) classifications from 1.684% to 14.918% over the previous study in decoding brain mechanisms. The main reasons are speculated as follows: 1) we considered that the multiple 3D average pooling layers can against the local feature redundancy in the feature extraction process and pass the global information to the classifier as much as possible; 2) through the branch connections, the model decision can depend on the merged features of the three 3D convolutions, thereby improving the model's robustness. In future research, we look forward to using grid-search to optimize the model hyperparameters and exploring the visual explanations based on the reached classification results.

Overall, this thesis proposed high-performance deep learning algorithms for P300 detection, EEG denoising, and fMRI classification, respectively. And they are expected to improve the efficiency in various brain decoding tasks in the future.

Contents

List of Figures

List of Tables

List of Appendixes

1. Introduction

1.1 Overview of EEG (Electroencephalography)

Electroencephalography (EEG) is a non-invasive tool for recording the brain's electrical activity. During EEG signal acquisition and processing, EEG electrodes are placed on the scalp to detect the synchronized activity of large groups of neurons in the brain. EEG signals are characterized by the frequency, amplitude, and time-domain [1].

The frequency characteristic indicates the number of oscillations or cycles of electrical activity per second, which is measured in Hertz (Hz). EEG signals are categorized into various frequency bands, each associated with distinct brain states, and the primary frequency bands include: Delta waves (0.5-4 Hz) are prevalent during sleep and in some pathological situations, such as coma or brain injury; Theta waves (4-8 Hz) relate to drowsiness, meditation, and specific cognitive functions, including memory and attention; Alpha waves (8-13 Hz) are dominant during relaxed wakefulness and closed eyes and can be suppressed by opening the eyes or engaging in mental activity; Beta waves (13-30 Hz) are associated with focused attention, mental activity, and motor activity; and Gamma waves (30-100 Hz) are linked to cognitive activities like perception, attention, and memory [1].

The amplitude characteristic of EEG signals indicates the height or strength of the electrical activity measured by EEG electrodes because of the synchronous activity of neuronal populations in the brain. Various factors can impact the amplitude of EEG signals, such as neuronal activity, electrode placement, signal-to-noise ratio (SNR), and individual variability.

The time-domain characteristic of EEG indicates the temporal dynamics of the electrical activity of EEG signals, which can be further investigated in the following perspectives:

1) Latency indicates the time delay between a stimulus or event and the corresponding neural response. For example, latency is a crucial feature in measuring the P300 component and a longer latency to P300 is generally related to slower information processing and may indicate deficits in cognitive function.
2) Event-related potentials (ERPs) [3] denote the specific waveforms in EEG signals that are time-locked to a particular stimulus or event, which is used to investigate the brain mechanisms, such as attention, perception, and memory.
3) Coherence indicates the degree of synchronization or correlation between different brain regions as represented in EEG signals, which can be used to analysis functional connectivity between brain regions.

Based on the above characteristics, EEG signals have a wide range of applications in various fields, and some representative examples of EEG applications are as follows:

1) Clinical diagnosis [4]: For example, clinicians can detect abnormal brain activity and identify the location and extent of seizure activity by analyzing EEG signals,
2) Cognitive neuroscience research [5]: EEG signals can be used to investigate the neural correlates of diverse cognitive processes, such as attention, perception, memory, and language.
3) Brain-computer interfaces (BCIs) [6]: EEG-based BCIs enable individuals to control external devices, such as prosthetic limbs or communication tools. This technology is particularly beneficial for individuals with disabilities.

However, some potential challenges are worth noting in EEG applications, for example, P300 detection studies. Environmental noise, movement artifacts, electrode impedance, and biological noise are some common noise types that can create artifacts or distortions in EEG signals. Therefore, advanced signal processing techniques such as deep learning are expected to play a significant role.

1.2 Overview of fMRI (Functional Magnetic Resonance Imaging)

Functional magnetic resonance imaging (fMRI) is a technique that measures brain activity by detecting changes in blood flow within the brain. This approach relies on the blood-oxygen-level-dependent (BOLD) contrast, which can capture the relationship between neuronal activation and cerebral blood flow [7]. In recent years, fMRI has become a crucial tool in cognitive neuroscience, psychology, and clinical research because of its several unique characteristics as follows:

1) High spatial resolution: Spatial resolution refers to the smallest distinguishable detail in an image or the ability to identify and differentiate between closely spaced structures or brain regions. fMRI is known for its high spatial resolution, allowing researchers to visualize and accurately localize the activation of specific brain regions and structures involved in various cognitive processes and tasks.
2) Whole-brain imaging: Whole-brain coverage means that fMRI measurement can capture activity across the whole brain, which enables researchers to study complex brain networks and interactions among different brain regions. A comprehensive view of brain activity is vital to understand the whole brain mechanisms and grasp the dysfunction in one region on other areas or the brain's overall functioning.
3) Adaptability: fMRI is highly adaptable to various experimental designs. Therefore, researchers can tailor their studies to specific research questions or investigate various cognitive processes.
4) Comparability in cross-subject: fMRI data is proved to have comparability in cross-subject studies. Therefore, researchers can examine and integrate findings from different experiments and populations, which is extremely valuable for understanding brain decoding's generalizability.
5) Data visualization: fMRI data can be visualized in a variety of ways. For example, researchers can analyze fMRI data at the level of large-scale brain networks. The visualization tools can help identify brain mechanisms and neural feedback processing for different stimuli or tasks.

Based on the characteristics of fMRI, several applications have emerged in lots of fields, such as:

1) Cognitive neuroscience research [8]: fMRI is used to study the neural basis of various cognitive processes, such as perception, attention, memory, and emotion.
2) Clinical diagnosis [8]: fMRI is used to understand the neural basis of neurological and psychiatric disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, depression, and autism.
3) Brain mapping: fMRI plays a significant role in advancing brain mapping by offering intricate depictions of the functional architecture within the brain.

4) Pre-surgical schedule: fMRI is employed in preoperative planning to identify crucial functional regions within the brain that should remain intact during surgery, especially when removing brain tumors, which helps minimize harm to vital areas and reduces the risk of postoperative impairments.

While fMRI has presented significant advantages in decoding brain mechanism, there are still some challenges that are worthy of attention:

1) Signal-to-noise ratio (SNR) [9]: fMRI information can be affected by multiple noise sources, including physiological factors (like heart rate and breathing), head movement, and scanner-related artifacts. These noise sources can negatively impact the accuracy and reliability of fMRI measurements.
2) Reproducibility: Variability in fMRI data acquisition, preprocessing, and analysis methods across studies can pose challenges when comparing and replicating results for other researchers.
3) Cost: The high costs of acquiring, maintaining, and operating fMRI scanners can limit the applications of the fMRI-based studies.
4) Contraindications: fMRI scanning requires participants to remain still in a confined, noisy environment, which may not be suitable for specific groups, such as young children or claustrophobic individuals, as well as those with implanted devices or metallic objects in their bodies.

In the future, we expect fMRI techniques to continue evolving, driven by technological advancements and methodological innovations. Advancements in fMRI technology may lead to increased spatial and temporal resolution, thereby enhancing our understanding of brain mechanisms. On the other hand, the multimodal neuroimaging approach, such as fMRI-EEG [11], has the potential to provide a more comprehensive understanding of brain function. Finally, the development and adoption of advanced data analysis methods, such as deep learning algorithms, will continue to expand our ability to extract meaningful information from fMRI data.

1.3 Overview of deep learning algorithms

Deep learning (DL) algorithms are a subfield of machine learning and artificial intelligence (AI), focusing on developing and applying multi-layer neural networks. DL-based techniques can model and learn hierarchical data representations to issue complex pattern recognition and regression problems. The key concepts of deep learning [12] can be summarized as follows:

1) Neural connection: DL models are composed of interconnected layers of artificial neurons or nodes that process and transform input data through a series of non-linear transformations. Deep neural networks (DNN) typically have multiple hidden layers between the input and output layers, allowing the network to capture hierarchical data information.
2) Convolutional layer: A convolutional layer indicates a set of filters to the input data, where each filter is designed to detect a specific feature or pattern. According to different data dimensionalities, there are various convolution operations applied to data, such as 1D, 2D, and 3D convolutions.
3) Weights and biases: Each connection between neural layers has a set of weights, which determines the strength of the connection. Biases are additional parameters that are used to shift the activation function. The weights and biases are adjusted during training to minimize the difference between the predictions and the target values.
4) Activation functions: Activation functions are non-linear functions applied to the output of each neural layer, introducing non-linearity into the network and then enabling it to learn and represent complex, non-linear relationships between inputs and outputs.
5) Loss functions and optimization: DL models are trained using loss functions, aiming to measure the difference between the predictions and the target values. Optimization strategies can adjust the model's weights and biases to minimize the loss function.
6) Backpropagation: Backpropagation is a crucial technique for training deep neural networks, which is used to minimize the difference between network outputs and ground truth labels. It can calculate the gradient of the loss function on the network parameters using the chain rule of differentiation. And it can update the network parameters in the opposite direction of the gradient.

DL techniques have seen tremendous success in many applications, as follows:

1) Image classification: DL algorithms have become the preferred method for image classification because of the data generalization ability. During the network training, the network would be presented with a

series of labeled images to learn the relationships between image features and class labels, which aims to make accurate predictions through the learning process.

2) Language translation: For the language translation task, the network would be presented with parallel texts in two languages and learn to translate between them during the network training.

3) Medical imaging: In the medical imaging tasks, the network would capture the correlations between medical image characteristics and medical conditions. By understanding intricate relationships within medical imaging data, it is expected to enhance precision in crucial tasks such as disease diagnosis and therapy planning.

4) Game AI: DL algorithms have been applied to game AI and have demonstrated the capability to enhance the intelligence and behavior of game characters and agents. In the task, the training data is game data, and the model would learn to make decisions based on the different rules and objectives in the game scenarios, which aims to make game AI more realistic and the game experience more engaging.

Based on the above, deep learning algorithms have the potential to greatly impact the field of fMRI and EEG signal processing, such as signal denoising and classification. For brain decoding tasks, deep learning is expected to improve the automatic capability through the fMRI and EEG data training [13]. Finally, here are servals perspectives to consider in studying fMRI and EEG signal processing based on deep learning.

1) Network performance: The network performance in EEG and fMRI signal processing is greatly influenced by training data quality, training data quantity, and network architecture. Noisy signals will bring risks for data feature learning, and a smaller size of dataset maybe not cover enough data features. Meanwhile, A well-designed network architecture would play a crucial role for improve the learning ability based on training data.

2) Interpretability [14]: Interpretability involves the level to which the network's decisions can be understood or explained by humans. It is important in medical applications because the network's decisions will have consequences for patients.

3) Ethical and Privacy Concerns [15]: The ethical and privacy concerns indicate that DL-based applications must ensure that the technology is used responsibly and ethically for EEG and fMRI signal processing. For example, personal medical information as training data is sensitive and must be protected. Researchers have an obligation to ensure that data is not misused.

1.4 Thesis significance of the study

This thesis introduces a study of deep learning algorithms for EEG and fMRI signal processing. EEG and fMRI signals are the analysis targets in the study because they have the advantages of measuring brain activities in temporal and spatial domains, respectively. Specifically, fMRI adopts the same principles of magnetic fields and radio waves as MRI, while it focuses on detecting and mapping brain activity and functional connectivity during specific tasks or stimuli. And EEG provides exceptional temporal resolution, capturing rapid changes in brain activity with millisecond precision, which makes it well-suited for studying dynamic processes and fast neural events, such as brain oscillations, event-related potentials (ERPs), and transient responses. Nowadays, the other techniques used for recording brain activities include positron emission tomography (PET), near-infrared spectroscopy (NIRS), computed tomography (CT), and magnetoencephalography (MEG). However, there are some limitations to them compared with EEG and fMRI, as follows:

1) PET measures the distribution of radioactive tracers injected into the bloodstream to track brain metabolism and neurochemical activity. However, PET requires the administration of radioactive tracers, which can be costly, time-consuming, and involve radiation exposure. This limits its suitability for repeated measurements over short time intervals.
2) NIRS measures the changes in blood oxygenation levels using near-infrared light. However, it has limited depth penetration into the brain. And its signals can be affected by the absorption and scattering of light in the scalp and skull, potentially leading to measurement artifacts or reduced accuracy in certain regions.
3) CT utilizes X-rays and computer processing to generate cross-sectional images of the brain. However, CT lacks functional data and has lower soft tissue resolution compared to fMRI techniques.
4) MEG measures magnetic fields generated by neuronal activity in the brain. However, compared to EEG, MEG is highly sensitive to external magnetic interference, such as metal objects or electrical equipment, which can lead to artifacts and affect the quality of recorded signals.

EEG and fMRI are chosen because of their complementary strengths in temporal and spatial resolution, brain coverage, non-invasiveness, direct measurement of neural activity, and cost-effectiveness or portability.

Deep learning algorithms can handle complex and high-dimensional data, which are thus used for EEG and fMRI signal processing. Specifically, deep learning models can effectively learn patterns and extract meaningful features from massive datasets, thereby identifying specific patterns or biomarkers and improving the accuracy of prediction and classification tasks. Therefore, deep learning algorithms are expected to resolve some challenges in EEG and fMRI signal processing. The rest of the thesis describes three proposed algorithms:

1) Spatial-temporal neural networks for P300 detection.

2) Multi-module neural networks for EEG denoising.

3) Multi-pooling 3D convolutional neural networks for fMRI classification.

Among them, the P300 components in EEG signals are widely associated with various cognitive processes, including attention, memory, and decision-making. However, P300 detection is a challenging task because of low signal-to-noise ratio, inter-individual variability, task and stimulus variability, limited trial data, temporal overlap, and the presence of artifacts and noise. Deep learning algorithms have emerged as a promising approach for detecting P300 signals because of their advantages in feature generalization and extraction. Furthermore, deep learning algorithms can be used for EEG denoising. First, the deep learning models are trained on large datasets by mapping from noisy EEG signals to clean signals, and then the trained model can be used to generalize and handle different types of noisy EEG signals. The application of deep learning-based denoising methods has shown significant improvements in enhancing the signal quality and preserving important brain activity, so we expected to improve the deep-learning model performance in EEG signal processing in this thesis.

From the perspective of fMRI classification, deep learning algorithms can play a significant role in capturing spatial correlations and maintaining computational efficiency. The main challenges include data availability, high dimensionality, inter-subject variability, spatial and temporal correlations, and interpretability concerns. This study will explore the design of high-performance classification models based on deep learning.

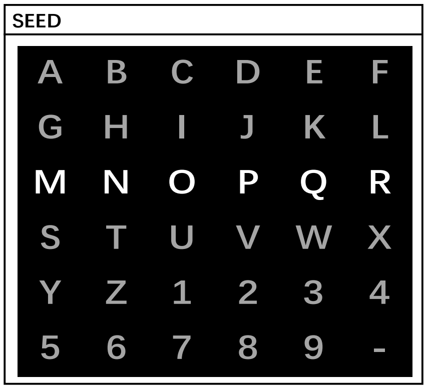
The background, materials, methods, experiments, and discussion about the proposed algorithms will be described in Chapter 2,3 and 4, respectively. The final chapter provides a conclusion to the whole paper.

2. Spatial-Temporal Neural Networks for P300 Detection

2.1 Background

Brain-computer interface (BCI) systems enable neural signals to control external devices directly. In recent years, BCIs have been applied in many fields, such as environmental control [16], communication [17], and neurofeedback rehabilitation [18]. Electroencephalography (EEG) monitoring is one of the most popular measurement tools in BCI applications because of its non-invasiveness, mobility, and relatively low cost [19].

The P300 speller, as an EEG-based BCI paradigm, was first proposed by Farwell and Donchin [20], as shown in Figure 2.1-1. During the spelling, the participants are required to focus their gaze on the lighted characters when the rows or columns of 36 alphanumeric characters are randomly intensified. In this process, the participants' brain activity changes evoked by the target characters, which are called event-related potentials (ERPs). Within the ERPs, the P300 signal is one of the most robust components that corresponds to a positive deflection, occurring 250-500ms after a target presentation [21].



SEED

A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z	1	2	3	4
5	6	7	8	9	-

Figure 2.1-1 Farwell and Donchin's paradigm.

An efficient P300 detection technique is a valuable contribution for the BCI community. Humans, particularly amyotrophic lateral sclerosis (ALS) patients, who suffer from progressive physical disabilities caused by the degeneration of the motor neuron system [22], will benefit from this research. The challenges, however, are that EEG signals inherently have a low signal-to-noise ratio (SNR) and differ significantly between individuals. Even for the same individual, EEG changes can differ in responding to the same target stimuli when affected by internal states and external surroundings. Thus, we usually average multiple EEG responses to a target stimulus to weaken noise and highlight features. However, it adds inconvenience for the participants, who are obliged to spend more time and endure multiple repeated stimuli for the same target. To cope with this challenge, researchers should make a reasonable tradeoff among time, cost, accuracy, and complexity when designing a P300 detection approach.

The current mainstream P3000 detection approaches can be categorized into two types: deep learning (DL) and traditional technologies using statistical features and classifiers. In the traditional ones, the feature extraction mainly includes measures such as independent component analysis (ICA) [23], canonical correlation analysis (CCA) [24], common spatial patterns (CSP) [25], and XDAWN spatial filter [26]. Commonly used classifiers include linear discriminant analysis (LDA) [27], support vector machine (SVM) [28], and Riemannian geometry classifier (RGC) [29] , among others. Of these, the combination of XDAWN and RGC is perhaps the most potent approach for P300 detection [30], which exhibits a strong generalization capability for variable EEG signals. Nevertheless, it is still not as competitive as DL approaches [31].

Convolutional neural network (CNN) as a representative DL framework has attracted widespread attention from the BCI community [32][33][34][35][36][37]. In 2010, Cecotti et al. [34] first proposed a CNN-based P300 detection approach that won the third BCI competition. This method adopts a four-layer CNN architecture to extract channel features and temporal features in sequence, demonstrating that CNN can capture both spatial peculiarities and latent serial dependencies from EEG signals. However, although CNN improved the detection accuracy to an unprecedented level, there are still two major obstacles that lie ahead for such methods. Firstly, the network accuracy depends on the quality and quantity of training data, while the amount of high-quality data commonly remains limited in P300 tasks because of the high cost of time and labor. Secondly, the P300 response is a relatively small potential change presented at a high resolution in the time domain, yet the CNN-based frameworks are not skilled at decoding sequential information with limited EEG data.

To resolve the above problems, some of the recent DL approaches tend to strengthen the learning capability of neural networks when limited data are available, such as [38][39][40], or adopt more advanced architectures to optimize the feature extraction procedure, such as [41][42][43][44], EEGNet as a generic DL network implemented by depth-wise and separable convolutions is proposed, which yields the satisfactory results in various EEG detections. This network extracts temporal features from the EEG signals and then performs spatial filtering on each temporal feature map. With this design, the network can directly perform sequential learning using raw EEG signals and then generalize the captured dependencies in the space domain. It is more competitive for P300 detection than other DL-based pure sequence models, such as recurrent neural networks and long-short-term memory networks. However, this network relies on multiple repeated stimuli to collect EEG signals.

In this paper, we proposed a spatial-temporal neural network (STNN) for P300 detection. It performs better and is more robust in various P300 detections with limited data and repeated stimuli. Our main contributions in the proposed network are as follows: 1) We proposed a parallel network consisting of a temporal unit and a spatial unit to simultaneously learn spatial and temporal features from raw EEG signals; 2) In our design, the spatial unit is mainly constructed using Conv1Ds and linear layers. It can generalize the spatial features of EEG signals recorded using different numbers of electrodes (for example, 8, 16, or 64); 3) We designed a temporal unit inspired by [45]. This unit is used to analyze brain potential changes in the time domain by stacking multiple temporal modules. The number of temporal modules can be adjusted according to the corresponding P300 detection task. We demonstrate the effectiveness of our model using three public databases: P300 speller with ALS patients [46], covert and overt ERP-based BCI [47], and BCI Competition III-dataset II [48].

2.2 Materials

2.2.1 Dataset 1: P300 speller with ALS patients [46]

In Dataset 1, EEG signals from eight ALS patients (five males and three females, mean age = 59.7 ± 12.3 years) were recorded using BCI2000 [49] and Farwell and Donchin's paradigm. The EEG signals were digitized at 256 Hz from eight channels (Fz, Cz, Pz, Oz, P3, P4, PO7, and PO8) according to 10-10 standard [50] and bandpass filtered between 0.1 and 30 Hz.

Every participant in the study went through 35 trials, with 10 rounds of repeated stimuli in each trial. Every round of stimuli contained two target stimuli and 10 nontarget stimuli, where a stimulus was a random intensification of a row or a column. Two target stimuli indicated the intensifications of the row and the column of the target character, respectively. The non-target stimuli were the intensifications of the rows and columns of the non-target characters. The time between the onset of two adjacent stimuli, called stimulus onset asynchrony (SOA), was 250 ms, where the intensification time and the inter-stimulus interval (ISI) were both 125 ms.

2.2.2 Dataset 2: Covert and overt ERP-based BCI [47]

In Dataset 2, 10 healthy subjects (six males and four females, mean age = 26.8 ± 5.6 years) took part in the experiment. The EEG signals were collected with BCI2000, digitized at 256Hz from 16 channels (Fz, FCz, Cz, CPz, Pz, Oz, F3, F4, C3, C4, CP3, CP4, P3, P4, PO7, and PO8) and bandpass filtered between 0.1 and 20 Hz. This study was performed on two speller paradigms: Farwell and Donchin's paradigm and the Geometric Speller (GeoSpell, Figure 2.2-1). The recordings using the two interfaces both included three sessions, with six trials in each session. Each trial contained eight rounds of repeated stimuli with 12 stimuli (two target stimuli and 10 nontarget stimuli) within every round of stimuli. The SOA and the ISI were 250 ms and 125 ms, respectively. For the stimulating patterns, the rows or columns were illuminated on Farwell and Donchin's interface as described in Dataset 1, whereas the GeoSpell interface displayed six characters per time interval until all 36 had appeared twice.

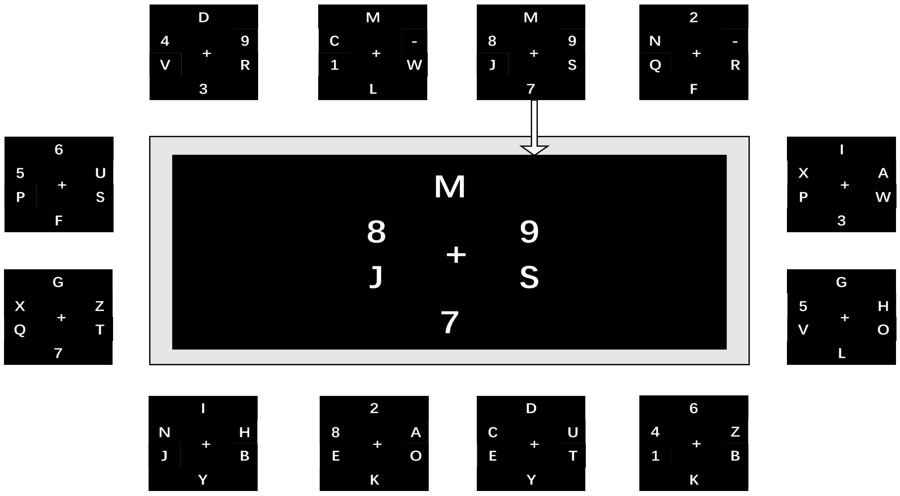


Figure 2.2-1 GeoSpell (Geometric Speller) paradigm.

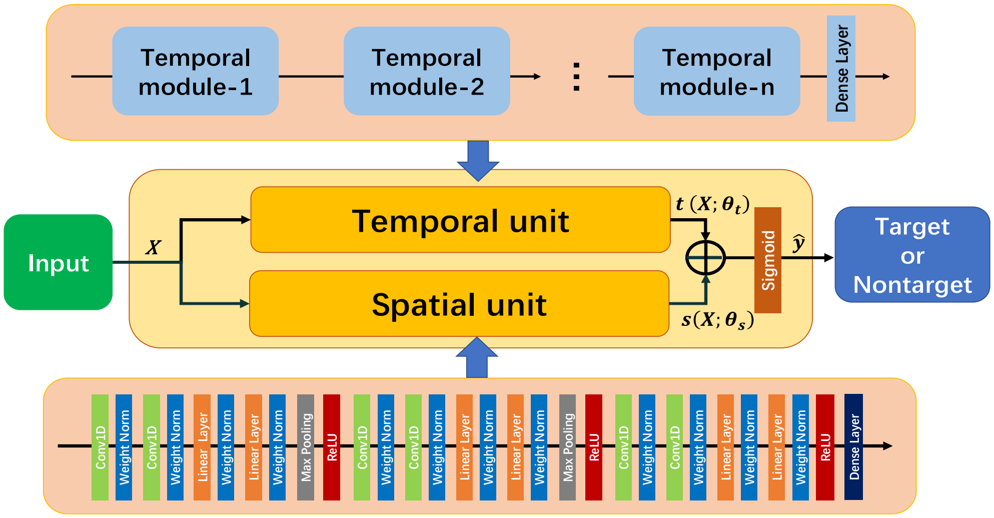


Figure 2.2-2 Overall architecture of the proposed spatial-temporal neural network (STNN).

2.2.3 Dataset 3: BCI Competition III-dataset II [48]

In Dataset 3, the EEG signals recorded using Farwell and Donchin's interface, were bandpass filtered between 0.1 and 60 Hz and digitized at 240 Hz from 64 channels. Both EEG signals from two subjects (A and B) were divided into a training set (85 trials) and a testing set (100 trials). Every trial contained 15 rounds of repeated stimuli, and the intensification time and the ISI were 100 ms and 75 ms in each round.

2.2.4 Data preprocessing

The EEG signals of Dataset 1-3 were down sampled to 128, 128, and 120 Hz, respectively. Then, they were bandpass filtered between 0.1 and 20 Hz with the fifth-order Butterworth filter [51] to remove the short-term fluctuations and leave the longer-term trends [52]. At last, they were extracted from 0 to 0.5 s after each stimulus onset, as shown in Table 2.3-1.

2.3 Methods

This section describes the proposed STNN, where the temporal unit and spatial unit are connected concurrently, as shown in Figure 2.2-2. The details are as follows.

2.3.1 Spatial unit Parallel mechanism

The proposed model adopts a parallel mechanism to perform simultaneous analysis of EEG information in the time and space domains, which is expressed as:

$$\hat{y} = Sigmoid\big(t\,(X;\theta_t) + \, s(X;\theta_s)\big), \qquad (2-1)$$

where X and $\hat{y}$ denote the input of EEG signals and the output of predicted results, the ideal output $\hat{y}$ is either 1 (target) or 0 (nontarget), $t(X;\theta_t)$ and $s(X;\theta_s)$ represent the functions of the temporal unit and the spatial unit, θ_t and θ_s are the network parameters, and $Sigmoid$ is an S-shaped activation function.

2.3.2 Spatial unit

The spatial unit utilizes the global features of the EEG signals in the space domain for P300 detection. It is composed of Conv1Ds, linear layers, weight norms (WNs) [38], max-pooling operations, rectified linear units (ReLUs), and a dense layer, as shown in Table 2.3-2. This unit generalizes spatial features from the horizontal and vertical dimensions by the combination of multiple Conv1Ds and linear layers. To improve the model's robustness in different P300 detections, a multistage feature generalization and compression (Conv1D to Linear layer to Max pooling) is built in the unit. Among them, Conv1Ds as single-dimensional filters, can generalize EEG channel features in the vertical dimension. By setting the kernel size to 1, the correlation between EEG electrodes at each time point is extracted; Linear layers can balance the extratced feature sizes in the horizontal and vertical dimensions, thus minimizing the information loss when compressing feature with max-pooling operations; ReLUs can improve the model's nonlinearity and avoid vanishing gradients, and WNs can accelerate network convergence; The dense layer is connected to the extracted features, producing the predicted results of the spatial unit.

Table 2.3-1 Data description and preprocessing procedure.

	Dataset 1	Dataset 2	Dataset 3
Original sampling rate (Hz)	256	256	240
# of Subjects	8 (ALS)	10 (Healthy)	2 (Healthy)
# of Trials each subject	35	18	185
# of Repeated stimuli	10	8	15
# of EEG channels	8	16	64
Target vs Non-target	1 vs 5	1 vs 5	1 vs 5
Speller paradigm	F	F & G	F
Data preprocessing procedure			
Subsampling rate (Hz)	128	128	120
Butterworth filter (Hz)	0.1-20	0.1-20	0.1-20
Selected duration (s)	0.5	0.5	0.5
EEG format (C × T)	8 × 64	16 × 64	64 × 60

* F/G: Farwell and Donchin's paradigm/GeoSpell paradigm; C/T: The number of channels / time points.

Table 2.3-2 Hyperparameters of the spatial unit.

Layer	# Params	Output
Input	/	C × T
Conv1D +WN	(C, 128, 1)	128 × T
Conv1D + WN	(128, 128, 1)	128 × T
Linear Layer + WN	(T, 128)	128 × 128
Linear Layer + WN	(128, 128)	128 × 128
Max Pooling + ReLU	(2, 2)	64 × 64
Conv1D +WN	(64, 64, 1)	64 × 64
Conv1D + WN	(64, 32, 1)	32 × 64
Linear Layer + WN	(64, 64)	32 × 64
Linear Layer + WN	(64, 32)	32 × 32
Max Pooling + ReLU	(2, 2)	16 × 16
Conv1D +WN	(16, 16, 1)	16 × 16
Conv1D + WN	(16, 4, 1)	4 × 16
Linear Layer + WN	(16, 16)	4 × 16
Linear Layer + WN + ReLU	(16, 4)	4 × 4
Dense Layer	(16, 1)	1 × 1
Output	/	1 × 1

* Conv1D: (Input channel, Output channel, Kernel size); * Linear layers: (Input channel, Output channel); * C/T: Number of channels/time points

Table 2.3-3 Hyperparameters of the temporal generalizer.

Layer	# Params	Output
Input	/	C × T
Conv1D +WN	(C, 128, 1)	128 × T
Linear Layer + WN	(T, 128)	128 × 128
Conv1D +WN	(128, C, 1)	C × 128
Linear Layer + WN	(128, T)	C × T
Output	/	C × T

* Conv1D: (Input channel, Output channel, Kernel size); Linear layers: (Input channel × Output channel); C/T: Number of channels/time points

2.3.3 Temporal unit

Temporal unit detects P300 signals by learning the features of temporal changes in EEG signals. It comprises *n* temporal modules and a dense layer, and the number of the temporal modules can be customized according to the input EEG signals.

Temporal module. As shown in Figure 2.3-1, each temporal module is assembled of a temporal analyzer and a global generalizer with a residual connection. The temporal analyzer performs sequence analysis, which is the core of learning the features of temporal changes. The global generalizer generalizes features from the raw EEG signals or the outputs from the former layer of the temporal module. It provides the global information for the next sequence analysis, which can be expressed as:

$$y_{:,T-1}^{(i)}, \dots, y_{:,0}^{(i)} = f_i\left(x_{:,T-1}^{(i-1)}, \dots, x_{:,0}^{(i-1)}\right), \qquad (2-2)$$

where $\left[x_{:,T-1}^{(i-1)}, \dots, x_{:,0}^{(i-1)}\right]$ and $\left[y_{:,T-1}^{(i)}, \dots, y_{:,0}^{(i)}\right]$ are the inputs and outputs of each temporal module, and both are the same size of $C \times T$; C and T indicate the number of channels and time points, respectively; and f_i represents the i_{th} temporal module.

Temporal analyzer. The temporal analyzer is composed of four components: a dilated Conv1D, a clipping operation, a weight norm, and a ReLU. Within the temporal analyzer of the i_{th} temporal module, the hyperparameters of the dilated Conv1D include input channel, output channel, kernel size, dilation, and zero-padding, where the input channel and output channel are the number of the electrodes of input EEG signals, and the kernel size, dilation, and zero-padding parameter are k, 2^{i-1}, and $(k-1) \times 2^{i-1}$, respectively. By them, the range of learning the temporal changes can be constantly extended. The function of clipping operation is to cut off $\left[y_{:,-1}^{(i)}, \dots, y_{:,-(k-1)\times 2^{i-1}}^{(i)}\right]$ for structural consistency between the inputs and outputs. The functions of the ReLU and weight norm are similar to those in the spatial unit. Figure *2.3-2* shows an example of two stacked temporal modules where the kernal size of the dilated Conv1D was set up to 2. The two temporal analyzers in Temporal module-1 and Temporal module-2 construct a sequential mapping from $\left[x_{:,3}^{(0)}, x_{:,2}^{(0)}, x_{:,1}^{(0)}, x_{:,0}^{(0)}\right]$ to $\left[y_{:,2}^{(1)}, y_{:,0}^{(1)}\right]$ to $\left[x_{:,2}^{(1)}, x_{:,0}^{(1)}\right]$ to $\left[y_{:,0}^{(2)}\right]$. The output $\left[y_{:,0}^{(2)}\right]$ represents the two-level temporal features of $\left[x_{:,3}^{(0)}, x_{:,2}^{(0)}, x_{:,1}^{(0)}, x_{:,0}^{(0)}\right]$.

As for a temporal unit stacked by *n* temporal modules, the output $\left[y_{:,0}^{(n)}\right]$ are the *n*-level mapping of the raw EEG signals $\left[x_{:,l-1}^{(0)}, \dots, x_{:,0}^{(0)}\right]$, which yields the final result of the temporal unit by connecting with a dense layer; l indicates the length of receiving field, as calculated in $(2-3)$.

$$l = k \times 2^{n-1}, \qquad (2-3)$$

where k is the kernel size of the dilated Conv1D within the i_{th} temporal module.

For EEG epochs in different P300 detection tasks, we can customize the temporal detection range by adjusting *n* (the number of stacked temporal modules) and *k* (the kernel size of the dilated Conv1D in the dilated Conv1D). Among them, the structure of the temporal generalizer is similar to the spatial unit, consisting of Conv1Ds, linear layers, and WNs, as shown in Table 2.3-3. The residual connection simplifies the network learning process, especially when multiple temporal modules are stacked.

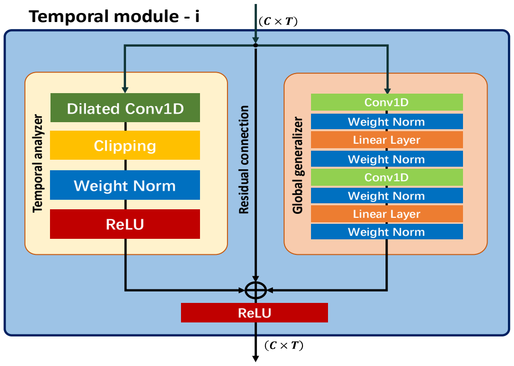


Figure 2.3-1 Internal structure of the temporal module.

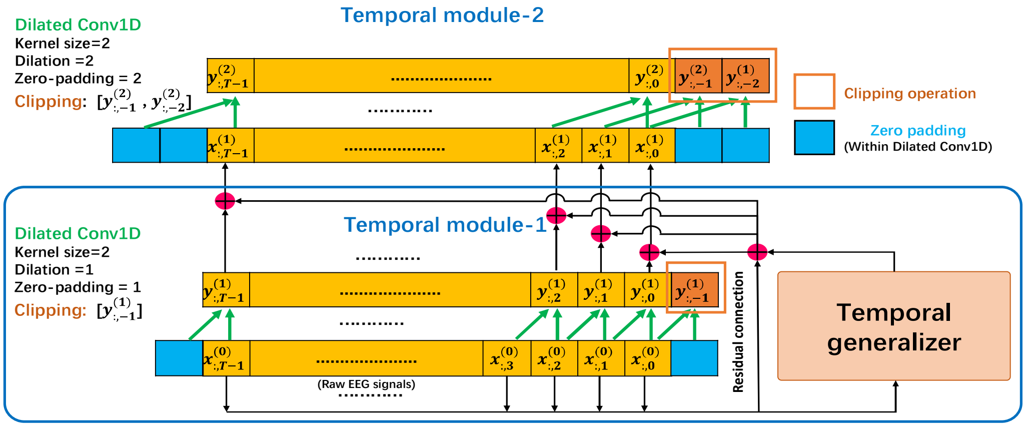


Figure 2.3-2 Mapping relations when stacking two temporal modules.

2.4 Experiments and Results

We performed three experiments to evaluate the proposed STNN: 1) P300 detection under multiple repeated stimuli with applications to ALS patients; 2) P300 detection on two speller paradigms to healthy subjects; 3) an ablation and combination study using BCI Competition III-dataset II.

All the experiments involved 30 iterations of network training within 5 mins, where the Adam optimizer (learning rate = 0.001) was used to minimize the binary cross entropy (BCE) [53] between the outputs and the labels in Pytorch environment. The evaluation metrics included accuracy, area under the receiver operating characteristic curve (AUC) [54], F1-score, and Kappa coefficient [55].

The reference approaches were 3D Input CNN [56], the winner in BCI Competition III [57], and EEGNet-t&p [58], where t and p were the number of temporal filters and pointwise filters, respectively.

2.4.1 Experiment 1

The first experiment explored our model performance under multiple rounds of repeated stimuli to ALS patients using Dataset 1. As described in Section III, Dataset 1 was composed of 8-channel EEG signals from eight ALS subjects, and there were 35 trials of each subject. Each trial was included of EEG signals under 10 rounds of repeated stimuli, there were 12 stimuli (two target and 10 nontarget stimuli) in each round. In a trial, by averaging the EEG epochs under the same stimuli from 1 to i rounds, there were 12 EEG epochs for training or testing the model performance under the i_{th} round of the repeated stimuli.
The first experiment explored our model performance under multiple rounds of repeated stimuli to ALS patients using Dataset 1. As described in Section III, Dataset 1 was composed of 8-channel EEG signals from eight ALS subjects, and there were 35 trials of each subject. Each trial was included of EEG signals under 10 rounds of repeated stimuli, there were 12 stimuli (two target and 10 nontarget stimuli) in each round. In a trial, by averaging the EEG epochs under the same stimuli from 1 to i rounds, there were 12 EEG epochs for training or testing the model performance under the i_{th} round of the repeated stimuli.

The proposed model was represented as STNN-$n\&k$, where n was the number of temporal modules and k was the kernel size of the dilated Conv1D in each module. STNN-3&15, 4&7, 4&8, 5&3, and 5&4 were used in the experiment, and they produced the receptive fields of length 60, 56, 64, 48 and 64, covering the main portion of the EEG signals (The data length is 64 in Databset 1) in the temporal domain. The reference models were EEGNet-4&2, 8&2, 16&2, 4&4, 8&4, 16&4, and the most used in [44] were EEGNet-8&2 and EEGNet-4&2.

We implemented a within-subject P300 detection and a cross-subject P300 detection, respectively. In the within-subject task, we randomly selected 20 trials (240 EEG epochs) for model training from each subject and the remaining 15 trials (180 EEG epochs) for testing the model. The average results of eight subjects with 1-10 rounds of repeated stimuli are shown in Table 2.4-1 and Table 2.4-2. From the performance comparison of these models, we can see that the average AUC and F1 scores of the proposed models are higher than its competitors under 1-10 rounds of repeated stimuli. All our models reach above 0.95 AUC scores using the EEG signals from the first five rounds of repeated stimuli, while EEGNet cannot reach it until at least the ninth round of stimuli. Moreover, the average F1 score of our models under the fifth round of repeated stimuli improved 25.3% than EEGNet in the same condition. This result is close to that of the reference models using 10 rounds of stimuli. It is demonstrated that the proposed models can attain the similar high detection accuracy using fewer repeated stimuli, thereby reducing the number of stimuli for ALS patients in applications.

In the cross-subject P300 detection, we utilized all trials of five random subjects for network training and the trials from the remaining three subjects to evaluate the network performance. The average AUC and F1-score results of five experiments following the above steps are listed in Table 2.4-3 and Table 2.4-4, where we can see that our models obtain improvements of 2% in the average AUC score and 12.4% in the average F1 score under 10 rounds of stimuli. Notably, our models using EEG signals of six rounds of stimuli can reach the similar performance compared to the reference ones of 10 rounds of stimuli.

The average results of the kappa coefficient are given in Figure *2.4-1*. According to [55], a study has substantial reliability when the kappa coefficient is greater than 0.6. STNN-average reached this standard under the 3rd and the 6th rounds of repeated stimuli in the with-subject detection and cross-subject detection, respectively. While EEGNet-average fulfilled the criterion under the 7th and the 10th rounds of stimuli, respectively. Overall, the proposed models achieved advantages both in the within-subject and cross-subject P300 detections, and they can reduce four rounds of repeated stimuli than the reference ones when gaining the similar results.

Table 2.4-1 The AUC results of 1-10 rounds of repeated stimuli in the within-subject P300 detection.

Method	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**	**7th**	**8th**	**9th**	**10th**
STNN-3&15	0.691	0.828	0.879	0.941	0.950	0.959	0.968	0.969	0.974	0.975
STNN-4&7	0.695	0.833	0.883	0.936	0.953	0.962	0.970	0.973	0.979	0.979
STNN-4&8	0.705	0.841	0.891	0.938	0.953	0.961	0.968	0.971	0.977	0.978
STNN-5&3	0.742	0.875	0.932	0.973	0.975	0.980	0.982	0.991	0.993	0.995
STNN-5&4	0.739	0.877	0.938	0.975	0.969	0.985	0.988	0.993	0.993	0.993
STNN-Average	**0.714**	**0.851**	**0.904**	**0.952**	**0.960**	**0.969**	**0.975**	**0.979**	**0.983**	**0.984**
EEGNet-4&2	0.677	0.788	0.836	0.877	0.905	0.918	0.909	0.946	0.963	0.967
EEGNet-8&2	0.709	0.813	0.874	0.940	0.934	0.944	0.951	0.977	0.970	0.970
EEGNet-16&2	0.709	0.810	0.877	0.935	0.935	0.941	0.949	0.979	0.975	0.975
EEGNet-4&4	0.677	0.788	0.841	0.881	0.917	0.921	0.915	0.950	0.965	0.970
EEGNet-8&4	0.679	0.793	0.845	0.891	0.921	0.925	0.936	0.955	0.968	0.969
EEGNet-16&4	0.705	0.811	0.867	0.940	0.938	0.944	0.953	0.976	0.971	0.973
EEGNet-Average	**0.692**	**0.801**	**0.857**	**0.911**	**0.925**	**0.932**	**0.935**	**0.963**	**0.968**	**0.971**

Table 2.4-2 The F1scores of 1-10 rounds of repeated stimuli in the within-subject P300 detection.

Method	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**	**7th**	**8th**	**9th**	**10th**
STNN-3&15	0.253	0.531	0.651	0.709	0.774	0.815	0.832	0.875	0.888	0.885
STNN-4&7	0.286	0.549	0.678	0.731	0.793	0.833	0.850	0.883	0.894	0.886
STNN-4&8	0.295	0.551	0.677	0.723	0.795	0.833	0.851	0.886	0.897	0.899
STNN-5&3	0.455	0.571	0.728	0.787	0.823	0.845	0.863	0.902	0.918	0.921
STNN-5&4	0.440	0.563	0.731	0.793	0.813	0.843	0.841	0.895	0.925	0.923
STNN-Average	**0.346**	**0.553**	**0.693**	**0.749**	**0.800**	**0.834**	**0.847**	**0.888**	**0.904**	**0.903**
EEGNet-4&2	0.053	0.128	0.241	0.410	0.433	0.541	0.641	0.773	0.748	0.772
EEGNet-8&2	0.095	0.233	0.347	0.454	0.587	0.643	0.732	0.790	0.787	0.751
EEGNet-16&2	0.107	0.258	0.372	0.471	0.591	0.648	0.739	0.791	0.795	0.785
EEGNet-4&4	0.062	0.134	0.271	0.399	0.456	0.571	0.643	0.783	0.745	0.781
EEGNet-8&4	0.116	0.245	0.357	0.463	0.597	0.638	0.741	0.783	0.792	0.789
EEGNet-16&4	0.131	0.241	0.371	0.453	0.620	0.645	0.735	0.790	0.789	0.789
EEGNet-Average	**0.094**	**0.207**	**0.326**	**0.442**	**0.547**	**0.614**	**0.705**	**0.785**	**0.776**	**0.778**

Table 2.4-3 The AUC results of 1-10 rounds of repeated stimuli in the cross-subject P300 detection.

Method	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**	**7th**	**8th**	**9th**	**10th**
STNN-3&15	0.734	0.788	0.815	0.885	0.888	0.910	0.913	0.913	0.917	0.918
STNN-4&7	0.779	0.813	0.843	0.889	0.908	0.915	0.921	0.920	0.923	0.925
STNN-4&8	0.778	0.805	0.839	0.893	0.907	0.921	0.923	0.925	0.924	0.925
STNN-5&3	0.792	0.815	0.845	0.901	0.911	0.919	0.925	0.938	0.935	0.937
STNN-5&4	0.789	0.809	0.851	0.899	0.909	0.918	0.923	0.935	0.936	0.936
STNN-Average	**0.774**	**0.806**	**0.836**	**0.893**	**0.904**	**0.916**	**0.921**	**0.926**	**0.927**	**0.928**
EEGNet-4&2	0.675	0.775	0.803	0.833	0.855	0.870	0.895	0.901	0.903	0.905
EEGNet-8&2	0.701	0.799	0.825	0.865	0.891	0.896	0.900	0.903	0.905	0.907
EEGNet-16&2	0.703	0.808	0.827	0.887	0.894	0.897	0.908	0.911	0.905	0.910
EEGNet-4&4	0.685	0.788	0.813	0.845	0.874	0.888	0.901	0.905	0.904	0.909
EEGNet-8&4	0.705	0.803	0.825	0.883	0.901	0.909	0.910	0.905	0.917	0.910
EEGNet-16&4	0.705	0.811	0.831	0.899	0.893	0.905	0.907	0.903	0.907	0.907
EEGNet-Average	**0.695**	**0.797**	**0.821**	**0.869**	**0.884**	**0.894**	**0.903**	**0.904**	**0.906**	**0.908**

Table 2.4-4 The F1scores of 1-10 rounds of repeated stimuli in the cross-subject P300 detection.

Method	**1st**	**2nd**	**3rd**	**4th**	**5th**	**6th**	**7th**	**8th**	**9th**	**10th**
STNN-3&15	0.328	0.401	0.486	0.543	0.592	0.632	0.639	0.688	0.721	0.755
STNN-4&7	0.356	0.414	0.505	0.560	0.611	0.640	0.645	0.700	0.735	0.774
STNN-4&8	0.356	0.407	0.499	0.574	0.605	0.639	0.641	0.705	0.740	0.776
STNN-5&3	0.381	0.445	0.503	0.599	0.635	0.669	0.671	0.711	0.753	0.785
STNN-5&4	0.385	0.443	0.510	0.607	0.643	0.671	0.681	0.709	0.761	0.779
STNN-Average	**0.361**	**0.422**	**0.501**	**0.577**	**0.617**	**0.650**	**0.656**	**0.703**	**0.742**	**0.774**
EEGNet-4&2	0.157	0.287	0.352	0.411	0.555	0.578	0.600	0.628	0.639	0.638
EEGNet-8&2	0.187	0.305	0.405	0.421	0.576	0.590	0.615	0.631	0.645	0.671
EEGNet-16&2	0.195	0.311	0.370	0.450	0.563	0.584	0.607	0.630	0.639	0.658
EEGNet-4&4	0.143	0.277	0.348	0.399	0.541	0.573	0.599	0.611	0.642	0.645
EEGNet-8&4	0.185	0.315	0.401	0.433	0.588	0.601	0.606	0.605	0.635	0.648
EEGNet-16&4	0.199	0.322	0.399	0.441	0.571	0.591	0.617	0.622	0.639	0.641
EEGNet-Average	**0.178**	**0.303**	**0.379**	**0.426**	**0.566**	**0.586**	**0.607**	**0.621**	**0.640**	**0.650**

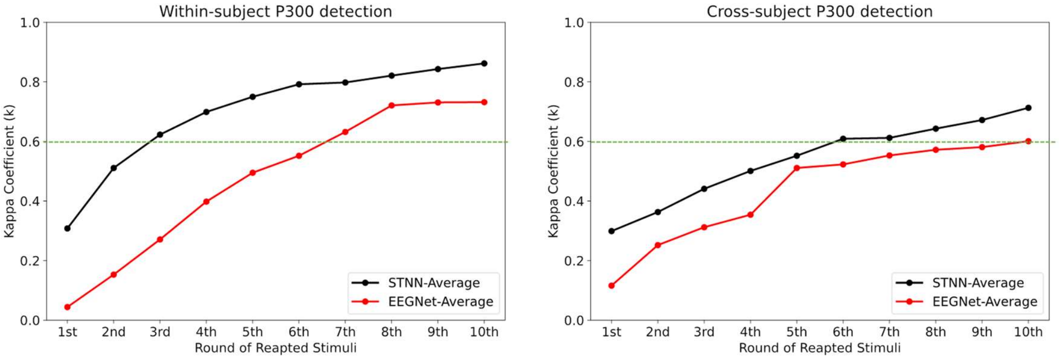


Figure 2.4-1 Kappa coefficient results in the within-subject P300 detection and cross-subject P300 detection.

2.4.2 Experiment 2

The second experiment studied our model performance on the two P300 speller paradigms to healthy subjects using Databset 2. Databset 2 were 16-channel EEG signals from 10 healthy subjects. There were three sessions (six trials in each session) in each subject's recordings. In each trial, the subject experienced eight rounds of repeated stimuli. Due to the fact that the data from Databets 1 and 2 have the same length in the time domain, we still used STNN-3&15, 4&7, 4&8, 5&3, and 5&4 to implement the P300 detection tasks.

In the within-subject experiment, two sessions were randomly chosen as the training set from each healthy subject, and the remaining one was used for testing the models. Figure 2.4-2 and Figure 2.4-3 give the AUC and F1 scores and Kappa coefficient results of the proposed models and reference models using Farwell and Donchin's paradigm and the GeoSpell paradigm. We can see that the proposed STNN-3&15, 4&7, 4&8, 5&3, and 5&4 all achieved perfect detection (AUC, F1 scores and Kappa coefficient results were equal to 1) under the second rounds of stimuli on Farwell and Donchin's paradigm and under the fourth rounds of stimuli on the GeoSpell paradigm, while the reference ones need at least four and six rounds of repeated stimuli to reach this goal on Farwell and Donchin's and the GeoSpell paradigm, respectively. It shows that our models reached better performance using fewer stimuli to healthy subjects. In the cross-subject experiment, we utilized all the trials from five random subjects to train network parameters. The trials from the remaining five subjects were used for the network testing. Figure 2.4-4 and Figure 2.4-5 show that our models always score higher than or equal to the reference ones under 1-8 rounds of repeated stimuli and reach perfect detection under the second round of stimuli over the two paradigms, while the reference ones fulfill this condition in the third or the fourth stimuli. Therefore, it can be seen that the proposed models can reduce repeated stimuli to healthy subjects and is robust to the two different P300 speller paradigms.

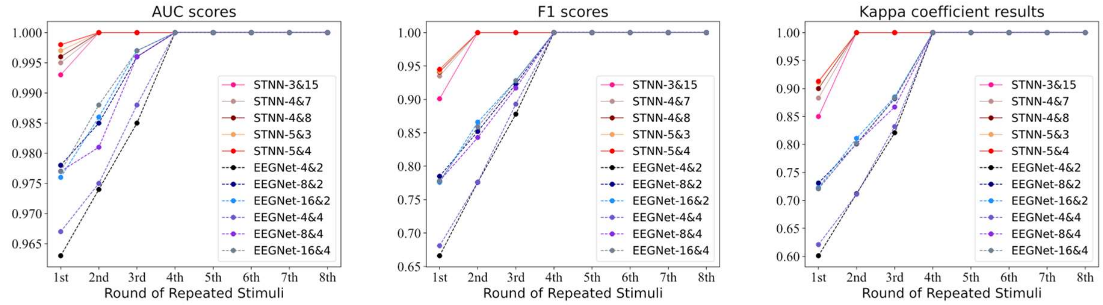


Figure 2.4-2 The AUC, F1 scores and Kappa coefficient results using Farwell and Donchin's paradigm in the within-subject P300 detection.

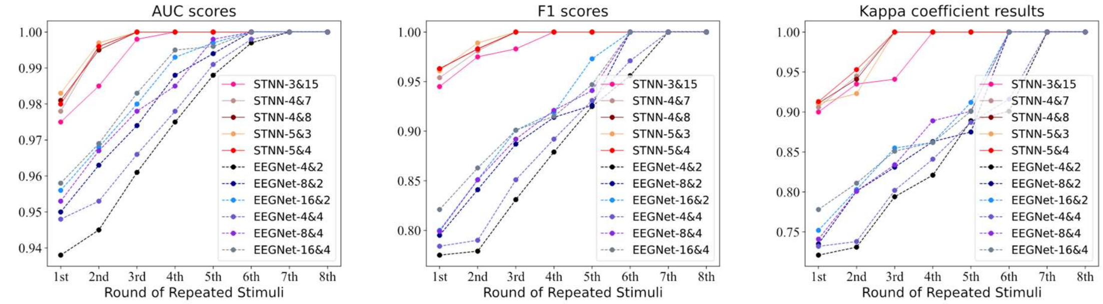


Figure 2.4-3 The AUC, F1 scores and Kappa coefficient results using the GeoSpell paradigm in the within-subject P300 detection.

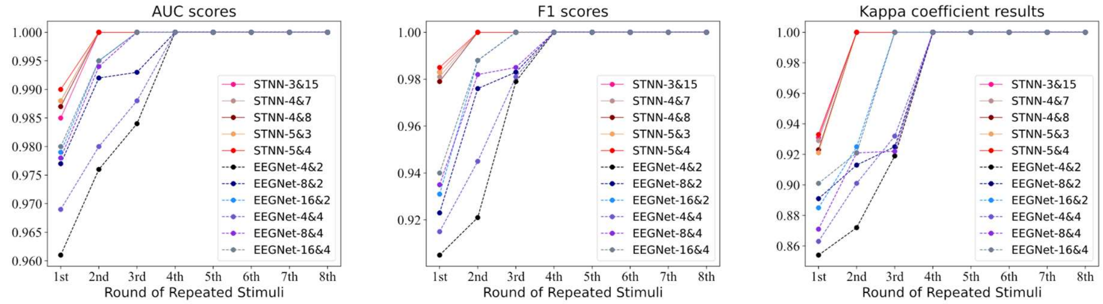


Figure 2.4-4 The AUC, F1 scores and Kappa coefficient results using Farwell and Donchin's paradigm in the cross-subject P300 detection.

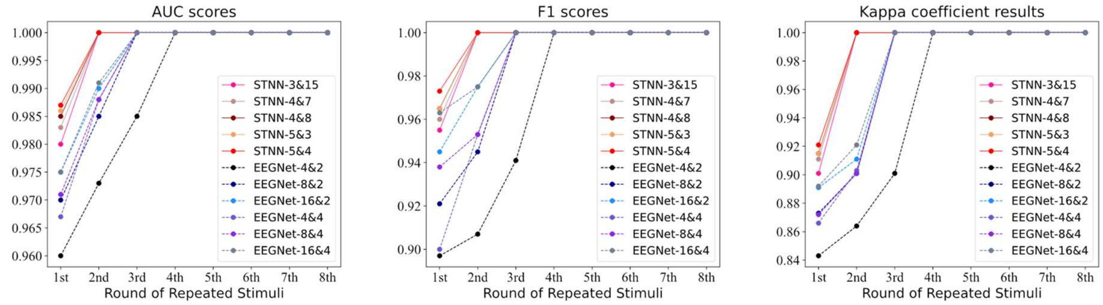


Figure 2.4-5 The AUC and F1 scores and Kappa coefficient results using the GeoSpell paradigm in the cross-subject P300 detection.

2.4.3 Experiment 3

To measure the contribution of individual components and component combinations on the model performance, the third experiment was an ablation and combination study on Dataset 3 (BCI Competition III-dataset II, 64 channels), where the training and testing sets of two subjects (A and B) were described in Chapter 2.2. In order to compare to other models, we implemented the P300 detection using the same evaluation metrics and rounds of stimuli as in the literatures [22][44].

In the combination study, we tested the accuracy of multiple combinations of STNN, including STNN-1&60, 2&30, 3&15, 4&7, 4&8, 5&3, 5&4, and 6&2. In the ablation study, the above combinations were compared with their temporal units and STNN assembled with only the spatial unit. The results are shown in Table 2.4-5, where we can see that 1) the network performance is continuously improved by stacking one to four temporal modules in the temporal unit, while the performance does not continue to be enhanced when stacking more than four modules; 2) the parallel mechanism of the temporal unit and the spatial unit improves the accuracy by 1-3% over the temporal unit working alone; 3) the temporal unit is superior to the spatial unit in terms of the average accuracy; 4) to the best of our knowledge, the proposed STNN stacked with four or more temporal units outperforms the best state-of-the-art model in the literatures by at least 9% in accuracy.

Table 2.4-5 Ablation and combination study: Accuracy under 5 rounds of repeated stimuli on Database 3.

Method	Subject-A	Subject-B	Average
STNN-1&60 (l = 60)	82.5%	89.1%	85.8%
STNN-1&60 - Only with the temporal unit	81.0%	86.2%	83.6%
STNN-2&30 (l = 60)	83.5%	88.5%	86.0%
STNN-2&30 - Only with the temporal unit	82.2%	87.4%	84.8%
STNN-3&15 (l = 60)	88.3%	89.5%	88.9%
STNN-3&15 - Only with the temporal unit	85.8%	89.2%	87.5%
STNN-4&7 (l = 56)	88.0%	90.0%	89.0%
STNN-4&7 - Only with the temporal unit	87.6%	87.7%	87.7%
STNN-4&8 (l = 64)	87.5%	90.5%	89.0%
STNN-4&8 - Only with the temporal unit	86.9%	87.5%	87.2%
STNN-5&3 (l = 48)	87.7%	90.5%	89.1%
STNN-5&3 - Only with the temporal unit	85.4%	88.8%	87.1%
STNN-5&4 (l = 64)	88.3%	89.7%	89.0%
STNN-5&4 - Only with the temporal unit	84.7%	87.9%	86.3%
STNN-6&2 (l = 64)	88.4%	89.8%	89.2%
STNN-6&2 - Only with the temporal unit	86.1%	85.9%	86.0%
STNN - Only with the spatial unit	82.5%	83.5%	83.0%
Competitors			
3D Input CNN (Best result in literatures)	74%	86%	80%
Winner in BCI Competition III	60%	87%	73.5%

*l represents the receiving field size.

2.5 Discussion and Conclusion

The results prove that STNN performs better than other DL model and reduces the number of repeated stimuli in different P300 detections. Both healthy subjects and ALS patients can benefit from this research, even with limited data. The main reasons are as follows: 1) the temporal unit, as a flexible DL-based network dedicated to time-domain analysis, can capture the temporal dependencies from brain potential changes by constructing an end-to-end multi-level sequential mapping, so it is more sensitive than the previously mentioned approaches when detecting P300 signals; 2) the spatial unit can constantly generalize and compress P300 features in the space domain, which hedges complex noise interference to a certain extent; 3) a joint decision-making mechanism is built into the network by connecting the temporal unit and the spatial unit concurrently, which can utilize the above advantages of the two units, thus achieving both better performance and stronger robustness, as shown in Experiments 1 and 2.

Furthermore, it should be emphasized that stacking multiple temporal modules within the temporal unit is critical for sequential modeling, as shown in Experiment 3. The network accuracy is constantly improved when one to four temporal modules are stacked, which demonstrates a more complicated multi-level sequence model is more suitable for characterizing temporal changes in human brain regions. Nevertheless, the over-stacking of temporal modules cannnot endlessly improve its performance but rather increase the model complexity because of the larger number of training parameters, as we can see that the accuracy scores of STNN-4&7, 4&8, 5&3, 5&4, and 6&2 are almost equivalent. Even so, our results still significantly outperform the best methods in the literature, to the best of our knowledge, in BCI Competition III. This is possible because, driven by the great success of 2D or 3D CNNs in image processing and video analysis, some current state-of-the-art DL frameworks are commonly obsessed with high-dimensional feature extraction from EEG data. However, the P300 signals present significant 1D features (the deflections in the time domain) rather than high-dimensional ones. While the 2D or 3D CNN frameworks are not skilled at decoding features from EEG signals recorded using a small number of channels, because the EEG data inherently lack the spatial resolution [57]. In contrast, our network focuses more on the temporal activities within the P300 signals and EEG channel generalization in the space domain, which thus can capture more hidden information from EEG signals at a low SNR.

In the future, the proposed STNN is predicted to reach a high information transfer rate (ITR) when implementing online P300 detection. Moreover, we consider that this network has potential for applications in EEG-BCI systems and some other areas of signal processing, such as Electrocardiogram (ECG) classification [58], seeing that it is designed with a flexible structure and can be fast training and testing with limitied data.

3. Multi-Module Neural Networks for EEG Denoising

3.1 Background

Electroencephalography (EEG) is a safe, reliable, and relatively non-invasive measurement tool to study human brain activity. However, noise and artifacts are always contained in EEG signals, and they are entangled with brain activity.

Eye movements [59] and facial muscle [60] activity are two common causes of noise and artifacts in EEG epochs. Eye movements distort the electric field around the eyes and over the scalp, thus causing ocular artifacts (OAs) [61]. Facial muscle activity responds to pressure changes in the upper airway, generating electrical amplitude signals, called myogenic artifacts (MAs) [62]. Recently, many approaches, such as regression [63][64][65], adaptive filtering [66][67], blind source separation (BSS) [68][69][70][71], and empirical mode decomposition (EMD) [72] have been proposed to remove OAs and MAs from EEG epochs. A high-performance denoising approach should be able to accurately remove artifacts in real-time without distorting the signal of interest and be sufficiently robust to reconstruct EEG data in various formats, especially the signals recorded using only a few electrodes. However, these methods are not fully competent in fulfilling these criteria. The regression and adaptive filtering techniques are not efficient for real-time applications, because they need to estimate the transfer and filtering coefficients before removing the artifacts. BSS and EMD are not functional for single-channel signals, because they remove artifacts by decomposing and reconstructing the EEG signals in the time and frequency domains. However, the signal decomposition relies on the independence between channels that the single-channel signals do not have.

Researchers have applied deep learning (DL) technologies to address these issues after witnessing their breakthroughs in the fields of computer vision [73][74][75][76][77][78][79] and natural language processing [80][81][82]. Recently proposed DL-based EEG denoising approaches [83][84][85][86][87] use fully connected (FC) layers, one-dimensional convolutions (Conv1Ds), and long short-term memory (LSTM) to build end-to-end learning models. Such models can automatically output real-time results and perform well even when both the multi-channel EEG information and reference signals are unavailable.

One of the key aspects of the DL models is that artifact removal strategies are not designed by human engineers but are learned from data, so the model performance is greatly influenced by training data [88]. However, in many cases, particularly in real applications, it is highly expensive to collect high-quality training data [89]. Therefore, there is value in studying the DL model performance on limited data. On the other aspect, DL models usually run as black boxes [90]. The lack of transparency may hinder DL applications in the medical field because it is difficult for humans to verify whether a complex DL model has expert medical or signal-processing knowledge. Thus, DL models that provide the explanations for their mechanism deserve to be further explored.

In this study, we propose a novel multi-module neural network (MMNN) for EEG denoising. This network can be implemented in real-time and applied to single-channel EEG data. Our contributions are as follows: 1) Artifact removal is defined as the detachment of pure EEG signals from signals containing additive noise, therefore we create a network flow that can constantly decompose and assemble EEG information using the proposed denoising modules. 2) We designed the denoising modules using Conv1Ds and FC layers, aiming to customize a solution specialized at separating OAs or MAs from noisy EEG signals by network learning. Conv1Ds were used to extract and generalize the informative features of brain activity, and FC layers were used to reconstruct the clean signals and artifacts. Their combination acted as an end-to-end trainable filter. 3) Referring to the work of EEGdenoiseNet that provided a publicly available structured database for EEG denoising studies, we compared the proposed network with the existing DL and conventional techniques under the same condition. 4) The model denoising performance using different amounts of learning data was explored, and the visualization of the model's component was discussed.

3.2 Materials

The database used in this study is summarized in Table 3.2-1. This database provided a large-scale dataset of clean EEG and artifact epochs, involving 4514 clean EEG epochs, 3400 EOG epochs, and 5598 EMG epochs. In the previous DL denoising studies [86][87], these epochs were used to synthesize the training and testing data. Their extraction process is briefly described as follows:

Table 3.2-1 The benchmark data used in our EEG denoising study.

	EEG	EOG	EMG
# of Epochs	4514	3400	5598
Sampling rate (Hz)	256 and 512	256	512
Bandpass filter (Hz)	1 to 80	0.3 to 10	1 to 120
# of EEG channels	1	1	1
Selected duration (s)	2	2	2
Notch filter (Hz)	50	50	50
Detrend	Yes	Yes	Yes
Manual check	Yes	Yes	Yes

3.2.1 Clean EEG data

The EEG data are composed of 4514 clean EEG epochs, and each epoch is a single-channel EEG segment of 2s. As described in [86], 64-channel EEG epochs were collected from a public database of motor-imagery BCI [91]. These epochs were then band-pass filtered between 1 and 80 Hz, notch-filtered (50 Hz), detrended, and processed using independent component analysis on ICLabel [92], Finally, the processed epochs were sampled at 256 and 512 Hz, respectively, cut into single-channel epochs, and manually checked to ensure that each one was clean.

3.2.2 Electrooculogram (EOG) data

The dataset of EOG data contains 3400 single-channel OA epochs with a sample rate of 256 Hz. These epochs were extracted from previous studies [93][94][95][96][97]. As described in[86], the data were bandpass filtered between 0.3 and 10 Hz, notch-filtered (50 Hz), and detrended. The extracted OAs were subsequently segmented into 2s per epoch and visually checked by experts.

3.2.3 Electromyography (EMG) data

The dataset of EMG data consists of 5598 MA epochs, and each epoch is a single-channel EEG epoch with a duration of 2s and a sampling rate of 512 Hz. These MA epochs were collected from [98] and band-pass filtered between 1 and 120 Hz. Afterwards, these epochs were notch-filtered (50 Hz), detrended, and visually checked by experts.

3.3 Methods

This section describes the proposed MMNN in detail. We first define the EEG denoising problem and then describe the denoising module, which serves as the basic component in our model, followed by the model structure. Finally, we introduce the synthesis process of noisy EEG signals, as well as training and testing data for OA and MA removals.

3.3.1 Problem definition

OAs and MAs belong to ambient noises, also called background noises [99][100]. They are generated independent of the clean signals, therefore the relationship among clean signals, OAs or MAs, and noisy signals in EEG recordings can be expressed as [101][102]:

$$Y = X + Z \qquad (3-1)$$

where X, Z and Y denote clean signals, ocular or myogenic artifacts, and noisy signals, respectively.

The essence of EEG denoising is to estimate the clean signals using the noisy signals Y. For DL denoising models, it is challenging to use the prior knowledge [103] learned from Z distribution to filter Y.

3.3.2 Denoising module

In our design, the denoising module is constructed by four Conv1Ds with rectified linear units (ReLUs), a residual connection, and two FC layers, as shown in Figure 3.3-1. Table 3.3-1 provides the details of the hyperparameters, and the parameter tuning process of c and k is given in Section III. Notably, the proposed denoising module outputs both clean signals and artifacts.

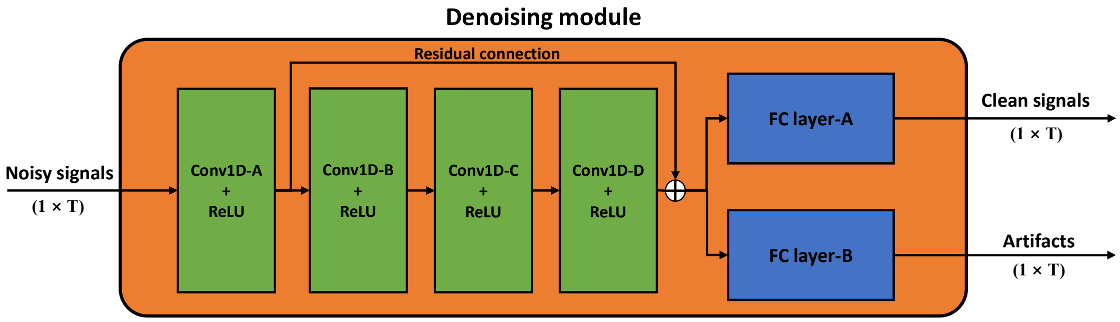


Figure 3.3-1. The internal structure of the denoising module.

Table 3.3-1 Hyperparameters of the denoising module.

Input	Noisy signals (1 × T)
Conv1d-A + ReLU	Input channel = 1; Output channel = c; Kernel size = k; Zero-padding = (k-1)/2
Conv1d-B + ReLU	Input channel = c; Output channel = c; Kernel size = k; Zero-padding = (k-1)/2
Conv1d-C + ReLU	Input channel =c; Output channel = c; Kernel size = k; Zero-padding = (k-1)/2
Conv1d-D +ReLU	Input channel = c; Output channel = c; Kernel size = k; Zero-padding = (k-1)/2
FC-layer-A	Input features = c × T; Output features = T
FC-layer-B	Input features = c × T; Output features = T
Outputs	Clean signals (1 × T) Artifacts (1 × T)

The objective of Conv1Ds is to decompose noisy EEG signals, and their parameters need to be learned from training data. Within each Conv1D, the output channel and kernel size determine the computational complexity and filter length of feature extraction, respectively. The zero-padding operation can maintain the structural consistency between the inputs and outputs. ReLUs can improve the model's nonlinearity and avoid the vanishing gradient problem in the learning stage. Conv1D-A first dismantles noisy single-channel EEG signals into features in multiple dimensions. Conv1D-B, C, and D, are subsequently used to continuously generalize the extracted features. The combination of multiple Conv1Ds with ReLUs can build complex mappings, thus dismantling EEG signals more finely. Residual connections can accelerate network convergence and improve the model's learning ability. The function of the FC layers is to reconstruct clean signals and artifacts by connecting the generalized features. The output clean signals and artifacts have the same data size as the input (data size: 1 × T).

3.3.3 Network structure

The proposed MMNN is built using multiple denoising modules, their number is flexible and can be adjusted according to different denoising tasks. An MMNN assembled using n denoising modules (MMNN-n) is shown in Figure 3.3-2, where the inputs and outputs of n denoising modules are Y, $Y - \hat{Z}_1$ to $Y - \hat{Z}_{n-1}$ and $\hat{X}_1, \hat{X}_2$ to $\hat{X}_n$. The final estimation of clean signals is the sum of $\hat{X}_1$, $\hat{X}_2$ to $\hat{X}_n$. In our designed structure, the proposed MMNN constantly purifies the inputs for each denoising module by removing the artifact estimation. Specifically, $Y - \hat{Z}_{i-1}$ replaces Y itself as the input for the i_{th} denoising module. According to (3–1), the former is a purer EEG signal than the latter. Therefore, there is a high probability that the outputs of the i_{th} denoising module, $\hat{X}_i$ and $\hat{Z}_i$, are closer to the ground truth of EEG signals and artifacts in theory.

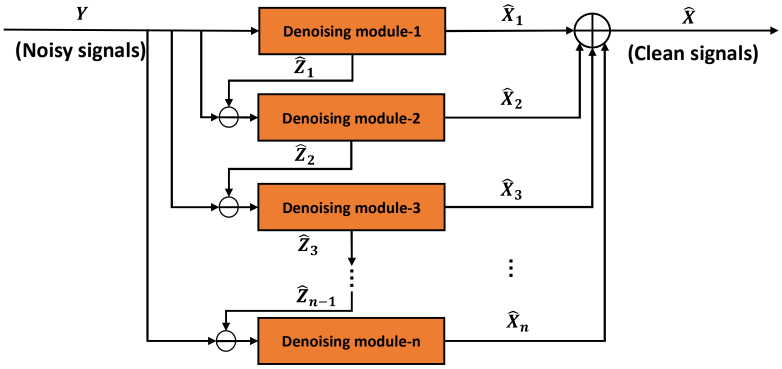


Figure 3.3-2 The structure of MMNN-n (Multi-Module Neural Network-n).

Based on the above, a workflow of multiple denoising modules can constantly improve the network's performance in theory. However, multi-stacking structures may lead to vanishing gradient problem during backpropagation [104]. To hedge this risk, the proposed model is designed as a parallel architecture, thus allowing the parameters from each denoising module to be updated synchronously. The network architecture is expressed as:

$$\hat{X}_i,\ \hat{Z}_i = \mathcal{F}_i(Y - \hat{Z}_{i-1}) \qquad (3-2)$$

$$\hat{X} = \mathcal{G}(Y) = \sum_{i=1}^{n} \hat{X}_i \qquad (3-3)$$

where $\hat{X}$ is the final estimation of the clean signal; $\mathcal{G}$ is the proposed MMNN, Y is the input noisy signal, and n is the number of the denoising modules; $\mathcal{F}_i$ indicates the i_{th} denoising module, $\hat{X}_i$ and $\hat{Z}_i$ are the reconstructed clean signals and artifacts, respectively, and $\hat{Z}_0 = 0$.

3.3.4 Noisy signal synthesis

Using the clean EEG, EOG, and EMG epochs from the mentioned database, we synthesized noisy EEG epochs for model training and testing. The synthesized noisy EEG epochs and clean EEG epochs were the data and labels, respectively. In the training stage, the Adam optimizer was adopted to minimize the mean squared error (MSE) [105] between the model outputs and labels. The details of the noisy signal synthesis are as follows.

To synthesize noisy signals with different noise levels, the signal-to-noise ratio (SNR) as a reference is first given, as shown in $(3-4)$. It describes the ratio of the true signal to the background noise and is widely used to evaluate noise levels.

$$SNR = 10\log\frac{RMS(X)}{RMS(Z)} \qquad (3-4)$$

where X and Z are the discrete-time clean EEG signal and artifact, respectively; and RMS is the root mean squared value, as defined:

$$RMS(P) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} p_i^2} \qquad (3-5)$$

where p_i indicates the i_{th} discrete time point in an epoch of P, and n is the number of time points in the epoch.

For signal synthesis with the given dataset, X denotes any clean single-channel EEG epoch, and Z is expressed as $\lambda \times N$, where N is any single-channel EOG or EMG epoch and λ is the parameter used to control the SNR of the noisy signal. By $(3-4)$, λ can be derived as:

$$\lambda = \frac{RMS(X)}{10^{0.1\times SNR} \times RMS(N)} \qquad (3-6)$$

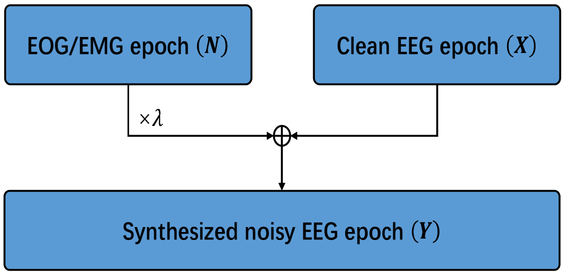
According to $(3-1)$, a clean EEG epoch X and an EOG or EMG epoch N can be simulated into a noisy EEG epoch Y with any SNR level, as shown in $(3-7)$:

$$Y = X + \lambda \times N \qquad (3-7)$$

3.3.5 Training data and testing data

Previous studies [106][108][109] have demonstrated that the SNR values of OAs and MAs are commonly between -7 and 2dB, thus the noisy signals were synthesized within this SNR range.

The OA removal task was implemented using 30000 pairs and 4000 pairs of training and testing samples, respectively. They were synthesized using 3400 clean EEG epochs (randomly selected from 4514 clean EEG epochs) and all 3400 EOG epochs, according to $(3-6)$ and $(3-7)$. For the noisy EEG synthesis of training dataset, the SNR values were followed a uniform distribution from -7 to 2dB, X and N were 3000 of 3400 clean epochs and EOG epochs, respectively. The testing set were synthesized using the remaining 400 pairs of EEG epochs and EOG epochs, and the SNR values ranged from -7dB to 2dB at an interval of one. (-7dB, -6dB, -5dB, -4dB, -3dB, -2dB, -1dB, 0dB, 1dB, 2dB). In the MA removal task, all 4514 clean EEG epochs and 5598 EMG epochs were utilized, where we randomly copied 1084 clean EEG epochs into original EEG epochs, thus producing 5598 clean EEG epochs. Finally, we used 5000 of 5598 EEG epochs and EMG epochs to construct 50000 pairs of training samples, and the remaining 598 pairs were used to construct 5980 pairs of testing samples. The synthesis process followed the OA removal task. Figure 3.3-3 briefly summarizes the above process.



OA (ocular artifact) removal
Training set:
30000 pairs of clean EEG epochs (Labels) and noisy EEG epochs (data)
Testing set:
4000 pairs of clean EEG epochs (Labels) and noisy EEG epochs (data)

MA (myogenic artifact) removal
Training set:
50000 pairs of clean EEG epochs (Labels) and noisy EEG epochs (data)
Testing set:
5980 pairs of clean EEG epochs (Labels) and noisy EEG epochs (data)

Figure 3.3-3 Noisy signal synthesis for model training and testing.

3.4 Experiments and Results

In this section, we first present the experimental hardware and evaluation metrics. Then, the hyperparameter tuning process of the denoising module is described. Finally, we compare the proposed model with other DL and conventional approaches through scoring and visualization.

3.4.1 Hardware and evaluation metrics

All the experiments were implemented using Pytorch and two GeForce GTX 1080 GPUs in a Linux system. The evaluation metrics included the temporal relative root mean square error (T-RRMSE), spectral relative root mean square error (S-RRMSE), and correlation coefficient (CC), as shown:

$$T - RRMSE = \frac{RMS(\mathcal{G}(y) - x)}{RMS(x)} \qquad (3-8)$$

$$S - RRMSE = \frac{RMS\left(PSD(\mathcal{G}(y)) - PSD(x)\right)}{RMS(PSD(x))} \qquad (3-9)$$

$$CC = \frac{Cov(\mathcal{G}(y), x)}{\sqrt{Var(\mathcal{G}(y)) \times Var(x)}} \qquad (3-10)$$

where x and y are the clean EEG epoch and input noisy EEG epoch, respectively; $\mathcal{G}$ indicates the proposed model; PSD is the power spectral density function; Cov and Var are short for the covariance function and variance function. In general, the smaller the T-RRMSE and S-RRMSE values are, the closer to 1 the CC value is, the better the performance is.

3.4.2 Hyperparameter tuning.

To select the hyperparameters (*c, k*) of the denoising module, we performed a 10-fold cross-validation on the given 30000 and 50000 pairs of training samples in the OA and MA removals, respectively. In comparison with the specified validation dataset, a cross-validation strategy can avoid the problems caused by the unreasonable division of the dataset.

In the hyperparameter tuning process, the number of output channels was set as 8, 16, 32, and 64. The filter length for feature extraction was set as 0.05s, 0.1s, 0.2s, 0.3s, 0.4s, and 0.5s, which corresponds to the kernel size of 13, 25, 51, 77, 103, and 127 in the OA (256Hz) task and the kernel size of 25, 51, 103, 155, 207, and 255 in the MA (512Hz) task, respectively.

The CC results of the 10-fold cross-validation from MMNN-1 to MMNN-6 are shown in Figure 3.4-1 and Figure 3.4-2. We can see that the model performance did not significantly improve, but the computational complexity increased when the number of output channels exceeded 32 in both the OA and MA removal tasks. Moreover, the filter lengths of 0.1s (kernel size = 25) and 0.2s (kernel size = 103) were obviously capable of improving the model performance with fewer training parameters in the OA and MA removals. Therefore, we separately chose (32, 25) and (32, 103) as the module's hyperparameters for removing OAs and MAs.

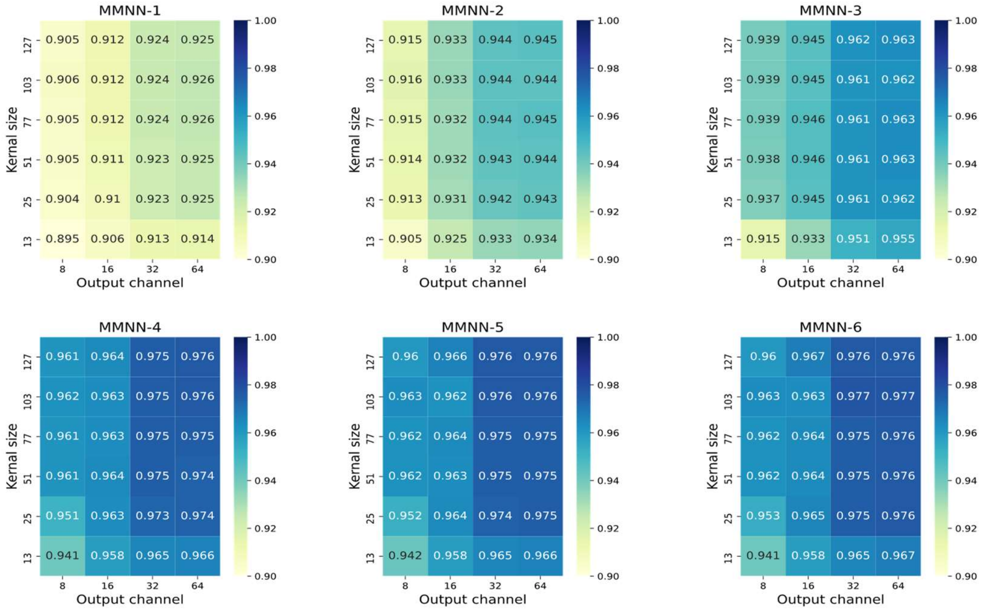


Figure 3.4-1. The CC results of 10-fold cross-validation in the OA removal task.

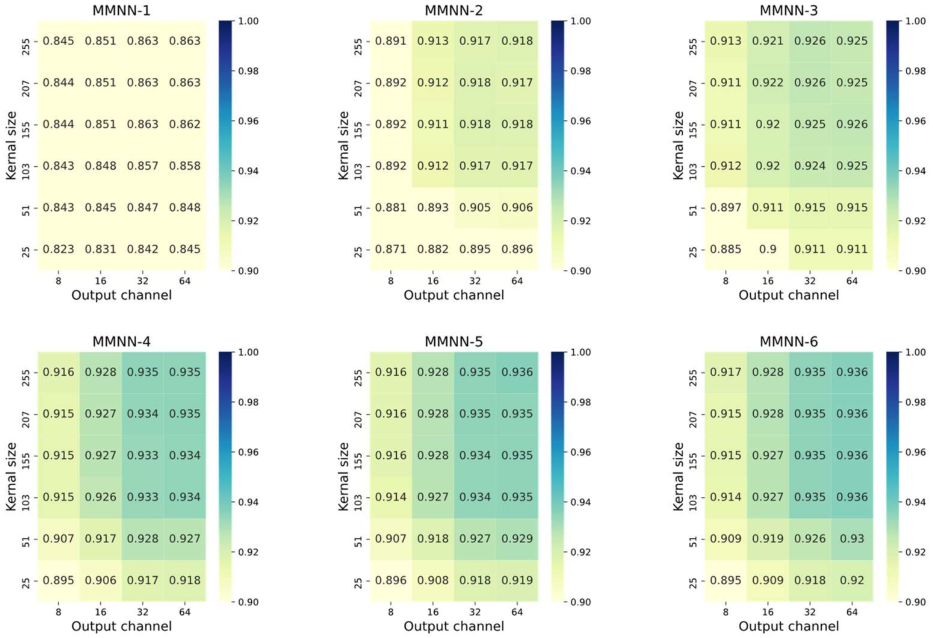


Figure 3.4-2 The CC results of 10-fold cross-validation in the MA removal task.

3.4.3 Results of OA and MA removals

We performed EEG denoising using MMNN-1 to MMNN-6. The reference DL models included fully connected neural network (FCNN), simple convolution neural network (Simple CNN), complex convolution neural network (Complex CNN), and recurrent neural network (RNN) from [86], and novel Convolutional Neural Network (Novel CNN) [87]. To fairly compare the model performance under the same condition, we trained and tested the models [86][87] using the same amount of dataset. The learning rate and batch size were 0.0001 and 128, respectively. The trained parameters of our models at the 10th iteration were used to test the denoising performance.

Table 3.4-1 and Table 3.4-2 show the denoising performance on the 4000 pairs of testing samples in the OA removal and 5980 pairs of testing samples in the MA removal. The results show that the scores of the proposed model can be constantly improved when using one to four denoising modules, whereas more than four denoising modules cannot significantly enhance its performance. Moreover, compared to the best results of the reference models, the proposed model (MMNN-4) reduced the T-RRMSE and S-RRMSE by at least 6.3% and 6.4%, respectively, and improved the CC by at least 3.5% when removing OAs. In the MA removal, it reduced the T-RRMSE and S-RRMSE by at least 6.2% and 6.4%, respectively, and improved the CC by at least 3.3%. These results illustrate that the proposed model performs well on the given database.

Subsequently, through the visualization of the denoised results, we compared the robustness between the proposed model and the top-scoring reference models for OA and MA removals (Complex CNN and Novel CNN). As shown in Figure 3.4-3 and Figure 3.4-4, we presented the signal deviation in the time and frequency domains between the denoised results and the sample labels by calculating the absolute values of the noise-free epoch minus the denoised epoch. From the deviation results of the OA and MA removals within the 95% confidence interval, we can observe that the signal deviation of the proposed model (MMNN-4) is closer to the horizontal axis (noise-free situation) and exhibits a smaller range of deviation than the other competitors in both the time and frequency domains, which confirms the relative robustness of the proposed model.

Table 3.4-1 Average denoising performance in the OA removal task.

Approach	T-RRMSE	S-RRMSE	CC
FCNN	0.367	0.387	0.906
Simple CNN	0.359	0.361	0.916
Complex CNN	0.336	0.343	0.923
RNN	0.411	0.389	0.900
MMNN-1	0.321	0.323	0.925
MMNN-2	0.301	0.309	0.936
MMNN-3	0.289	0.295	0.941
MMNN-4	0.273	0.279	0.958
MMNN-5	0.273	0.277	0.959
MMNN-6	0.273	0.276	0.959

Table 3.4-2 Average denoising performance in the MA removal task.

Approach	T-RRMSE	S-RRMSE	CC
FCNN	0.367	0.387	0.906
Simple CNN	0.359	0.361	0.916
Complex CNN	0.336	0.343	0.923
RNN	0.411	0.389	0.900
MMNN-1	0.321	0.323	0.925
MMNN-2	0.301	0.309	0.936
MMNN-3	0.289	0.295	0.941
MMNN-4	0.273	0.279	0.958
MMNN-5	0.273	0.277	0.959
MMNN-6	0.273	0.276	0.959

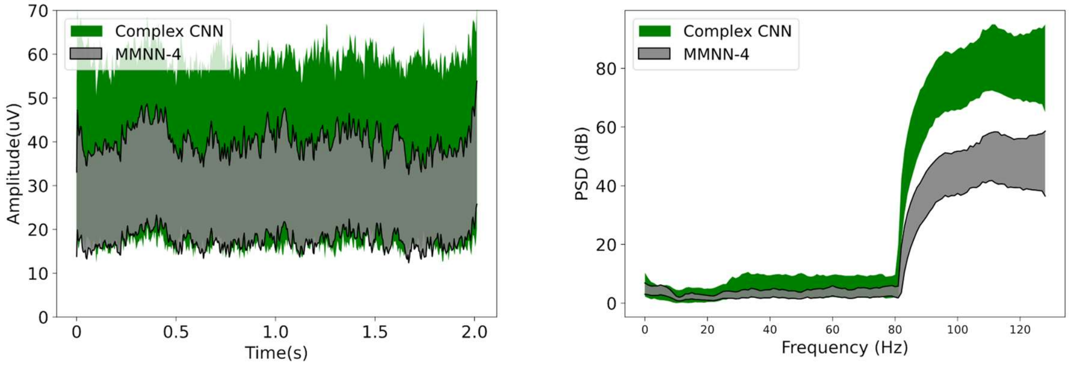


Figure 3.4-3 Distribution of signal deviation in the time and frequency domains for OA removal (Confidence interval = 0.95).

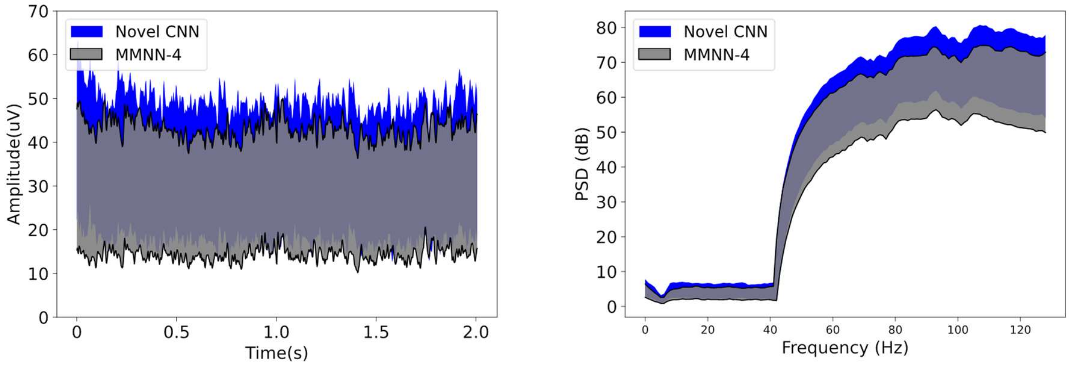


Figure 3.4-4 Distribution of signal deviation in the time and frequency domains for MA removal (Confidence interval = 0.95).

3.4.4 Proposed models vs. Conventional models

We further compared the proposed model with three conventional models: Regression[64], ICA[109], and SSP[111]. These models are classic EEG denoising approaches applied to MNE toolbox[112]. Figure 3.4-5 and Figure 3.4-6 show the score distributions of the OA removal (4000 testing epochs) and MA removal (5980 testing epochs), respectively, where the proposed model achieves higher CC and smaller T-RRMSE and S-RRMSE scores than the conventional ones. According to the ANOVA results with Holm-Bonferroni correction, the performance differences between the proposed model and the classical approaches are significant (all *p*-values < 0.001) in both the OA and MA removals.

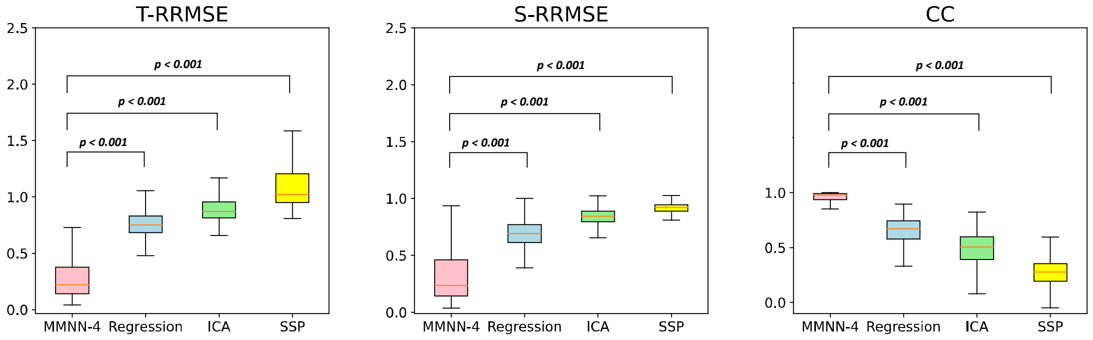


Figure 3.4-5 Performance comparison between the proposed model and traditional models in the OA removal.

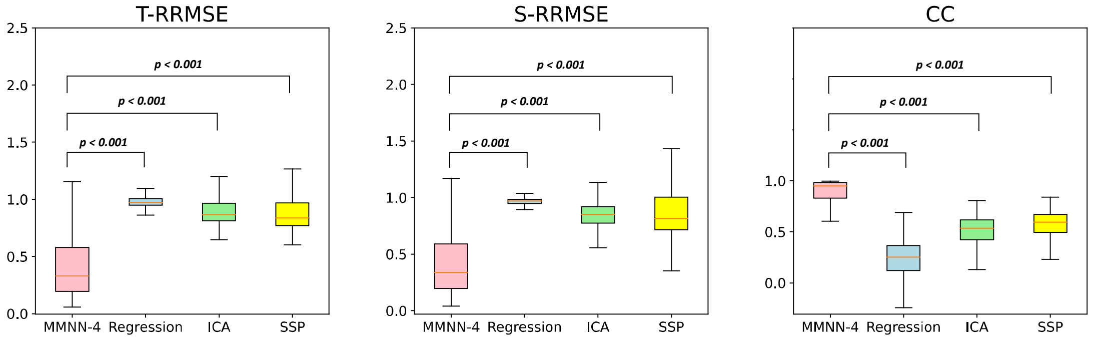


Figure 3.4-6 Performance comparison between the proposed model and traditional models in the MA removal.

3.4.5 OA and MA removals on limited training data

In the applications of DL-based EEG denoising, sufficient high-quality training data are usually unavailable. Therefore, we investigated the robustness of DL models when using limited training data, as shown in Figure 3.4-5 and Figure 3.4-6. The proposed MMNN-4 was compared with the top-scoring reference models for the OA and MA removals, Complex CNN, and Novel CNN, respectively, where 10 to 100% of the training data were separately selected from the given database for network learning, and the training iterations and parameters were consistent with the former settings. The results show that the proposed model always has a superior performance over its competitors when using the same amount of training data both for the OA and MA removals. Notably, our model can reach scores similar to the reference ones using only 60% of the training data when removing OAs and MAs.

Table 3.4-3 Denoising performance with limited training data in the OA removal.

Training data	Testing data	Complex CNN (T-RRMSE/S-RRMSE/CC)	MMNN-4 (T-RRMSE/S-RRMSE/CC)
3000	4000	0.641/0.557/0.737	0.562/0.501/0.790
6000	4000	0.543/0.507/0.810	0.495/0.459/0.840
9000	4000	0.481/0.462/0.849	0.437/0.417/0.871
12000	4000	0.445/0.454/0.872	0.382/0.375/0.902
15000	4000	0.427/0.442/0.879	0.346/0.351/0.916
18000	4000	0.410/0.438/0.892	0.322/0.340/0.929
21000	4000	0.381/0.393/0.902	0.307/0.328/0.935
24000	4000	0.366/0.377/0.908	0.298/0.303/0.945
27000	4000	0.342/0.361/0.917	0.287/0.297/0.950
30000	4000	0.336/0.343/0.923	0.273/0.279/0.958

Table 3.4-4 Denoising performance with limited training data in the MA removal.

Training data	Testing data	Novel CNN (T-RRMSE/S-RRMSE/CC)	MMNN-4 (T-RRMSE/S-RRMSE/CC)
5000	5980	0.770/0.680/0.597	0.615/0.579/0.745
10000	5980	0.597/0.554/0.751	0.537/0.518/0.804
15000	5980	0.555/0.537/0.776	0.503/0.494/0.826
20000	5980	0.516/0.515/0.814	0.474/0.453/0.843
25000	5980	0.499/0.508/0.829	0.446/0.437/0.855
30000	5980	0.485/0.489/0.840	0.430/0.430/0.865
35000	5980	0.491/0.473/0.844	0.420/0.410/0.872
40000	5980	0.462/0.464/0.851	0.415/0.404/0.874
45000	5980	0.457/0.452/0.857	0.395/0.395/0.883
50000	5980	0.448/0.442/0.863	0.386/0.378/0.896

3.5 Discussion and Conclusion

In this paper, we proposed a novel DL-based EEG denoising model called MMNN. This model achieved smaller T-RRMSE and S-RRMSE scores and higher CC scores than the other models when removing OAs and MAs. It can reach a performance similar to that of the reference DL models with only 60% of the training data. Through the visualization of signal deviation distribution, the performance differences between the reference and the proposed models are clearly observed in the time and frequency domains.

Overall, the proposed model has a superior performance compared to the reference DL models. There are two reasons for this. First, the proposed model enables constantly providing more purified input signals for denoising modules. As shown in Figure 3.5-1 and Figure 3.5-2, we present the signal deviation distribution between the inputs of four denoising modules and clean signals. In Figure 3.5-1, the OAs in the range of 0-80 Hz were rapidly suppressed using two denoising modules, and the OAs above 80 Hz were gradually reduced when more modules were used. In Figure 3.5-2, the MAs of the input signals were suppressed by degrees from the first to the fourth module. Second, the parallel architecture of the proposed model allows the gradients to flow through each denoising module directly in the backpropagation, thereby avoiding the vanishing gradient problem and enhancing the network learning ability.

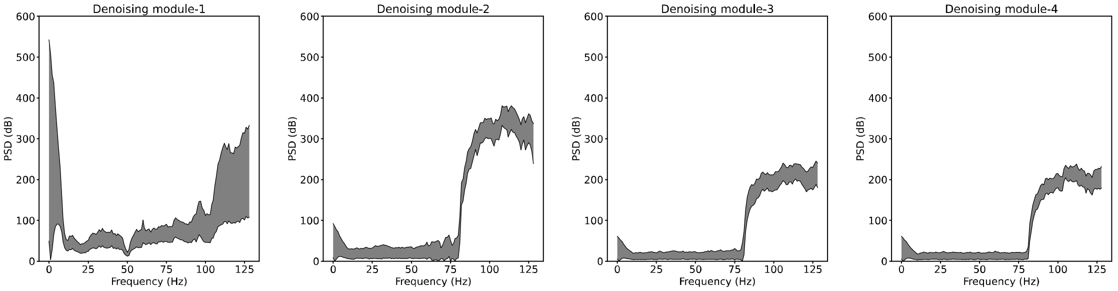


Figure 3.5-1 Deviation distribution between the input signals of denoising modules and clean signals for OA removal (Confidence interval = 0.95).

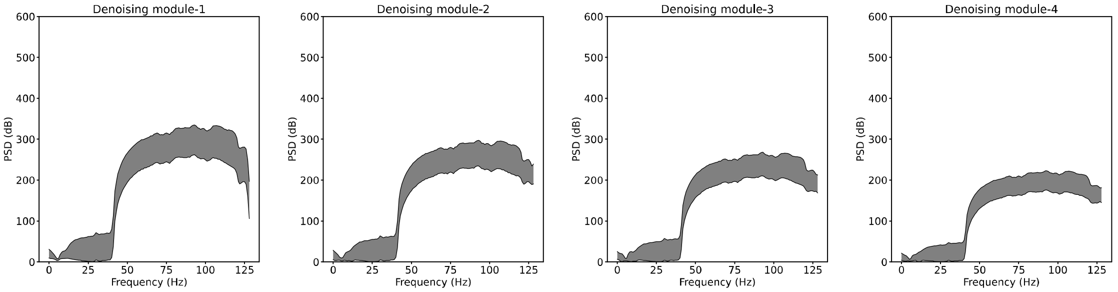
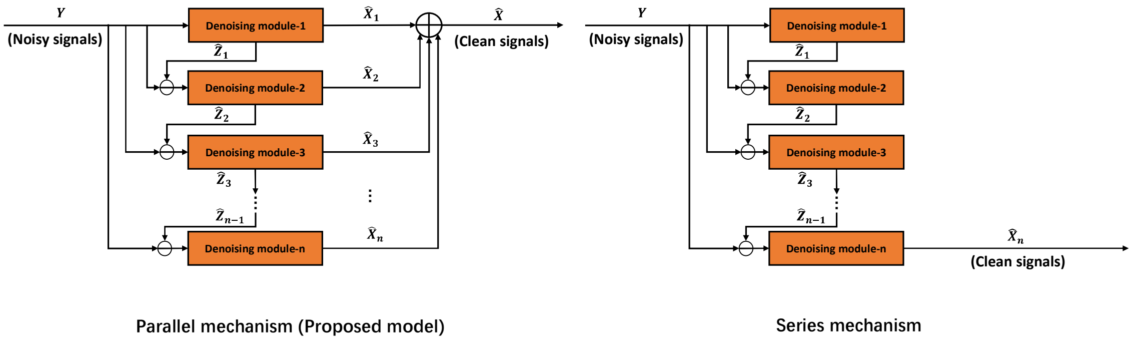


Figure 3.5-2 Deviation distribution between the input signals of denoising modules and clean signals for MA removal (Confidence interval = 0.95).

For further clarification, the parallel and series mechanisms of the denoising modules are presented in Figure 3.5-3. The training and testing losses of the two mechanisms (batch size = 128 and learning rate = 0.0001) using the given database are shown in Figure 3.5-4 and Figure 3.5-5. The followings were observed, respectively: 1) both the two models converge in 10 iterations; 2) the training and testing losses of the parallel mechanism are smaller than those of the series mechanism when more than two denoising modules are assembled within our model, which illustrates that the parallel mechanism of the proposed model possesses a stronger learning capacity; 3) for the series mechanism, the network learning capacity is weakened when more denoising modules are stacked in the model, which is possibly caused by the vanishing gradient problem in the learning process; 4) in contrast with the series mechanism, the parallel mechanism can improve the learning capacity when more denoising modules are used. However, there was a limitation to the improvement of network learning. We can see that the training and testing losses of MMNN-4, MMNN-5, and MMNN-6 are almost the same for the parallel model, which explains their similar scores in the experiment. Furthermore, the loss comparison between the proposed MMNN-4 and the other DL models is given in Figure 3.5-6, where our model has smaller training and testing losses and can converge faster than the others, which is possibly the reason why it performs well with fewer training data in both the OA and MA removals.

In the future, there are some challenges worth exploring using the proposed model. Specifically, we used the denoising modules of the different filter sizes in the OA and MA removals. Whether the filter size of feature extraction is caused by noise feature differences should be further studied. Moreover, OAs and MAs are entangled with motion artifacts in a real EEG epoch, however, for the mixed signals, there is no available public database to evaluate the model performance [112]. Given that the proposed model offers significant advantages over the conventional and DL models in this study, the related research is within the scope of further work.



Parallel mechanism (Proposed model)

Series mechanism

Figure 3.5-3 Parallel and series mechanisms.

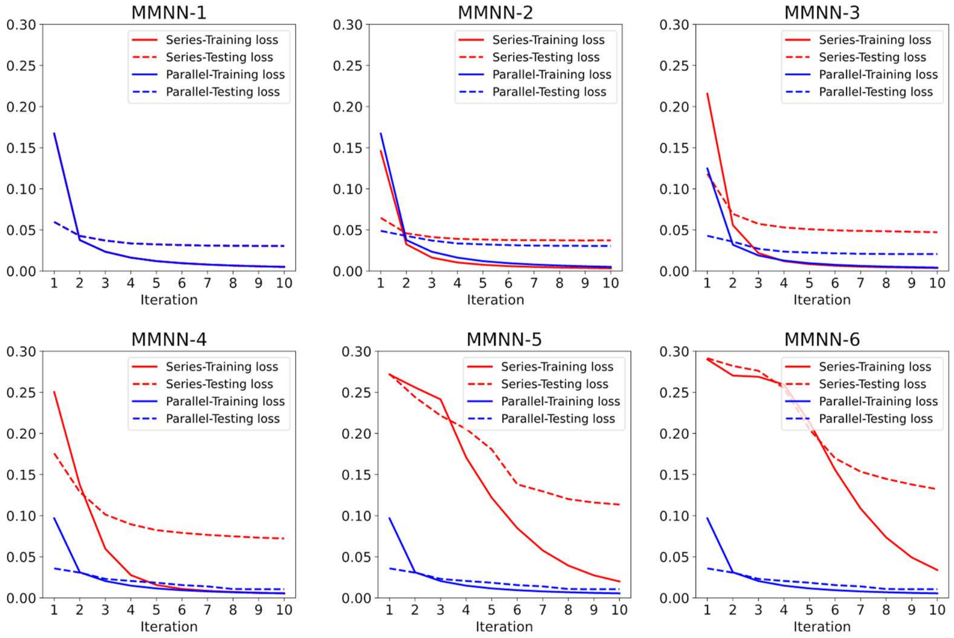


Figure 3.5-4 Training and testing losses using the given database in the OA removal.

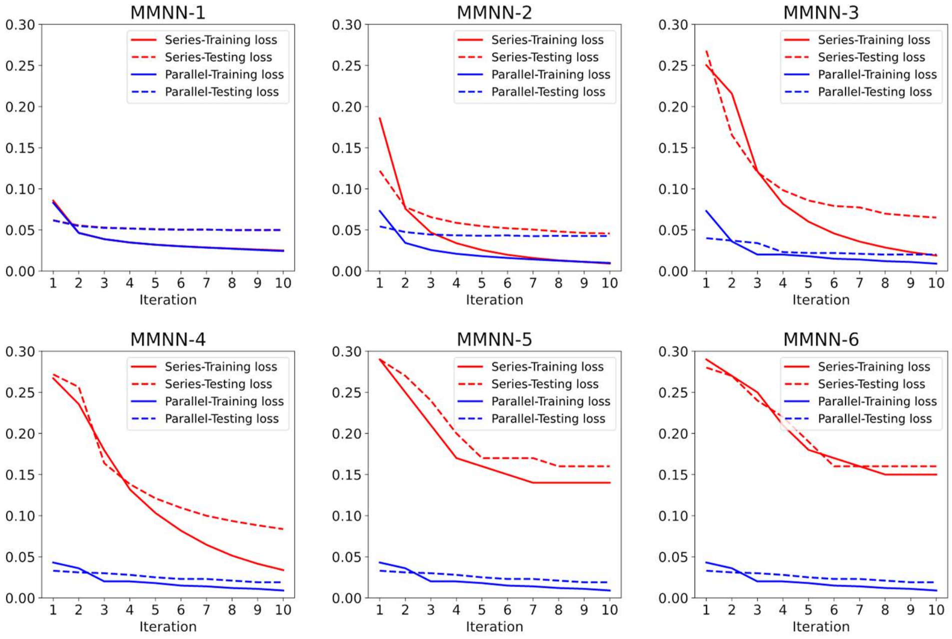


Figure 3.5-5 Training and testing losses using the given database in the OA removal.

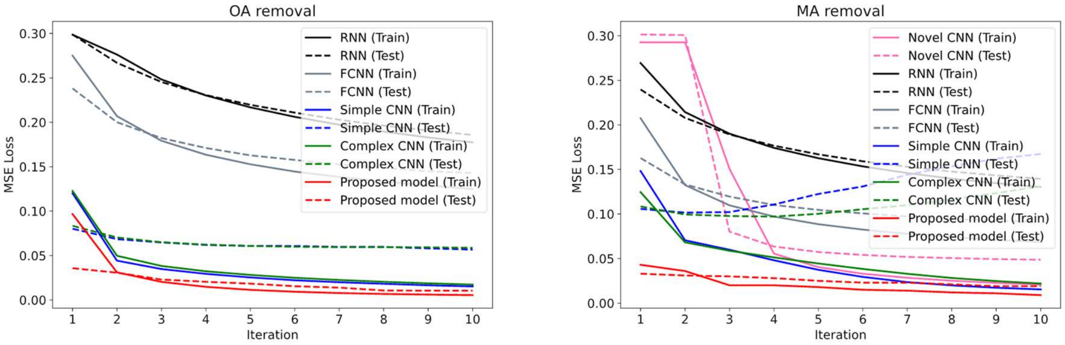


Figure 3.5-6 . Loss comparison between the proposed model and the other DL models.

4. Multi-pooling 3D Convolutional Neural Networks for fMRI Classification

4.1 Background

Functional magnetic resonance imaging (fMRI) is a non-invasive and reliable technique that measures the small changes in blood flow caused by brain activities. fMRI data have relatively high spatial resolution and availability, thus providing a solution to reveal neural activities of brain regions under visual stimuli [113][114][115]. However, there are many difficulties in analyzing fMRI data because the data are highly dimensional, structurally complex, and have intra- and inter-subject variability [116][117][118]. For fMRI classification tasks, the challenges are four aspects as follows: 1) background noise; 2) the complex spatial relevance between features; 3) feature redundancy; 4) incomprehensible but rich brain features [119][120][121].

Various techniques have been proposed to resolve these problems, such as general linear model (GLM) [122], multivoxel pattern analysis (MVPA) [123], and principal component analysis (PCA) [124]. GLM is skilled at separating stimulus-induced signals from noise; MVPA can analyze spatial activity patterns from the region of interest (ROI) of the cerebral cortex; PCA removes irrelevant features through dimensionality reduction.

Compared with the conventional techniques, CNN is an end-to-end learning model that can leave out all the steps from the input fMRI data to the classification results. In the learning process, CNN architecture can capture sufficiently general features from the training data and has a relatively high tolerance for data quality. Moreover, CNN is a multi-layer perception mechanism that presents a generalization like human visuals [125][126]. Therefore, it can process massive complex spatial features and understand the deep relevance among them. CNN architecture can process one-dimensional (1D) sequences, two-dimensional (2D) images, or 3D volumes, corresponding to 1D, 2D, or 3D CNNs for brain decoding, respectively. Among them, the 3D CNN structure can process neural activity patterns from the raw spatial domain of fMRI voxels, exhibiting a superior capability to decode visual brain states. As one of the 3D CNN-based applications, Watanabe et al. [127] revealed the flow of human visual processing of categorical and sub-categorical information. In detail, the researchers collected the brain activities from several subjects under three visual stimuli tasks (faces vs. objects, male faces vs. female faces, and natural objects vs. artificial objects). Afterward, they implemented CNN-based classification to fMRI and electroencephalogram (EEG) data and investigated the relevant brain regions regarding the classification results.

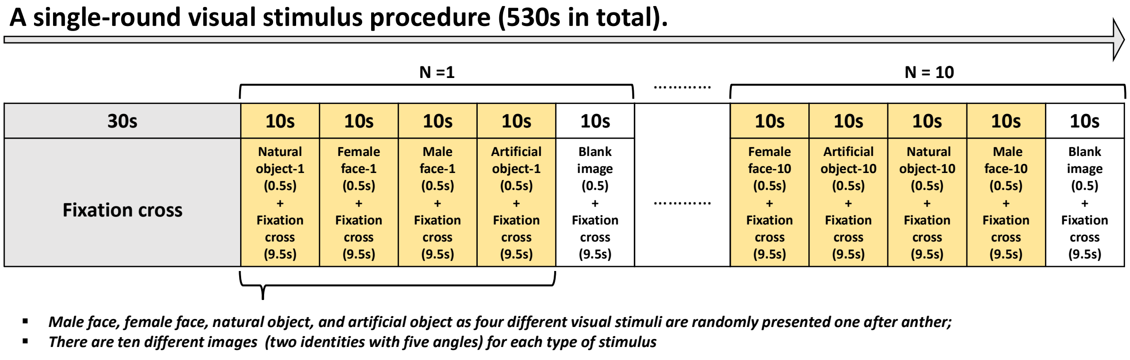
Considering that the classification results are critical to explaining human visual processing, the fMRI classification model used in [127] attracts our attention in two aspects: 1) the training data and test data need to be processed by averaging the trials under the same visual stimuli, otherwise the classification accuracy would drop significantly; 2) the brain neural activities are relatively slight in the face sub-categorical and object sub-categorical classifications tasks, so the classification model does not reach statistical significance in accuracy even when features are enhanced by averaging nine trials.

Therefore, this paper proposed a novel multi-pooling 3D convolutional neural network (MP3DCNN) to improve the classification performance for categorical and sub-categorical classifications. Our solutions are as follows: 1) we design a feature extraction unit using three 3D convolutions with average 3D pooling layers. The 3D pooling layers perform constant generalization and compression to the rich extracted features from the 3D convolutions, which aims to fight against the local feature redundancy in feature extraction. 2) moreover, the first two 3D convolutions each have a further branch of pooling connection, respectively. The features extracted from the branches are merged into the mainchain of the model. Finally, the classifier makes a joint decision based on the extracted features from all 3D convolutions.

4.2 Materials

4.2.1 Visual stimulus task

In their work, 53 healthy subjects (34 males and 19 females, ages 18 to 26) participated the visual stimulus task, and each one experienced averaged 9.96 ± 2.88 rounds of visual stimuli. In each round of the visual stimulus procedure, the participant perceived four types of visual stimuli, involving male face, female face, natural object, and artificial object. And there were ten images (two identities with five angles) for each type of stimulus.


A single-round visual stimulus procedure (530s in total).

	N =1					....	N = 10				
30s	10s	10s	10s	10s	10s		10s	10s	10s	10s	10s
Fixation cross	Natural object-1 (0.5s) + Fixation cross (9.5s)	Female face-1 (0.5s) + Fixation cross (9.5s)	Male face-1 (0.5s) + Fixation cross (9.5s)	Artificial object-1 (0.5s) + Fixation cross (9.5s)	Blank image (0.5) + Fixation cross (9.5s)	....	Female face-10 (0.5s) + Fixation cross (9.5s)	Artificial object-10 (0.5s) + Fixation cross (9.5s)	Natural object-10 (0.5s) + Fixation cross (9.5s)	Male face-10 (0.5s) + Fixation cross (9.5s)	Blank image (0.5) + Fixation cross (9.5s)

- *Male face, female face, natural object, and artificial object as four different visual stimuli are randomly presented one after anther;*
- *There are ten different images (two identities with five angles) for each type of stimulus*

Figure 4.2-1 Example of a single-round stimulus procedure.

Figure 4.2-1 is an example of a single-round stimulus procedure, where the fixation cross was presented in the first 30s. Next, the participant perceived a random visual stimulus (an image of a male face, female face, natural object, or artificial object) for 0.5s followed by a fixation cross for 9.5s. In this period, he or she needed to click the button corresponding to the stimulus. After four types of stimuli, a blank image for 0.5s and a fixation cross for 9.5s were presented to regulate the cerebral hemodynamic responses. By ten times from four stimuli to one blank image, the participant perceived 40 stimuli, costing 530s in a single-round stimulus procedure.

4.2.2 fMRI data collection and pre-processing

A 3T MRI scanner (Siemens Prisma, Germany) with a 64-channel head coil scanned the brain activities of all subjects in the visual stimulus task. The T1-weighted 3D fMRI volumes were acquired using a magnetization-prepared rapid gradient-echo (MP-RAGE) sequence (TR: 2,400ms; TE = 2.32ms; FA: 8 deg; 192 slices; slice thickness: 0.9 mm; in-plane resolution: 0.93 × 0.93 mm). Through SPM12 [128] , the collected fMRI volumes were realigned, co-registered, normalized to the standard Montreal Neurological Institute (MNI) template, and resampled to 2-mm isotropic voxels.

4.2.3 fMRI data cleaning and averaging

CNN-based classification tasks used the fMRI volume collected from 4s after each visual stimulus presentation. The collected fMRI dataset was cleaned according to the following criteria: 1) the subject was excluded from the study whose mean reaction time exceeded 1.5s or whose error rate was higher than 20%; 2) the fMRI volume was discarded when a subject responded to a visual stimulus for longer 2s or gave an error response; 3) considering the possibility of combining EEG analysis, if excessive (EEG) noises were detected manually from the heil coil, the fMRI volume in this range was not adopted. Finally, 17306 fMRI volumes from 50 subjects were available for our CNN-based classification tasks, including 4453, 4399, 4214, and 4240 volumes corresponding to the visual stimuli of man face, female face, natural object, and artificial object, respectively.

During the fMRI scanning, intense noises originated from irrelevant neural activities and the scanner. The fMRI dataset of each subject was thus multi-fold averaged to improve the data quality used for classification tasks. To compare the performance of the classification model between unaveraged and averaged data, the averaged data needed to be augmented to the original data amount.

4.3 Methods

The proposed MP3DCNN is composed of feature extraction, feature combination, and classifier, as shown in Figure 4.3-1. This model is an end-to-end learning approach to decode visual brain states using fMRI data.

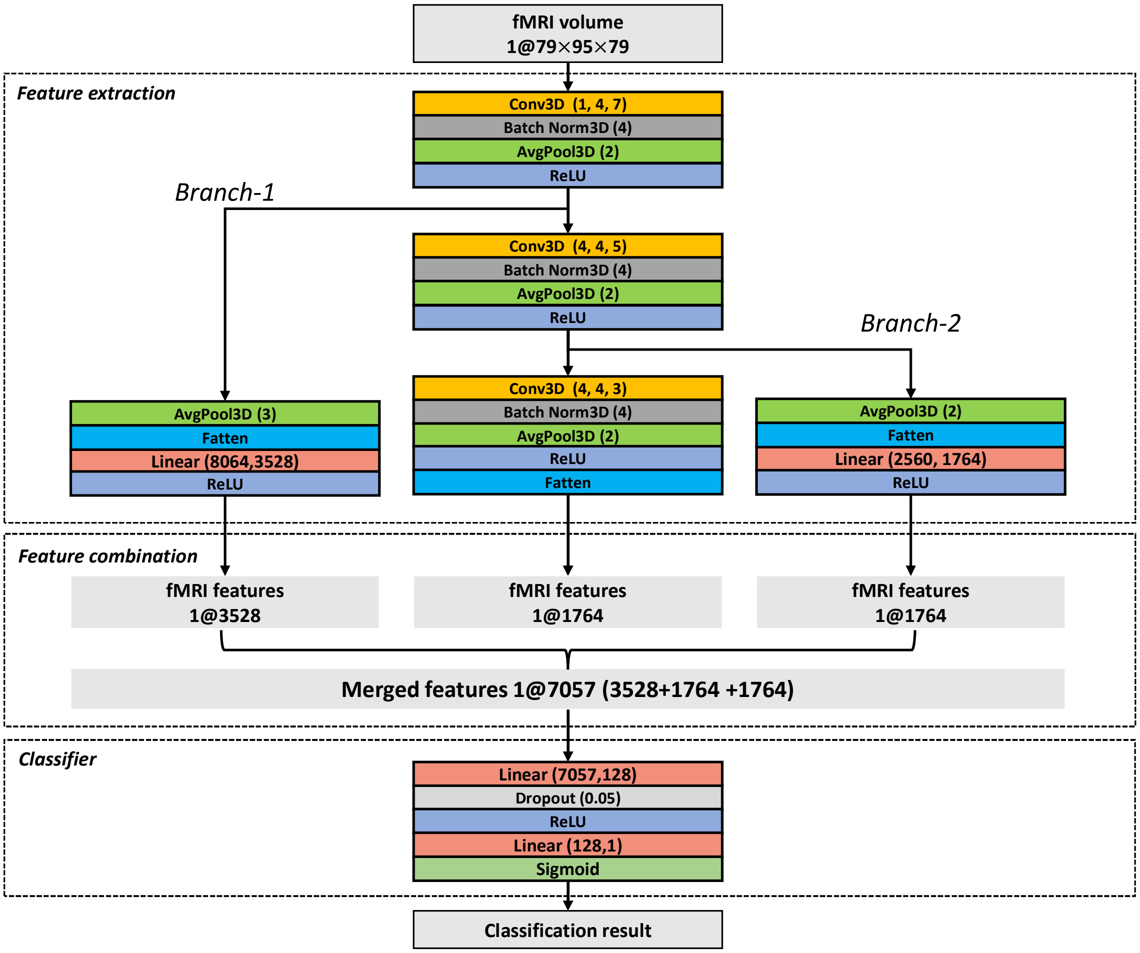


Figure 4.3-1 The structure of the proposed multi-pooling 3D convolutional neural network (MP3DCNN).

4.3.1 Feature extraction

The feature extraction has a mainchain and two branches. Its mainchain is three 3D convolutions connected in series, and each 3D convolution combines with a batch normalization [25], an average 3D pooling layer, and a rectified linear unit (ReLU). Batch normalization can accelerate network learning by reducing the internal covariate shift. The average 3D pooling can compress the extracted features, and the ReLU can improve the model's nonlinearity.

The two branches connect with the first and the second 3D convolution, respectively. They combine an average 3D pooling layer and a linear layer to generalize further and compress features. The relevant hyperparameters are as follows: 1) **3D Convolution:** An input fMRI volume can be represented as a matrix with the size of $x \times y \times z$, where x, y, and z are 79, 95, and 79, respectively. An 3D convolution is a 3D tensor with the hyperparameter of $(c_{in},\ c_{out}, k_{conv})$, where c_{in} depends on the number of the input feature maps (default 1 to the input fMRI volume), c_{out} determines the number of output feature maps, and $k_{conv} \times k_{conv} \times k_{conv}$ is the kernel size of 3D convolutional calculation. Batch normalization is applied on c dimension, where c equals to the c_{out} of the former-layer 3D convolution. The hyperparameters of the three 3D convolutions in the feature extraction are $(1,4,7)$, $(4,4,5)$ and $(4,4,3)$, respectively. 2) **Average 3D pooling.** Suppose the i_{th} 3D convolution outputs a 3D tensor with the size of $c_i \times h_i \times w_i \times d_i$, where c_i is the number of feature maps, h_i, w_i, and d_i indicate the height, width and depth of each feature map, respectively. An average 3D pooling with a hyperparameter of k_{pool} can perform the average pooling on each feature map using a kernel size of $k_{pool} \times k_{pool} \times k_{pool}$. k_{pool} is set to 2 for the three 3D convolutions in the mainchain. In the branch-1 and branch-2, k_{pool} is set to 3 and 2, respectively. 3) **Linear layer.** Linear layer with the hyperparameters of (f_{in}, f_{out}) can convert 1D tensors of arbitrary size. The output of the third 3D convolution is $c_3 \times h_3 \times w_3 \times d_3$, which is flattened into a 1D tensor with the size of 1×1764. To obtain the features with the size of 1×3528 and 1×1764 from the first and second 3D convolutions, the hyperparameters of the linear layer are set to $(8064,3528)$ and $(2569,1764)$ respectively.

4.3.2 Feature combination and classifier

In the proposed model, the mainchain of the feature extraction outputs the third-level feature representations of fMRI data, and the branch-1 and -2 make the first and second feature representations to be fed into the classifier also. Based on this structure, the network has the potential to provide a sophisticated decision using multi-level features in classification problems. Moreover, this design adds a further learning path for the first and second 3D convolutions in the backpropagation, respectively, thereby improving the model's efficiency.

The classifier is composed of two linear layers with the parameters of (7057,128) and (128,1), respectively. The dropout is used to prevent overfitting in the training processing, and the probability is set to 0.05. Sigmoid activation [130] is the last layer of the proposed model, and binary cross entropy (BCE) is the loss function for fMRI classification in our study.

4.3.3 Model training, validation, and testing

The proposed model is implemented in Pytorch, trained and tested on an NVIDIA GeForce RTX 3060 (12GB) GPU in a Linux system. To evaluate the model performance independently, the fMRI volumes from 5 of 50 subjects are detached as the test dataset, which would not be used for tuning the model's hyperparameters.

According to the previous study, the amount of the test dataset is equalized by discarding data, and there are 1468 (734 faces vs. 734 objects), 772 (386 male faces vs. 386 female faces), and 718 (359 natural objects vs. 359 artificial objects) trials for the categorical classification, face sub-categorical classification, and object sub-categorical classification, respectively.

We performed 9-fold cross-validation on the training dataset from the 45 leftover subjects. Specifically, 45 subjects were randomly divided into nine groups according to each group of five subjects, the data from each group only were used for model validation once, and data from the left eight groups were used for model training.

In the training stage, each of the nine models was iterated 25 times with a batch size of 64, and the learning rate was 0.00001. Within the 25 iterations, the model parameters with the highest accuracy score on the validation dataset are retained for model testing. A majority voting scheme [132] was applied for fMRI classification on the test dataset. As is shown in Figure 4.3-2, the final prediction is the output when more than four models predict consistently.

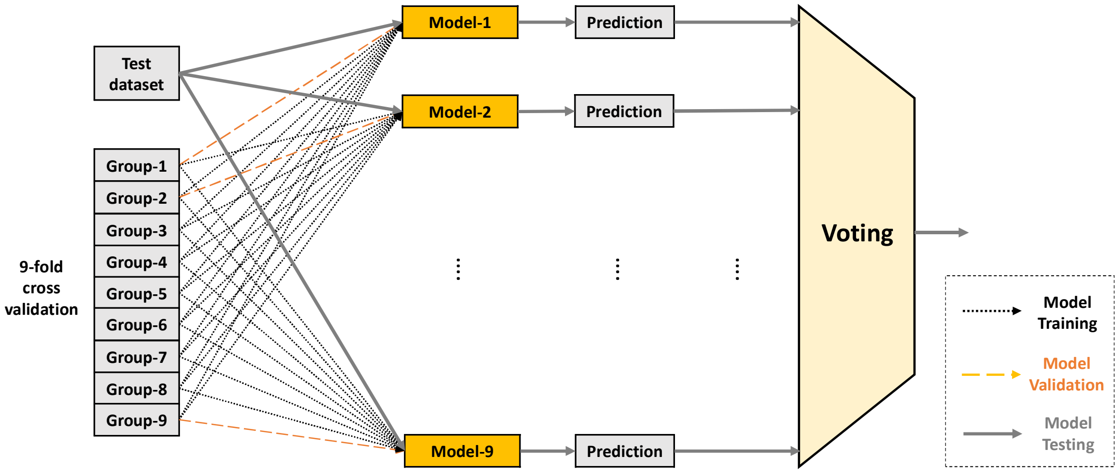


Figure 4.3-2 Model training, validation, and testing.

4.4 Experiments and Results

We compared the results of the categorical (face vs. object), face sub-categorical (male face vs. female face), and object sub-categorical (natural object vs. artificial object) classification, respectively. Thanks to the open-source code provided by the study [127], the proposed and existing models were able to be trained and tested in our hardware environment.

4.4.1 Face vs. Object

Table 4.4-1 and Table 4.4-2 are the classification results of the unaveraged and 9-fold averaged fMRI data on the categorical (face vs. object) classification, respectively. And Figure 4.4-1 and Figure 4.4-2 reveal the model training processes of the proposed and existing models, where the model parameters of the best validation accuracy within 25 epochs are used for model testing.

Table 4.4-1 Classification performance comparison on Faces vs. Objects (unaveraged fMRI data).

Group	Validation accuracy (Proposed / Existing)		Test accuracy (Proposed / Existing)	
1	**72.256%**	54.222%	**71.458%**	55.381%
2	**71.709%**	54.022%	**69.346%**	48.978%
3	**70.962%**	56.35%	**71.866%**	54.087%
4	**69.526%**	54.182%	**68.801%**	51.294%
5	**71.812%**	54.484%	**72.207%**	52.793%
6	**68.851**%	55.271%	**69.959%**	54.768%
7	**74.609%**	55.473%	**72.548%**	52.112%
8	**74.857%**	54.951%	**71.117%**	53.542%
9	**76.369%**	53.383%	**70.368%**	52.997%
Voting	\		**71.458%**	56.54%

***Bold**: Better score in comparison between the proposed and existing models

Table 4.4-2 Classification performance comparison on Faces vs. Objects (9-fold averaged fMRI data).

Group	Validation accuracy (Proposed / Existing)		Test accuracy (Proposed / Existing)	
1	**94.27%**	84.741%	**90.599%**	79.36%
2	**92.367%**	84.278%	**86.853%**	79.223%
3	**87.3%**	84.649%	**90.736%**	82.766%
4	**83.591%**	81.034%	**89.986%**	79.36%
5	**93.228%**	82.245%	**89.237%**	79.973%
6	**88.267%**	80.226%	**87.193%**	83.583%
7	**94.267%**	90.841%	**90.054%**	82.698%
8	**93.831%**	87.921%	**89.578%**	83.106%
9	**94.307%**	87.594%	**85.627%**	82.834%
Voting	\		**89.578%**	85.49%

***Bold**: Better score in comparison between the proposed and existing models

In Table 4.4-1, the validation accuracy is improved from 13.58% to 22.986% and the test accuracy is improved from 15.191% to 20.436% using the unaveraged data. Figure 4.4-1 presents that the proposed model has a smaller training and validation losses in all nine training groups. It shows that the proposed model has a stronger learning ability than the existing model for high-noisy fMRI data. Finally, the voting result of the proposed model reaches 71.458% using the unaveraged data, which improves 14.918% over the existing one in the test accuracy.

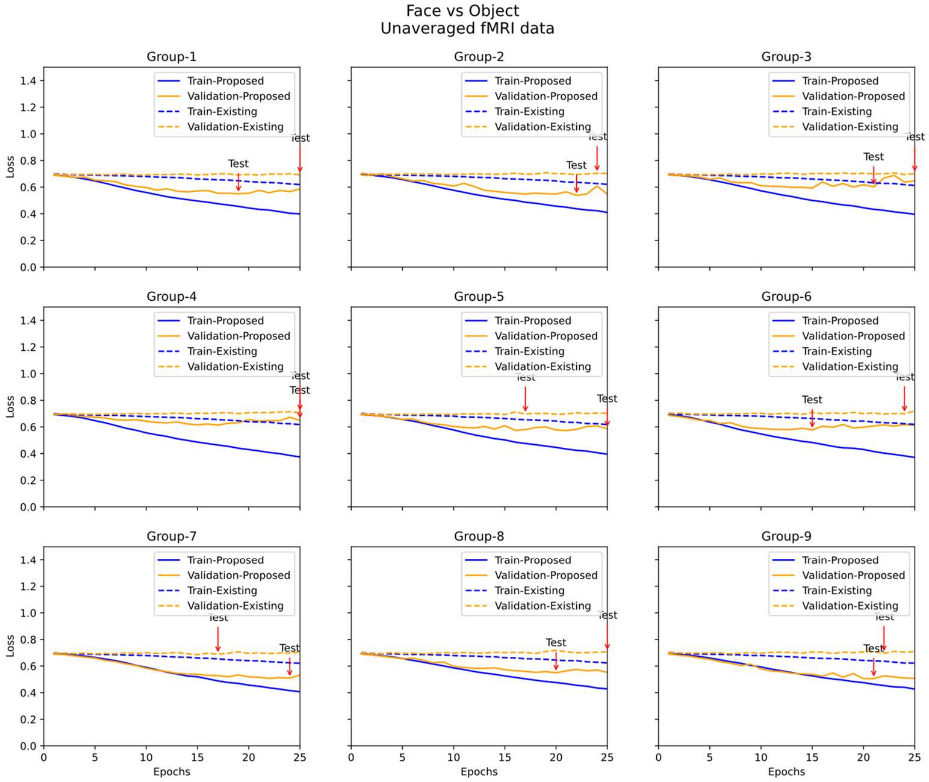


Figure 4.4-1 Training and validation losses within 25 iterations on Faces vs. Objects (unaveraged data).

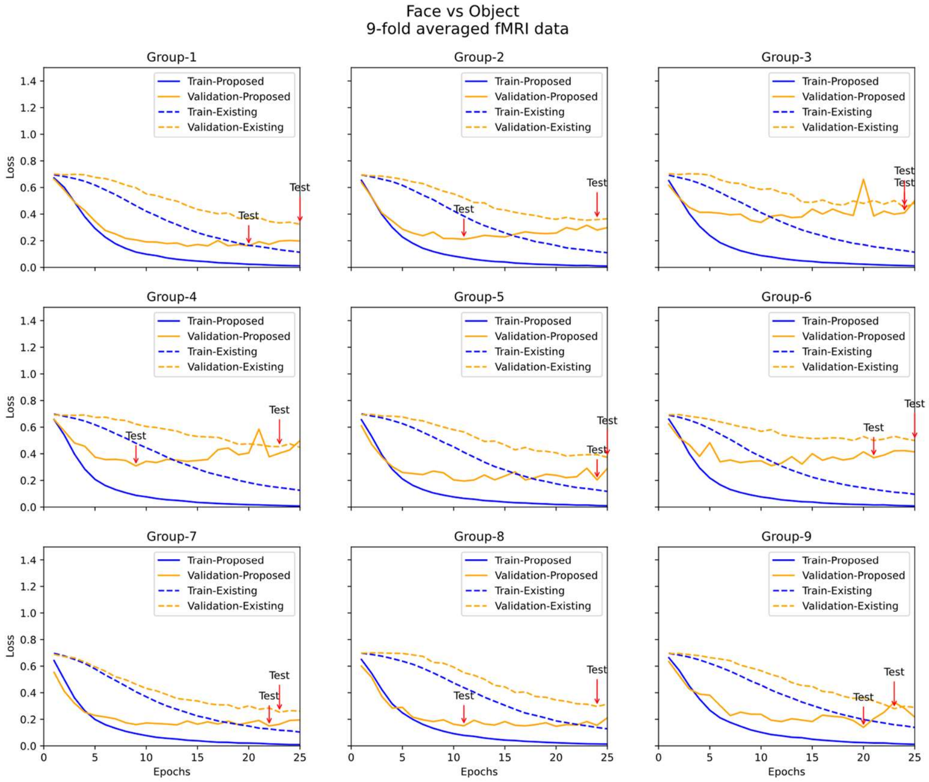


Figure 4.4-2 Training and validation losses within 25 iterations on Faces vs. Objects (9-fold averaged data).

When the 9-fold averaged fMRI data are used, the training and validation losses are as shown in Figure 4.4-2. The proposed model improves the validation accuracy and test accuracy from 2.557% to 10.983% and from 2.793% to 11.239%, respectively, as shown in Table 4.4-2. The proposed model achieves the test accuracy of 89.578% by voting, which is 4.088% higher than the existing one. By comparing the voting accuracies between Table 4.4-1 and Table 4.4-2, we can see that the existing model has degraded by 28.95% without 9-fold data averaging. By contrast, the proposed model has devalued by 18.12% for the voting accuracy, which shows the robustness of the proposed model in the categorical (face vs. object) classification.

4.4.2 Male face vs. Female face

Table 4.4-3 and Table 4.4-4 are the classification results of the unaveraged and 9-fold averaged fMRI data on the face sub-categorical (male face vs. female face) classification, respectively. Compared to the existing model, the proposed model improves the validation accuracy from 1.114% to 1.788% in the 1st, 3rd, 5th, 6th, and 8th training groups, but it shows a decrease from 0.663% to 1.765% for the 2nd, 4th, 7th, and 9th groups.

Table 4.4-3 Classification performance comparison on Male face vs. Female face (unaveraged fMRI data).

Group	Validation accuracy (Proposed / Existing)		Test accuracy (Proposed / Existing)	
1	**52.984%**	51.832%	**50.777%**	48.964%
2	52.038%	**52.701%**	**52.073%**	47.539%
3	**54.291%**	52.595%	49.223%	**51.943%**
4	53.984%	**55.357%**	**50.13%**	47.15%
5	**55.432%**	54.318%	**52.073%**	47.28%
6	**54.022%**	52.234%	**51.036%**	48.316%
7	52.647%	**54.412%**	51.425%	**51.943%**
8	**55.942%**	54.493%	**55.57%**	52.591%
9	53.215%	**53.991%**	**51.554%**	49.611%
Voting	\		**52.72%**	49.352%

***Bold:** Better score in comparison between the proposed and existing models

Table 4.4-4 Classification performance comparison on Male face vs. Female face (9-fold averaged fMRI data).

Group	Validation accuracy (Proposed / Existing)		Test accuracy (Proposed / Existing)	
1	**59.476%**	54.764%	**52.979%**	48.964%
2	**61.611%**	55.735%	50.389%	**56.477%**
3	**63.373%**	54.79%	52.073%	**54.922%**
4	**58.242%**	53.434%	**54.663%**	49.87%
5	**61.838%**	54.875%	**55.57%**	52.073%
6	**60.973%**	53.128%	**49.741%**	48.964%
7	**59.02%**	52.059%	**53.109%**	49.482%
8	**61.304%**	56.812%	55.829%	**56.347%**
9	**62.417%**	56.098%	**56.606%**	52.332%
Voting	\		**55.311%**	53.627%

***Bold:** Better score in comparison between the proposed and existing models

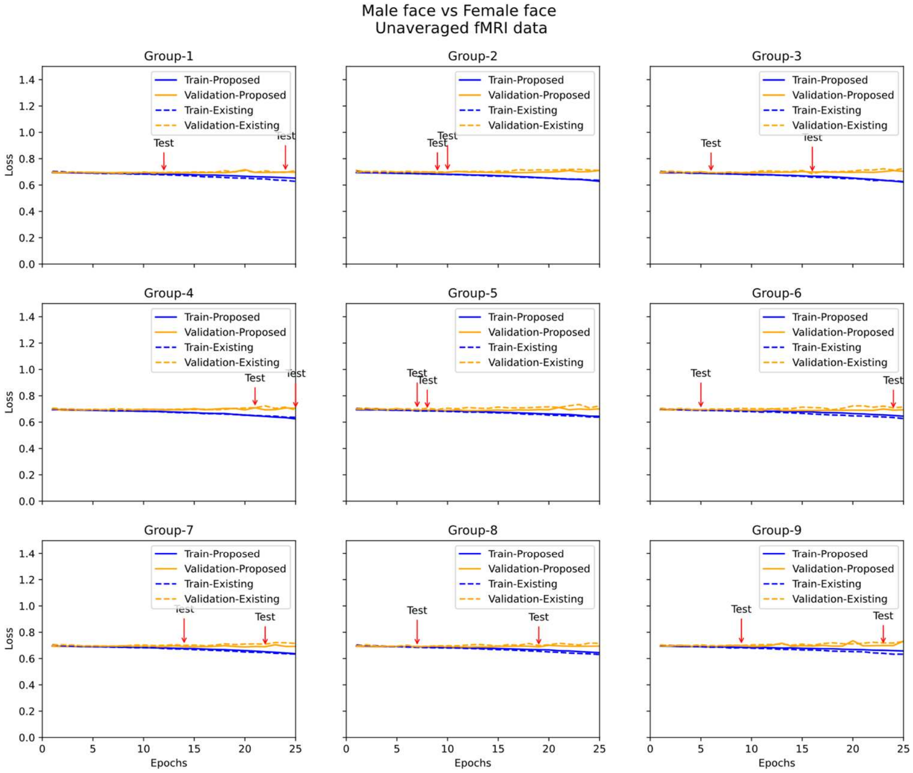


Figure 4.4-3 Training and validation losses within 25 iterations on Male face vs. Female face (unaveraged fMRI data).

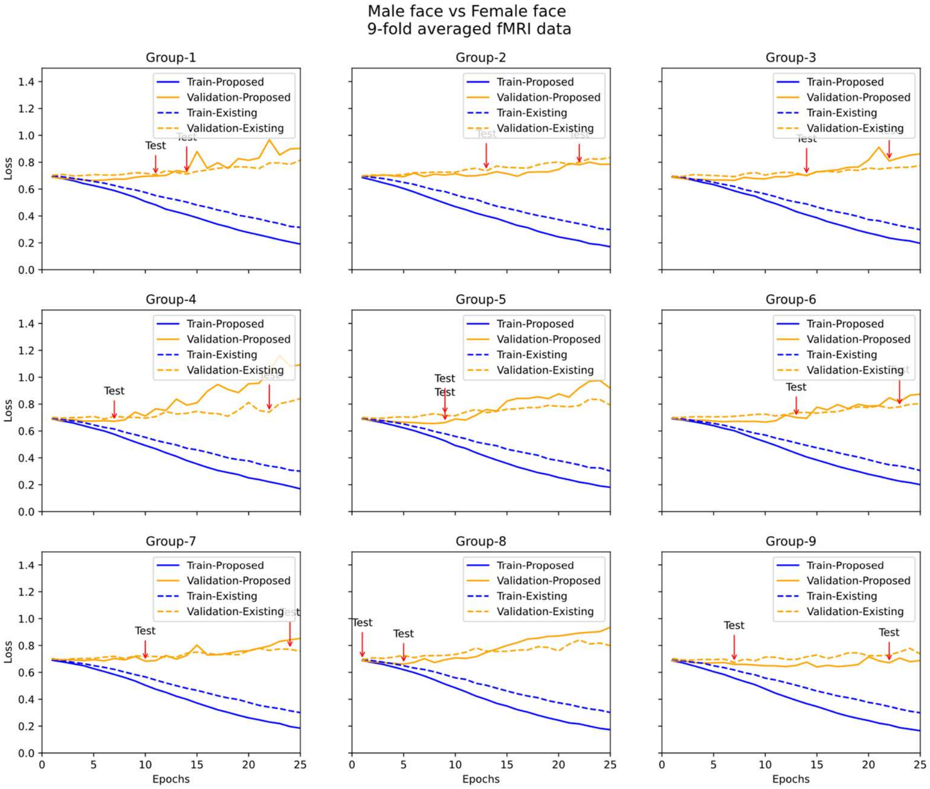


Figure 4.4-4 Training and validation losses within 25 iterations on Male face vs. Female face (9-fold averaged fMRI data).

In Figure 4.4-3, we can see that the proposed and existing model performs very close in the training and validation losses, which shows that the proposed model does not present a better learning ability than the existing one for high-noisy fMRI data in the face sub-categorical classification. As for the test accuracy, the proposed model reaches over 50% in eight training groups, while the existing one reaches it in only three groups. Finally, the voting result of the proposed model is 52.72%, and the existing one is 49.352% which does not exceed 50%. It reveals that the proposed model is more robust than the existing one for the face sub-categorical classification using high-noisy fMRI data.

When the 9-fold averaged fMRI data are used, the proposed model improves the validation accuracy from 4.492% to 8.583% and always produces a smaller training loss than the existing one within 25 iterations of training. As for the test accuracy, the proposed model reaches over 50% in eight training groups, while the existing one reaches it in only five groups. The voting result of the proposed model is 55.311%, which improves 1.684% compared to the existing one. And the test accuracy is improved from 0.777% to 4.793% in the 1st, 4th, 5th, 6th, 7th, and 9th groups while decreases from 0.518% to 6.088% in the 2nd, 3rd, and 8th groups. However, we can see that the proposed and existing models both have a pronounced upward trend in the validation loss, as shown in Figure 4.4-4. The reason is that the training loss continuously approaches zero in the training process, making the model overconfidence [132]. For example, the ideal model output is 0 or 1, and the model output would gradually approach either of them through training, then the wrong predictions increase the calculation loss. In an extreme case, the wrong predictions would affect the test results. Compared to the existing model, the proposed model produced results closer to 0 or 1, so its upward trend was more apparent when the proposed and existing models were overconfident.

Overall, the proposed model has slightly improved over the existing model in face sub-categorical classification. After voting, the classification accuracies are 3.368% and 1.684% over the existing one when the unaveraged and 9-fold averaged fMRI data, respectively. Notably, it always enables more groups to reach a classification accuracy above 50% than the existing model, which presents the robustness of the proposed model.

4.4.3 Natural object vs. Artificial object

Table 4.4-5 and Table 4.4-6 are the classification results of the unaveraged and 9-fold averaged fMRI data on the object sub-categorical (natural object vs. artificial object) classification. When the unaveraged fMRI data are used, the proposed and existing models exhibit a similar learning performance, as shown in Figure 4.4-5. We can see that the proposed model improves the validation accuracy by 0.349-2.43% over the existing one for the 1st, 2nd, 6th, 8th, and 9th groups, while the validation accuracy is decreased by 1.72-3.65% for the 3rd, 4th, 5th, and 7th groups. For test accuracy, the proposed model all reaches above 50% in the test accuracy except the 8th group, while the existing one reaches it only for the 2nd, 5th, 7th, and 8th groups. Finally, the voting accuracy of the proposed model is 51.532%. However, the existing model still does not exceed 50% by voting. When 9-fold averaged fMRI data are used, the proposed model significantly improves the classification accuracy over the existing one. In Table 4.4-6, the proposed model improves the validation accuracy by 0.973-10.931% and the test accuracy by 4.178-16.853% for all groups. The voting result of the proposed model is 61.978%, which is 13.649% higher than the existing one. Figure 4.4-6 is the training and test processes of 9-fold averaged fMRI data, where we can see that the proposed model always presents a smaller training loss but is trending upwards in the validation loss for the 4th, 7th, and 8th groups. The possible reason is because the model is overconfident, as explained in the face sub-categorical classification.

Table 4.4-5 Classification performance comparison on Natural object vs. Artificial object (unaveraged fMRI data).

Group	Validation accuracy (Proposed / Existing)		Test accuracy (Proposed / Existing)	
1	**54.273%**	52.636%	**50.418%**	47.772%
2	**54.861%**	52.431%	**55.014%**	50.139%
3	53.906%	**56.25%**	**52.089%**	48.607%
4	52.581%	**54.301%**	**53.621%**	49.443%
5	51.338%	**54.988%**	**51.671%**	50.139%
6	**54.419%**	54.07%	**52.228%**	47.493%
7	53.081%	**55.332%**	**53.482%**	50.836%
8	**54.0%**	53.385%	49.861%	**51.253%**
9	**54.111%**	53.448%	**52.646%**	49.721%
Voting	\		**51.532%**	48.886%

***Bold:** Better score in comparison between the proposed and existing model

Table 4.4-6 Classification performance comparison on Natural object vs. Artificial object (9-fold averaged fMRI data).

Group	Validation accuracy (Proposed / Existing)		Test accuracy (Proposed / Existing)	
1	**58.273%**	51.545%	**59.053%**	53.064%
2	**60.648%**	53.009%	**56.685%**	49.582%
3	**59.598%**	54.911%	**59.749%**	47.354%
4	**54.194%**	51.935%	**61.699%**	45.125%
5	**53.163%**	52.19%	**60.167%**	55.989%
6	**66.512%**	55.581%	**61.003%**	49.025%
7	**54.858%**	52.962%	**59.889%**	51.253%
8	**54.0%**	52.923%	**62.256%**	48.747%
9	**57.56%**	52.52%	**66.017%**	49.164%
Voting	\		**61.978%**	48.329%

***Bold:** Better score in comparison between the proposed and existing model

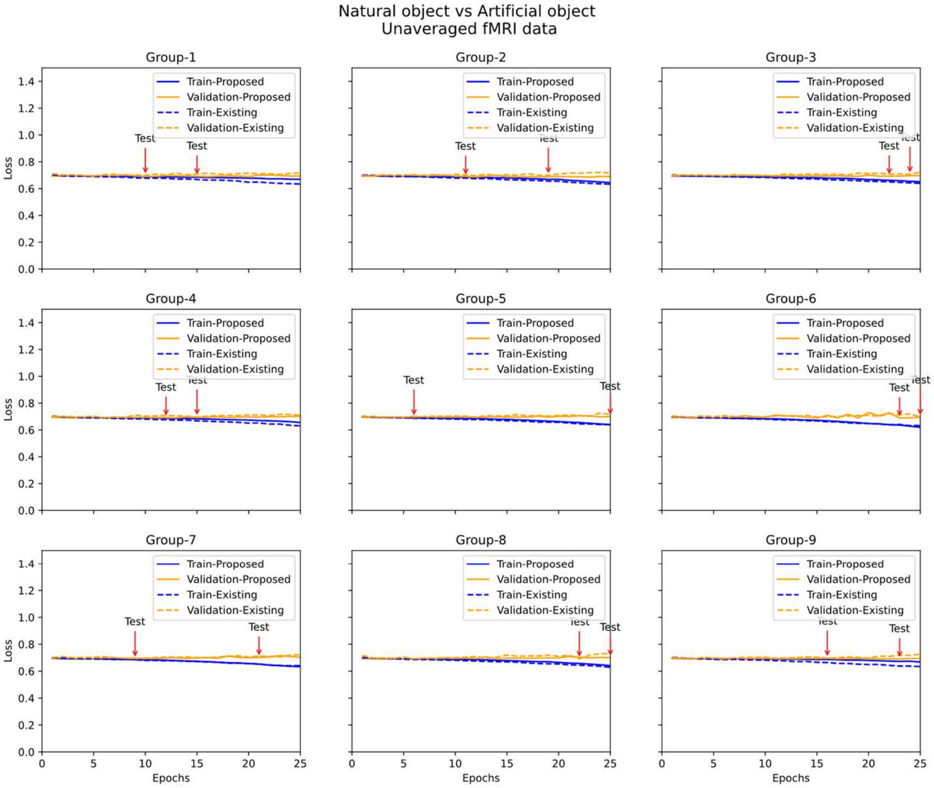


Figure 4.4-5 Training and validation losses within 25 iterations on Natural object vs. Artificial object (unaveraged fMRI data).

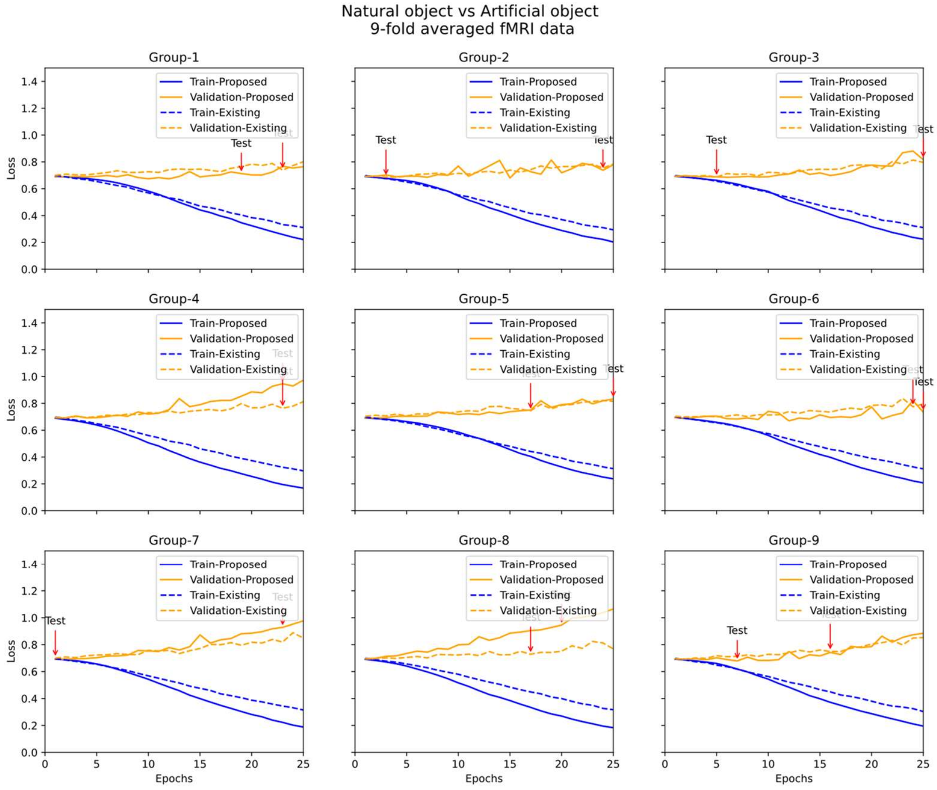


Figure 4.4-6 Training and validation losses within 25 iterations on Natural object vs. Artificial object (9-fold averaged fMRI data).

4.5 Discussion and Conclusion

This study proposed a novel fMRI classification model called MP3DCNN. Its mainchain is constructed with three-layer 3D convolutions, and the first and the second convolutions each have a branch connection. There are multiple 3D average pooling layers used in the mainchain and branches.

Compared to the existing model in the categorical classification (face vs. object), face sub-categorical classification (male face vs. female face), and object sub-categorical classification (natural object vs. artificial object), the proposed model improved the voting accuracy by 14.918%, 3.368%, and 2.646% when using the unaveraged fMRI data, and by 4.088%, 1.684% and 13.649 % when using the 9-fold averaged fMRI data, respectively.

To investigate the role of the model component in improving the classification performance, we further compared the existing model, the proposed model that contains only the mainchain, and the proposed model, as shown in Figure 4.5-1.

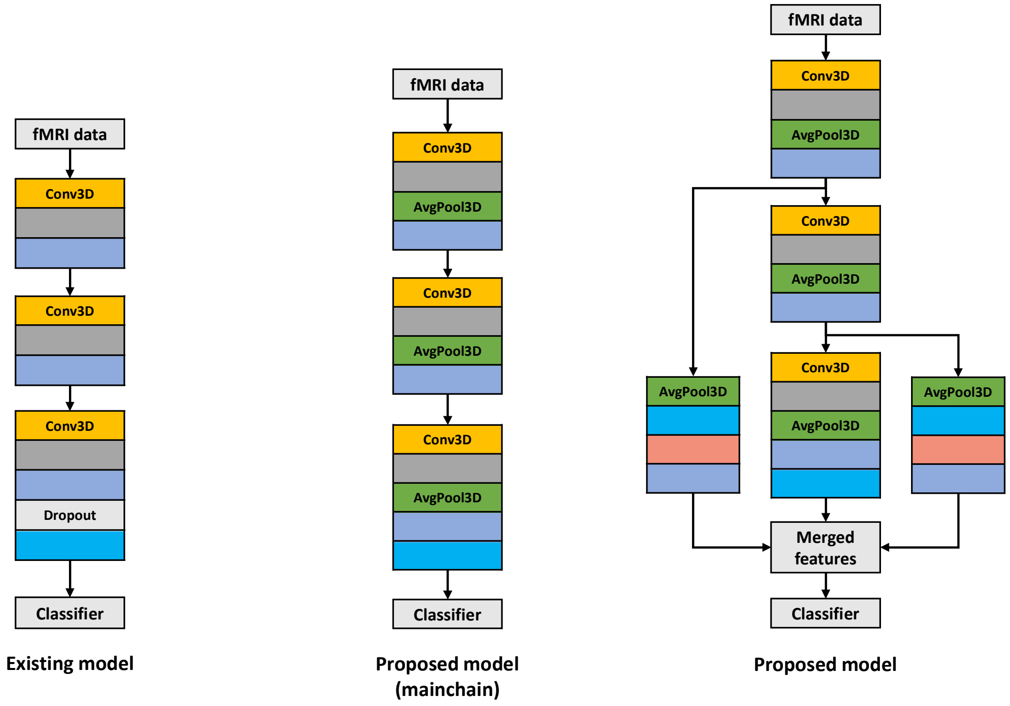


Figure 4.5-1 Existing model vs. Proposed model (mainchain) vs. Proposed model.

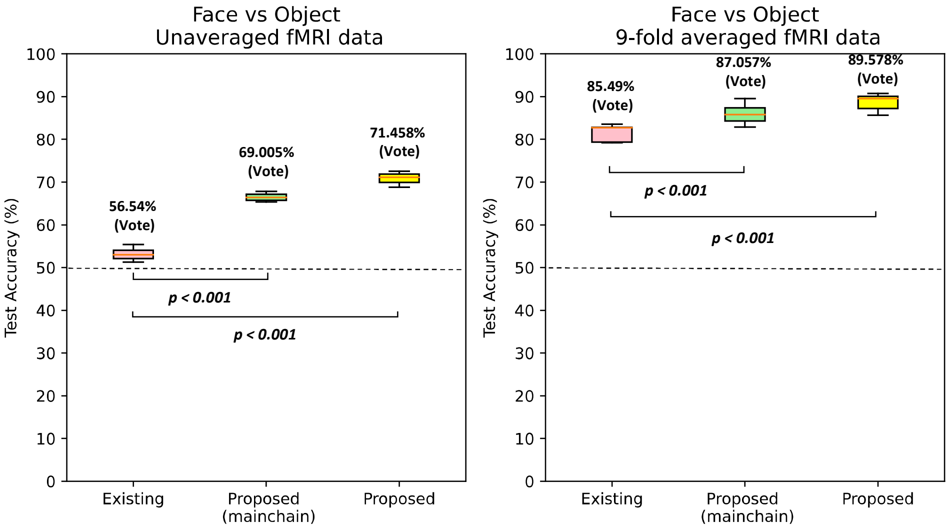


Figure 4.5-2 Performance comparison among the existing model, the proposed model (mainchain), and the proposed model in the categorical classification (face vs. object).

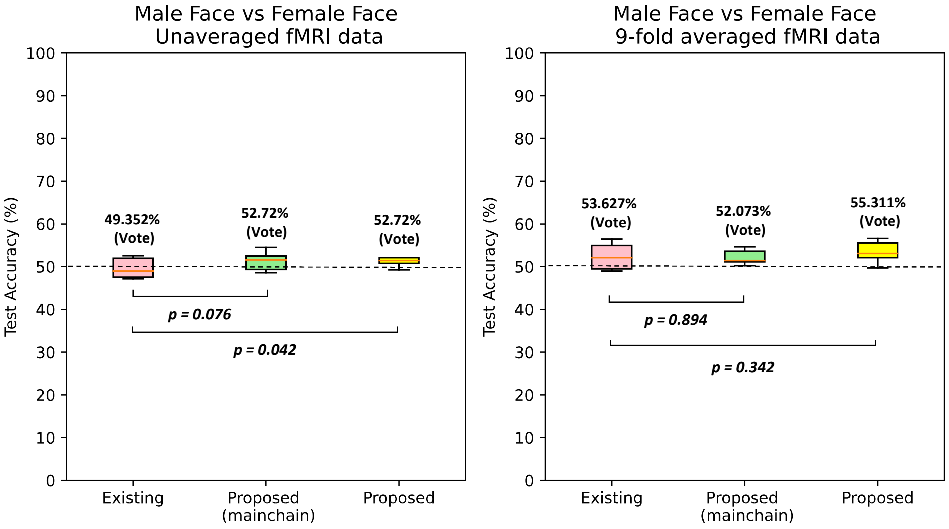


Figure 4.5-3 Performance comparison among the existing model, the proposed model (mainchain), and the proposed model in the face sub-categorical classification (male face vs. female face).

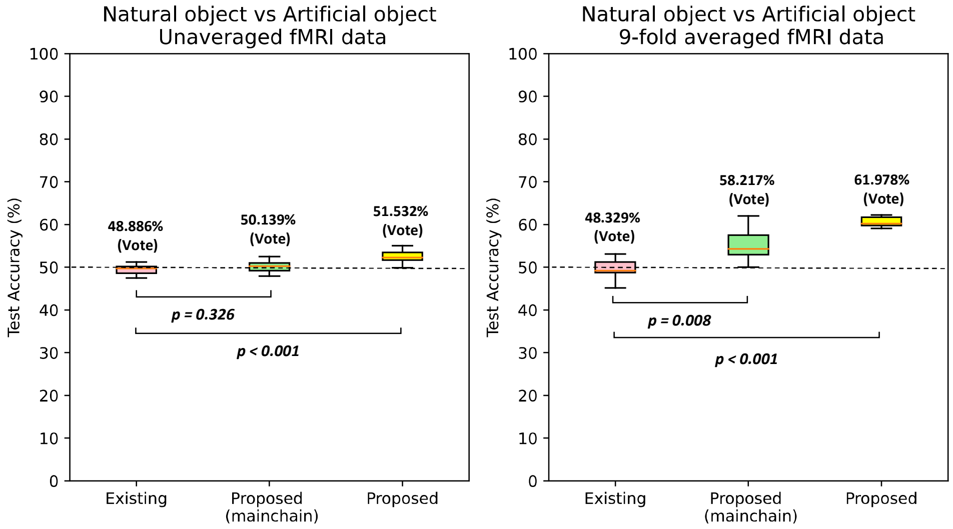


Figure 4.5-4 Performance comparison among the existing model, the proposed model (mainchain), and the proposed model in the object sub-categorical classification (natural object vs. artificial object).

The following are the accuracy distributions of nine test groups in the categorical, face sub-categorical, and object sub-categorical classifications. In Figure 4.5-2, the mainchain of the proposed model improved the voting accuracy by 12.465% when using the unaveraged data, and by 1.567% when using the 9-fold averaged data over the existing model. In the categorical classification, we can see that the model only with the mainchain still improved the classification performance significantly when using the high-noisy data.

In Figure 4.5-3, according to the ANOVA results, the performance differences among the three models are insignificant (p-values > 0.001) for the face sub-categorical classification. The mainchain and the complete body of the proposed model have the same improvement in the voting accuracy for the unaveraged data. However, the mainchain of the proposed model had an accuracy decay of 1.554% for the 9-fold averaged data, but its complete body still improved by 1.684% in the voting accuracy. It suggests that the branch connection can improve the model's robustness, especially when the mainchain of the model is not significantly superior in the face sub-categorical classification.

In Figure 4.5-4, the mainchain of the proposed model improved the voting accuracy by 1.253% for the unaveraged data and by 9.888% for the 9-fold averaged data, however, the performance difference between it and the existing model is insignificant (p-values > 0.001). In contrast, the complete body of the proposed model can reach higher voting accuracy and always has statistical significance to the existing model, which proves that the branch connections are critical for the object sub-categorical classification.

Based on the above, we speculated that the proposed model reached higher performance than the existing model for two reasons: first, brain activity in the categorical classification is relatively strong, which leads to a more pronounced distinction between the global features of the different categories. In the mainchain of the proposed model, the 3D average pooling layer can against the local feature redundancy in the feature extraction process and pass the global information to the classifier as much as possible. Therefore, the 3D pooling within the model's mainchain is the main reason for the improved classification accuracy. Second, in the face and object sub-categorical classifications, the mainchain of the proposed model presented a higher classification error rate because the brain activity became relatively slight. Through the branch connections, the model decision can depend on the merged features of the three 3D convolutions, thereby improving the model's robustness. Moreover, the branch connections provided extra learning paths for the first two convolutions, thereby improving the learning efficiency in the backpropagation. The results demonstrate that the branch connections can be used as an effective supplement for enhancing the model's robustness and accuracy.

5. Conclusion

5.1 Contribution & Summary

This thesis reported the study on deep learning algorithms for EEG and fMRI signal processing, including 1) STNN (spatial-temporal neural network) for P300 detection; 2) MMNN (multi-module neural network) for EEG denoising; and 3) MP3DCNN (multi-pooling 3D Convolutional Neural Networks) for fMRI Classification. Chapter 1 brief provided a brief overview of EEG, fMRI, and deep learning algorithms, involving the key concepts and EEG and fMRI based deep learning. Chapter 2 described the proposed STNN for P300 detection network, which is a parallel architecture consisting of a temporal unit and a spatial unit that can perform EEG channel generalization and analyze the brain's potential changes simultaneously. The results using three public datasets reveal that the proposed model performed better with fewer rounds of stimuli than other competitors. It is robust with limited data and is suitable for decoding EEG data recorded with various electrodes. Chapter 3 described the proposed MMNN for EEG denoising, which is assembled with multiple denoising modules. The results revealed that the proposed model can automatically remove OAs and MAs from single-channel noisy EEG signals. Compared to the existing models, it achieved higher signal reconstruction accuracy and reached this goal with less training data, which is expected to have applications in real tasks. Chapter 4 described the proposed MP3DCNN for fMRI classification. This model's mainchain is a three-layer 3DCNN, where the first and the second layers of convolution each have a branch connection. In the mainchain and branches, the extracted fMRI featured are multiple pooled. The results show that the proposed model reached higher classification accuracies for the categorical, face sub-categorical, and object sub-categorical classifications. From the study in this thesis, the model structure, the amount and quality of data, data diversity, and the training process have proven to be critical for EEG and fMRI signal processing using deep learning algorithms. To design a high-performance model for EEG and fMRI signal processing in the future, some strategies need to be considered as follows:

1) The specific object in tasks should be clearly defined first. Specifically, the output variable, the model's prediction, and the evaluation metrics should be set up carefully. For example, the EEG denoising task is defined as a regression problem, the model's output variable is continuous or numerical, and the model predicts values within a constant range. The evaluation metric calculates the mean squared error (MSE) between the labels and the model's outputs. On the other hand, the P300 detection and fMRI classification tasks are defined as the classification problems, where the model assigns input instances to predefined classes, and the evaluation metrics include accuracy, precision, recall, or F1 score.

2) The model’s components should be customized based on the data type of EEG or fMRI signals, aiming to handle the specific signal characteristics. EEG data are sequential or time series data, then the model’s components should be able to capture long-term dependencies and temporal patterns. Therefore, Conv1Ds are the base units in P300 detection and EEG classification tasks. However, the fMRI classification task is required to obtain the spatial information from different brain regions in three dimensions, therefore Conv3Ds are commonly used to capture local and global patterns of brain activities. It is worth noting that the experiences about the component choice are based on our specific task, available data, and computational resources. In the different tasks, the experimentation, iteration, and careful evaluation are often necessary to refine and optimize the architecture and techniques for optimal performance.

3) Deep learning models usually benefit from high-quality and diverse EEG and fMRI datasets. More data from different subjects can provide a broader representation of the underlying distribution, enabling the model to generalize better to unseen instances. Furthermore, the high-quality and diverse dataset can reduce the risk of overfitting and contribute to more accurate and robust models.

4) The capacity of deep learning models should be increased to improve the model’s performance. The implementation approaches include adding more parameters or layers to a model. The larger capacity models are better equipped to handle large and diverse datasets, capturing a wider range of features and variations present in the large dataset. However, the larger capacity models will bring the model overfitting or gradient diffusion problem. Regularization can help prevent overfitting and improve the generalization ability of deep learning models. Therefore, we used dropout, batch normalization, and early stopping to improve the model's generalization in the training process. Meanwhile, the gradient diffusion is a common challenge during the training of deep learning models. To deal with it, some activation functions such as ReLU and the skip connections, are widely used in the proposed models. Furthermore, the parallel connections are adopted in the proposed models, which can improve the gradient efficiency and reach more stable training, faster convergence, as well as higher performance across EEG and fMRI signal processing.

5) The fine-tuning of hyperparameters can explore the right balance between the model complexity and overfitting. This strategy includes the tuning of learning rate, batch size, regularization techniques, activation functions, optimizer algorithms, and network initialization. The systematic experiments should be used to find the optimal combination of hyperparameters through techniques like grid search or random search.

5.2 Prospects of this study

5.2.1 Prospects of the proposed STNN

When users intend to select specific targets or communicate through the BCI, the proposed high-performance P300 detection model can ensure that their intentions are correctly interpreted. Improved accuracy results in more precise identification of P300 responses, leading to enhanced user experience and faster information transfer rates. Moreover, the high-performance P300 detection model can reduce the occurrence of false positives or false negatives, thereby ensuring the reliable and robust operation of BCIs in real-world scenarios. In the field of clinical diagnostics, P300 responses have been extensively studied to understand cognitive processes, diagnose neurological disorders, and assess cognitive impairments. High-performance P300 detection approaches hold significant potential for improving the reliability and sensitivity of diagnostic tools. Accurate identification and quantification of P300 signals can aid in the early detection, monitoring, and treatment evaluation of conditions such as Alzheimer's disease, attention-deficit/hyperactivity disorder (ADHD), and other cognitive disorders. By achieving high accuracy, clinicians and researchers can extract valuable insights from P300 signals, enabling them to make more accurate diagnoses, personalize treatment plans, and track disease progression more effectively.

Furthermore, the underlying methodology and principles of P300 detection can be adapted and extended to various domains where the accurate detection of specific patterns within time series data is essential. For example, time series data from sensors, such as accelerometers or environmental sensors, often contain specific events or patterns of interest. The proposed P300 detection model can be adapted to detect these events. By training the model on labeled data with desired event patterns, the proposed model can identify and localize occurrences of these events in real-time sensor data, enabling applications in areas like activity recognition, anomaly detection, or environmental monitoring. For financial data, the proposed P300 detection model is expected to identify anomalous patterns that may indicate market manipulation, unusual trading activities, or fraudulent behavior. By training the model on historical financial data and defining the target anomalies, the proposed model is expected to accurately detect such events and assist in early warning systems or fraud detection mechanisms. For abnormality detection in health monitoring, the proposed model is expected to identify abnormal patterns in such data, enabling early detection of health issues or abnormalities. This model is expected to detect and alert healthcare providers or individuals to potential health risks by training the model on labeled health data and defining the target abnormalities.

5.2.2 Prospects of the proposed MMNN

The proposed EEG denoising model is expected to be applied to the field of neuroscience research and clinical diagnostics. High-performance EEG denoising approaches can help researchers obtain cleaner and more reliable data, allowing them to analyze and interpret brain activity with higher precision. Meanwhile, this technique can be used to enable more accurate identification and analysis of relevant biomarkers and abnormalities. Therefore, we look forward to applying this approach for individuals to control prosthetic limbs, interact with computers, or communicate through brain signals in the future.

Notably, the essence of the EEG denoising task is the regression study based on deep learning. Therefore, the proposed model structure will have the potential for other fields, such as financial forecasting, energy demand forecasting, environmental analysis, customer behavior analysis, recommender systems, quality control and manufacturing, and traffic flow prediction. Specifically, the proposed model is expected to predict financial metrics such as stock prices, market indices, exchange rates, and future sales. By training historical data, we expected the proposed model to capture complex patterns and relationships and provide financial planning, investment strategies, and risk management. Also, the data about weather patterns, historical energy consumption, and demographic information can be used to forecast energy demand for utilities and energy companies, which aims to predict future energy demand, facilitating efficient resource allocation, load balancing, and energy planning. In environmental analysis, the proposed model is expected to be trained by atmospheric data and pollutant levels, thereby providing predictions for ecological changes and guiding policy decisions. For customer behavior analysis, the training data include customer segmentation, churn prediction, and demand forecasting. By integrating various customer-related data, such as purchase history, demographics, and browsing patterns, our models have the potential to capture customer preferences, predict future behavior, and enable targeted marketing strategies. Also, this model can potentially be trained as a recommender system for personalized recommendations. In this case, the training data would come from user preferences, historical interactions, and contextual data. In quality control and manufacturing, the proposed model is expected to be used to enhance productivity and reduce defects through analyzing sensor data and historical performance. At the end, the proposed model as a regression approach based on deep learning, is expected to predict traffic flow patterns, congestion levels, and travel time estimation. By analyzing historical traffic data, weather conditions, and real-time information, deep regression models can forecast traffic patterns, enabling efficient route planning, traffic management, and urban planning.

5.2.3 Prospects of the proposed MP3DCNN.

In the future, the proposed fMRI classification model will have a wide range of applications, which are not limited to the applications in studying brain function and cognition. For example, fMRI classification can be used to understand consumers' preferences and responses to marketing stimuli. By analyzing brain activity patterns, researchers can identify neural correlates of positive or negative responses to advertisements, product designs, or brand experiences. This information can inform marketers and advertisers on how to optimize their strategies to elicit desired consumer reactions. Similarly, fMRI classification can help in understanding consumer decision-making processes and preferences. By examining brain activity, researchers can gain insights into the neural mechanisms underlying consumer choices and identify factors that influence purchasing behavior. This knowledge can be valuable for designing effective marketing campaigns, product development, and pricing strategies. In the education field, fMRI classification can help improve educational strategies by identifying neural markers associated with effective learning. By studying brain activity patterns, researchers can develop classifiers to assess cognitive engagement, attention, and comprehension during learning tasks. This information can inform the development of optimized teaching methods and personalized educational approaches. And fMRI classification also can be applied to understand the neural mechanisms associated with sports performance and expertise. By comparing brain activity patterns between athletes of different skill levels, classifiers can identify neural signatures related to superior performance. This information can be used to optimize training protocols, talent identification, and performance enhancement strategies.

5.3 Future challenges

Deep learning techniques are proven to offer promising opportunities for EEG and fMRI signal processing, but they also come with associated risks in the future. The key concerns include limited data availability, interpretability challenges, model complexity, generalization to new domains, data quality, and ethical considerations. Therefore, the further study can consider combining data augmentation and transfer learning with the proposed models. Furthermore, deep learning models trained on specific EEG or fMRI datasets may not generalize well to new datasets acquired from different populations or experimental conditions. Variability across individuals, equipment, and experimental setups can affect the performance of the trained models on unseen data. Therefore, we should be cautious about the generalizability of their models and validate their performance on diverse datasets. Meanwhile, we should seek for a balance between precision and complexity in deep learning, aiming to ensures that the deep learning model is not only accurate but also efficient, practical, and adaptable to various real-world scenarios.

In medical applications, the model's explanations are essential for clinical decision-making but the model structure is very complex, then a huge challenge is that how to ensure transparency and interpretability of deep learning models for EEG and fMRI signal processing. Nevertheless, we believe that deep learning algorithms are revolutionizing EEG and fMRI signal processing. Unlike traditional techniques that require manual feature engineering, deep learning eliminates the need for domain-specific knowledge and handcrafted feature extraction, making the analysis process more efficient and less subjective. EEG and fMRI signals are inherently complex, and traditional techniques often struggle to model their nonlinear dynamics. Deep learning models, with their hierarchical architectures and powerful computational capacity, can effectively model and capture the intricate dependencies and subtle patterns in brain signals, leading to improved accuracy and understanding. Therefore, we look forward to going deep into cross-study for EEG and fMRI signal processing based on deep learning in the future. For example, we expect to explore the technical feasibility of fMRI denoising based on deep learning after witnessing the application in EEG denoising. Compared with EEG denoising, fMRI denoising task encompasses not only temporal but also spatial artifacts, therefore fMRI denoising will pose more challenges in various factors, such as how to distinguish true neural signals from noise, choose optimal denoising techniques, avoid overfitting, and address subject variability.

References

[1] Biasiucci, B. Franceschiello, and M. M. Murray, "Electroencephalography," Curr. Biol., vol. 29, no. 3, pp. R80-R85, 2019.

[2] S. Siuly, Y. Li, and Y. Zhang, "Electroencephalogram (EEG) and its background," in EEG Signal Analysis and Classification: Techniques and Applications, 2016, pp. 3-21.

[3] S. Sur and V. K. Sinha, "Event-related potential: An overview," Ind. Psychiatry J., vol. 18, no. 1, p. 70, 2009.

[4] S. J. M. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy," J. Neurol. Neurosurg. Psychiatry, vol. 76, suppl. 2, pp. ii2-ii7, 2005.

[5] N. Srinivasan, "Cognitive neuroscience of creativity: EEG based approaches," Methods, vol. 42, no. 1, pp. 109-116, 2007.

[6] D. J. McFarland and J. R. Wolpaw, "EEG-based brain–computer interfaces," Curr. Opin. Biomed. Eng., vol. 4, pp. 194-200, 2017.

[7] N. K. Logothetis, "The neural basis of the blood–oxygen–level–dependent functional magnetic resonance imaging signal," Philos. Trans. R. Soc. Lond. B Biol. Sci., vol. 357, no. 1424, pp. 1003-1037, 2002.

[8] R. A. Poldrack, "The future of fMRI in cognitive neuroscience," Neuroimage, vol. 62, no. 2, pp. 1216-1220, 2012.

[9] P. M. Matthews, G. D. Honey, and E. T. Bullmore, "Applications of fMRI in translational medicine and clinical practice," Nat. Rev. Neurosci., vol. 7, no. 9, pp. 732-744, 2006.

[10] M. Welvaert and Y. Rosseel, "On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data," PLoS ONE, vol. 8, no. 11, p. e77089, 2013.

[11] I. Neuner, J. Arrubla, C. J. Werner, K. Hitz, F. Boers, W. Kawohl, and N. J. Shah, "The default mode network and EEG regional spectral power: a simultaneous fMRI-EEG study," PLoS ONE, vol. 9, no. 2, p. e88214, 2014.

[12] D. H. Murphree, P. Puri, H. Shamim, S. A. Bezalel, L. A. Drage, M. Wang, et al., "Deep learning for dermatologists: Part I. Fundamental concepts," J. Am. Acad. Dermatol., vol. 87, no. 6, pp. 1343-1351, 2022.

[13] I. H. Sarker, "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions," SN Comput. Sci., vol. 2, no. 6, p. 420, 2021.

[14] M. P. Hosseini, T. X. Tran, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, "Deep learning with edge computing for localization of epileptogenicity using multimodal rs-fMRI and EEG big data," in Proc. 2017 IEEE Int. Conf. Autonomic Comput. (ICAC), Jul. 2017, pp. 83-92.

[15] Q. S. Zhang and S. C. Zhu, "Visual interpretability for deep learning: a survey," Front. Inf. Technol. Electron. Eng., vol. 19, no. 1, pp. 27-39, Jan. 2018.

[16] C. Panigutti, A. Monreale, G. Comandé and D. Pedreschi, "Ethical, societal and legal issues in deep learning for healthcare," in Deep Learning in Biology and Medicine, pp. 265-313, 2022.

[17] Y. Liu, Y. Liu, J. Tang, E. Yin, D. Hu, and Z. Zhou, “A self-paced BCI prototype system based on the incorporation of an intelligent environment-understanding approach for rehabilitation hospital environmental control,” Comput. Biol. Med., vol. 118, pp. 103618, 2020.

[18] M. Balconi, and G. Fronda, “The Use of Hyperscanning to Investigate the Role of Social, Affective, and Informative Gestures in Non-Verbal Communication. Electrophysiological (EEG) and Inter-Brain Connectivity Evidence,” Brain Sci., vol. 10, no. 1, pp. 29, 2020.

[19] A. Mishra, S. Sharma, S. Kumar, P. Ranjan, and A. Ujlayan, “Effect of hand grip actions on object recognition process: a machine learning-based approach for improved motor rehabilitation,”, Neural. Computi. Appl., vol. 33, no. 7, pp. 2339-2350, 2020.

[20] X. Wang, G. Gong, N. Li, and Y. Ma, “A survey of the BCI and its application prospect,” in Proc. Asian Simulation Conference., Springer. Singapore, 2016, pp. 102-111.

[21] L. Farwell, and E. Donchin, “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials,” Electroencephalogr. Clin. Neurophysiol., vol. 70, no. 6, pp. 510-523, 1988.

[22] C. Başar-Eroglu, T. Demiralp, M. Schürmann, and E. Başar, “Topological distribution of oddball ‘P300’ responses,” Int. J. of Psychophysiol., vol. 39, no. 2-3, pp. 213-220, 2001.

[23] M. Abidi, G. Marco, A. Couillandre, M. Feron, E. Mseddi, N. Termoz, G. Querin, P. Pradat, and P. Bede, “Adaptive functional reorganization in amyotrophic lateral sclerosis: coexisting degenerative and compensatory changes,”, Eur. J. Neurol., vol. 27, no. 1, pp. 121-128, 2019.

[24] J. Cheng, L. Li, C. Li, Y. Liu, A. Liu, R. Qian, and X. Chen, “Remove Diverse Artifacts Simultaneously From a Single-Channel EEG Based on SSA and ICA: A Semi-Simulated Study,” IEEE Access., vol. 7, pp. 60276-60289, 2019.

[25] T. Wang, P. Liu, X. An, Y. Ke, J. Xu, M. Xu, L. Kong, W. Liu, and D. Ming, “Modeling Strategies and Spatial Filters for Improving the Performance of P300-speller within and across Individuals,” in Proc IEEE Int. Conf. Comput. Intell. Virtual Environ. Meas. Syst. Appl., 2019, pp. 1–5.

[26] M. Jalilpour Monesi, and S. Hajipour Sardouie, “Extended common spatial and temporal pattern (ECSTP): A semi-blind approach to extract features in ERP detection,” Pattern Recogn., vol. 95, pp. 128-135, 2019.

[27] P. Schembri, R. Anthony, and M. Pelc, “The feasibility and effectiveness of P300 responses using low fidelity equipment in three distinctive environments,” in Proc. 5th Int. Conf. Physiological Comput. Syst., 2018, pp 77–86.

[28] J. Jin, S. Li, I. Daly, Y. Miao, C. Liu, X. Wang, and A. Cichochi, “The study of generic model set for reducing calibration time in P300-based brain–computer interface,” IEEE Trans. Neural Syst. Rehabil. Eng., vol. 28, no. 1, pp. 3-12, Jan. 2020.

[29] S. Kundu, and S. Ari, “P300 based character recognition using convolutional neural network and support vector machine,” Biomed. Signal Proc. Control., vol. 55, pp. 101645, 2020.

[30] D. Krzemiński, S. Michelmann, M. Treder, and L. Santamaria, “Classification of P300 Component Using a Riemannian Ensemble Approach,” in Proc. 15th Medit. Conf. Med. Biol. Eng. Comput., 2019, pp 1885–89.

[31] F. Li, Y. Xia, F. Wang, D. Zhang, X. Li, and F. He, “Transfer Learning Algorithm of P300-EEG Signal Based on XDAWN Spatial Filter and Riemannian Geometry Classifier,” Appl. Sci., vol. 10, no. 5, pp. 1804, 2020.

[32] M. Simões, D. Borra, E. Santamaría-Vázquez, M. Bittencourt-Villalpando, D. Krzemiński, A. Miladinović, T. Schmid, H. Zhao, C. Amaral, B. Direito, J. Henriques, P. Carvalho, and M. Castelo-Branco, “BCIAUT-P300: A Multi-Session and Multi-Subject Benchmark Dataset on Autism for P300-Based Brain-Computer-Interfaces,” Front. Neurosci., vol. 14, 2020.

[33] X. Zhang, L. Yao, X. Wang, J. Monaghan, D. McAlpine, and Y. Zhang, “A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers,” J. Neural Eng., vol. 18, no. 3, pp. 031002, 2021.

[34] Z. Cao, “A review of artificial intelligence for EEG‐based brain–computer interfaces and applications,” Brain Sci. Adv., vol. 6, no. 3, pp. 162-170, 2020.

[35] H. Cecotti, and A. Graser, “Convolutional neural networks for P300 detection with application to brain-computer interfaces,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 3, pp. 433-445, Mar. 2011.

[36] S. Ghiselli, F. Gheller, P. Trevisi, E. Favaro, A. Martini, and M. Ermani, “Restoration of auditory network after cochlear implant in prelingual deafness: a P300 study using LORETA,” Acta Otorhinolaryngolo. Ital, vol. 40, no. 1, pp. 64-71, 2020.

[37] M. Awais, X. Long, B. Yin, S.F. Abbasi, S. Akbarzadeh, C. Lu, X. Wang, L. Wang, J. Zhang, J. Dudink, and W. Chen, “A hybrid DCNN-SVM model for classifying neonatal sleep and wake states based on facial expressions in video,” IEEE J. Biomed. Health Inform., vol. 5, no. 25, pp. 1441-1449, 2021.

[38] L. Wang, and W. Chen, ‘‘EEG-based neonatal sleep-wake classification using multilayer perceptron neural network,’’ IEEE Access, vol. 8, pp. 183025–183034, 2020.

[39] M. Liu, W. Wu, Z. Gu, Z. Yu, F. Qi, and Y. Li, “Deep learning based on Batch Normalization for P300 signal detection,” Neurocomput, vol. 275, pp. 288-297, 2018.

[40] D. Borra, S. Fantozzi, and E. Magosso, “Convolutional neural network for a P300 brain-computer interface to improve social attention in autistic spectrum disorder,” in Proc. 15th Medit. Conf. Med. Biol. Eng. Comput., vol. 76, 2019, pp. 206-212.

[41] Z. Lu, Q. Li, N. Gao, T. Wang, J. Yang, and O. Bai, “A Convolutional Neural Network based on Batch Normalization and Residual Block for P300 Signal Detection of P300-speller System,” in Proc. IEEE Int. Conf. Mechatronics Autom, 2019, pp. 2303-2308.

[42] O. Tal, and D. Friedman, “Recurrent neural networks for P300-based BCI,” arXiv:1901.10798, 2019, [online] Available: https://arxiv.org/abs/1901.10798.

[43] A. Ditthapron, N. Banluesombatkul, S. Ketrat, E. Chuangsuwanich, and T. Wilaiprasitporn, “Universal Joint Feature Extraction for P300 EEG Classification Using Multi-Task Autoencoder,” IEEE Access., vol. 7, pp. 68415-68428, 2019.

[44] R. Maddula, J. Stivers, M. Mousavi, S. Ravindranand, and V. de Sa, “Deep Recurrent Convolutional Neural Networks for Classifying P300 BCI signals,” in Proc. 7th Graz Brain-Comput. Interface, Conf., 2017, pp 18–22.

[45] V. Lawhern, A. Solon, N. Waytowich, S. Gordon, C. Hung, and B. Lance, “EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces,” J. Neural Eng., vol. 15, no. 5, pp. 056013, 2018.

[46] S. Bai, JZ. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” arXiv: 1803.01271, 2018, [online] Available: https://arxiv.org/abs/1803.01271.

[47] A. Riccio, L. Simione, F. Schettini, A. Pizzimenti, M. Inghilleri, M. Belardinelli, D. Mattia, and F. Cincotti, “Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis,” Front. Hum. Neurosci., vol. 7, 2013.

[48] F. Aloise, P. Aricò, F. Schettini, A. Riccio, S. Salinari, D. Mattia, F. Babiloni, and F. Cincotti, “A covert attention P300-based brain-computer interface: Geospell,” Ergonomics, vol. 55, no. 5, pp. 538-551, 2012.

[49] B. Blankertz, K. Muller, D. Krusienski, G. Schalk, J. Wolpaw, A. Schlogl, G. Pfurtscheller, J. Millan, M. Schroder, and N. Birbaumer, "The BCI competition III: validating alternative approaches to actual BCI problems," IEEE Trans. on Neural Syst. Rehabil. Eng., vol. 14, no. 2, pp. 153-159, 2006.

[50] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw, "BCI2000: A General-Purpose Brain-Computer Interface (BCI) System," IEEE Trans. Biomed. Eng, vol. 51, no. 6, pp. 1034-1043, 2004.

[51] G. Chatrian, E. Lettich, and P. Nelson, "Ten Percent Electrode System for Topographic Studies of Spontaneous and Evoked EEG Activities," Am. J. EEG Technol, vol. 25, no. 2, pp. 83-92, 1985.

[52] I. Selesnick, and C. Burrus, "Generalized digital Butterworth filter design," IEEE Trans. Signal Proces, vol. 46, no. 6, pp. 1688-1694, 1998.

[53] S. F. Abbasi, M. Awais, X. Zhao and W. Chen, "Automatic denoising and artifact removal from neonatal EEG," in Proc. 3rd Int. Conf. Biol. Inf. Biomed. Eng., 2019, pp. 1–5.

[54] A. Creswell, K. Arulkumaran, and AA. Bharath, "On denoising autoencoders trained to minimise binary cross-entropy," arXiv: 1708.08487, 2014, [online] Available: https://arxiv.org/abs/1708.08487.

[55] J. Lobo, A. Jiménez-Valverde, and R. Real, "AUC: a misleading measure of the performance of predictive distribution models," Glob. Ecol. Biogeogr., vol. 17, no. 2, pp. 145-151, 2008.

[56] M. L. McHugh, ''Interrater reliability: The kappa statistic,'' Biochem.

[57] Z. Oralhan, "3D Input Convolutional Neural Networks for P300 Signal Detection," IEEE Access., vol. 8, pp. 19521-19529, 2020.

[58] F. Miwakeichi, E. Martnez-Montes, P. Valds-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space–time–frequency components using parallel factor analysis," NeuroImage., vol. 22, no. 3, pp. 1035–1045, 2004.

[59] J. Zhang, A. Liu, M. Gao, X. Chen, X. Zhang, and X. Chen, ''ECG-based multi-class arrhythmia detection using spatio-temporal attention based convolutional recurrent neural network,'' Artif. Intell. Med., vol. 106, Jun. 2020, Art. no. 101856.

[60] J. L. Whitton, F. Lue, and H. Moldofsky, "A spectral method for removing eye movement artifacts from the EEG," Electroencephalogr. clin. neurophysiol., vol. 44, no. 6, pp. 735–741, 1978.

[61] R. D. O'Donnell, J. Berkhout, and W. R. Adey, "Contamination of scalp EEG spectrum during contraction of cranio-facial muscles," Electroencephalogr. clin. neurophysiol., vol. 37, no. 2, pp. 145–151, 1974.

[62] M. A. Shaffer, "Problem record of the month, No. 3: Asymmetrical eye- blink artifact," Am. J. Electroneurodiagnostic Technol., vol. 10, no. 4, pp. 153–156, 1970.

[63] O. P. Mathew, Y. K. Abu-Osba, and B. T. Thach, “Influence of upper airway pressure changes on genioglossus muscle respiratory activity,” J. Appl. Physiol., vol. 52, no. 2, pp. 438–444, 1982.

[64] R. J. Croft and R. J. Barry, “Removal of ocular artifact from the EEG: a review,” Neurophysiol Clin., vol. 30, no. 1, pp. 5–19, 2000.

[65] G. Gratton, M. G. Coles, and E. Donchin, “A new method for off-line removal of ocular artifact,” Electroencephalogr. clin. neurophysiol., vol. 55, no. 4, pp. 468–484, 1983.

[66] O. L. Frost, “An algorithm for linearly constrained adaptive array processing,” Proc. IEEE, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[67] P. He, G. Wilson, C. Russell, and M. Gerschutz, “Removal of ocular artifacts from the EEG: a comparison between time-domain regression method and adaptive filtering method using simulated data,” Med. Biol. Eng. Comput., vol. 45, no. 5, pp. 495– 503, 2007.

[68] S. Boudet, L. Peyrodie, G. Forzy, A. Pinti, H. Toumi, and P. Gallois, “Improvements of adaptive filtering by optimal projection to filter different artifact types on long duration EEG recordings,” Comput. Methods Programs Biomed., vol. 108, no. 1, pp. 234–249, 2012.

[69] H. Hallez, M. D. Vos, B. Vanrumste, P. V. Hese, S. Assecondi, K. V. Laere, P. Dupont, W. V. Paesschen, S. V. Huffel, and I. Lemahieu, “Removing muscle and eye artifacts using blind source separation techniques in ictal EEG source imaging,” Clin. Neurophysiol., vol. 120, no. 7, pp. 1262–1272, 2009.

[70] M. A. Klados, C. Papadelis, C. Braun, and P. D. Bamidis, “REG-ICA: a hybrid methodology combining blind source separation and regression techniques for the rejection of ocular artifacts,” Biomed. Signal Process. Control., vol. 6, no. 3, pp. 291–300, 2011.

[71] L. Shoker, S. Sanei, and M. A. Latif, “Removal of eye blinking artifacts from EEG incorporating a new constrained BSS algorithm,” in Proc. IEEE Sensor Array Multichannel Signal Process. Workshop, 2004, pp. 177–181.

[72] K. H. Ting, P. C. W. Fung, C. Q. Chang, and F. H. Y. Chan, “Automatic correction of artifact from single-trial event-related potentials by blind source separation using second order statistics only,” Med. Eng. Phys., vol. 28, no. 8, pp. 780–794, 2006.

[73] K. T. Sweeney, S. F. McLoone, and T. E. Ward, “The use of ensemble empirical mode decomposition with canonical correlation analysis as a novel artifact removal technique,” IEEE. Trans. Biomed., vol. 60, no. 1, pp. 97–105, 2012.

[74] J.-E. Liu and F.-P. An, ‘‘Image classification algorithm based on deep learning-kernel function,’’ Sci. Program., vol. 2020, pp. 1–14, Jan. 2020.

[75] M. Kim, G. S. Jeng, I. Pelivanov, and M. O'Donnell, "Deep-learning image reconstruction for real-time photoacoustic system," IEEE Trans. Med. Imaging., vol. 39, no. 11, pp. 3379–3390, 2020.

[76] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super resolution: A survey," CoRR, vol. abs/1902.06068, pp. 1–24, 2019.

[77] D. A. Wood, S. Kafiabadi, A. A. Busaidi, E. L. Guilhem, J. Lynch, M. K. Townend, A. Montvila, M. Kiik, J. Siddiqui, and N. Gadapa, "Deep learning to automate the labelling of head MRI datasets for computer vision applications," Eur. Radiol., vol. 32, no. 1, pp. 725–736, 2022.

[78] M. H. Alkinani, W. Z. Khan, and Q. Arshad, "Detecting human driver inattentive and aggressive driving behavior using deep learning: Recent advances, requirements and open challenges," IEEE Access., vol. 8, pp. 105 008–105 030, 2020.

[79] S. Oprea, P. Martinez-Gonzalez, A. Garcia-Garcia, J. A. Castro-Vargas, S. Orts-Escolano, J. Garcia-Rodriguez, and A. Argyros, "A review on deep learning techniques for video prediction," IEEE Trans. Pattern Anal. Mach. Intell., 2020.

[80] J. Guo, H. He, T. He, L. Lausen, M. Li, H. Lin, X. Shi, C. Wang, J. Xie, and S. Zha, "GluonCV and GluonNLP: Deep Learning in Computer Vision and Natural Language Processing." J. Mach. Learn. Res., vol. 21, no. 23, pp. 1–7, 2020.

[81] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," IEEE Comput. Intell. Mag., vol. 13, no. 3, pp. 55–75, 2018.

[82] D. W. Otter, J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 2, pp. 604–624, 2020.

[83] M. Gardner, J. Grus, M. Neumann, O. Tafjord, P. Dasigi, N. Liu, M. Peters, M. Schmitz, and L. Zettlemoyer, "Allennlp: A deep semantic natural language processing platform," 2018, arXiv:1803.07640.

[84] B. Yang, K. Duan, C. Fan, C. Hu, and J. Wang, "Automatic ocular artifacts removal in EEG using deep learning," Biomed Signal Process Control., vol. 43, pp. 148–158, 2018.

[85] W. Sun, Y. Su, X. Wu, and X. Wu, "A novel end-to-end 1D-ResCNN model to remove artifact from EEG signals," Neurocomputing, vol. 404, pp. 108– 121, 2020.

[86] C. Hanrahan, "Noise reduction in EEG signals using convolutional autoencoding techniques," 2019. Masters University of Dublin (https://arrow.tudublin.ie/scschcomdis/188/)

[87] H. Zhang, M. Zhao, C. Wei, D. Mantini, Z. Li, and Q. Liu, “Eegdenoisenet: A benchmark dataset for deep learning solutions of EEG denoising,” J. Neural Eng., vol. 18, no. 5, pp. 056057, 2021.

[88] H. Zhang, C. Wei, M. Zhao, Q. Liu, and H. Wu, ‘‘A novel convolutional neural network model to remove muscle artifacts from EEG,’’ in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Jun. 2021, pp. 1265–1269.

[89] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” nature., vol. 521, no. 7553, pp. 436–444, 2015.

[90] X. Li, H. Jiang, K. Zhao, and R. Wang, “A deep transfer nonnegativity-constraint sparse autoencoder for rolling bearing fault diagnosis with few labeled data,” IEEE Access., vol. 7, pp. 91216-91224, 2019.

[91] W. Samek, T. Wiegand, and K.-R. Müller, “Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models,” 2017 arXiv:1708.08296.

[92] H. Cho, M. Ahn, S. Ahn, M. Kwon, and S. C. Jun, ‘‘EEG datasets for motor imagery brain computer interface,’’ Giga Sci., vol. 6, no. 7, pp. 1–8, 2017.

[93] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, “ICLabel: An auto- mated electroencephalographic independent component classifier, dataset, and website,” NeuroImage., vol. 198, pp. 181–197, 2019.

[94] S. Kanoga, M. Nakanishi, and Y. Mitsukura, “Assessing the effects of voluntary and involuntary eyeblinks in independent components of electroencephalogram,” Neurocomputing., vol. 193, pp. 20–32, 2016.

[95] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, “EMG and EOG artifacts in brain computer interface systems: A survey,” Clin Neurophysiol., vol. 118, no. 3, pp. 480–494, 2007.

[96] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller, “Separability of four-class motor imagery data using independent components analysis,” J. Neural Eng., vol. 3, no. 3, p. 208, 2006.

[97] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, “A fully automated correction method of EOG artifacts in EEG recordings,” Clin. Neurophysiol., vol. 118, no. 1, pp. 98–104, 2007.

[98] C. Brunner, M. Naeem, R. Leeb, B. Graimann, and G. Pfurtscheller, “Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis,” Pattern Recognit. Lett., vol. 28, no. 8, pp. 957–964, 2007.

[99] V. Rantanen, M. Ilves, A. Vehkaoja, “A survey on the feasibility of surface EMG in facial pacing,” 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, 2016, pp. 1688–1691.

[100] Y. Yasui, “A brainwave signal measurement and data processing technique for daily life applications,” J. Phys. Anthr., vol. 28, no. 3, pp. 145–150, 2009.

[101] J. T. Jacobson and C. A. Jacobson, “The effects of noise in transient EOAE newborn hearing screening,” Int. J. Pediatr. Otorhinolaryngol., vol. 29, no. 3, pp. 235–248, 1994.

[102] D. Safieddine, A. Kachenoura, L. Albera, G. Birot, A. Karfoul, A. Pasnicu, A. Biraben, F. Wendling, L. Senhadji, and I. Merlet, “Removal of muscle artifact from EEG data: comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches,” EURASIP J Adv Signal Process., vol. 2012, no. 1, pp. 1–15, 2012.

[103] G. Gomez-Herrero, W. De Clercq, H. Anwar, O. Kara, K. Egiazarian, S. Van Huffel, and W. Van Paesschen, “Automatic removal of ocular artifacts in the EEG without an EOG reference channel,” in Proc. 7th Nordic Signal Process. Symp., Jun. 2006, pp. 130–133.

[104] P. J. Allen, G. Polizzi, K. Krakow, D. R. Fish, and L. Lemieux, “Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction,” Neuroimage., vol. 8, no. 3, pp. 229–239, 1998.

[105] S. Hochreiter, “The vanishing gradient problem during learning recurrent neural nets and problem solutions,” Int. J. Uncertain. Fuzziness Knowlege-Based Syst., vol. 6, no. 02, pp. 107–116, 1998.

[106] M. Tuchler, A. C. Singer, and R. Koetter, “Minimum mean squared error equalization using a priori information,” IEEE Trans. Signal Process., vol. 50, no. 3, pp. 673–683, 2002.

[107] G. Wang, C. Teng, K. Li, Z. Zhang, and X. Yan, “The removal of EOG artifacts from EEG signals using independent component analysis and multivariate empirical mode decomposition,” IEEE J Biomed Health Inform., vol. 20, no. 5, pp. 1301–1308, 2015.

[108] X. Chen, H. Peng, F. Yu, and K. Wang, “Independent vector analysis applied to remove muscle artifacts in EEG data,” IEEE Trans Instrum Meas., vol. 66, no. 7, pp. 1770–1779, 2017.

[109] W. D. Clercq, A. Vergult, B. Vanrumste, W. V. Paesschen, and S. V. Huffel, “Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram,” IEEE. Trans. Biomed., vol. 53, no. 12, pp. 2583–2587, 2006.

[110] J. Dammers, M. Schiek, F. Boers, C. Silex, M. Zvyagintsev, U. Pietrzyk, and K. Mathiak, “Integration of amplitude and phase statistics for complete artifact removal in independent components of neuromagnetic recordings,” IEEE. Trans. Biomed., vol. 55, no. 10, pp. 2353– 2362, 2008.

[111] M. A. Uusitalo and R. J. Ilmoniemi, “Signal-space projection method for separating MEG or EEG into components,” Med Biol Eng Comput., vol. 35, no. 2, pp. 135–140, 1997.

[112] A. Gramfort et al., ‘‘MNE software for processing MEG and EEG data,’’ Neuroimage, vol. 86, pp. 446–460, Feb. 2014

[113] A. Craik, Y. He, and J. L. Contreras-Vidal, “Deep learning for electroencephalogram (EEG) classification tasks: a review,” J. Neural Eng., vol. 16, no. 3, p. 031001, 2019.

[114] D. J. Heeger and D. Ress, "What does fMRI tell us about neuronal activity?," Nat. Rev. Neurosci., vol. 3, no. 2, pp. 142-151, 2002.

[115] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fMRI signal," Nature, vol. 412, no. 6843, pp. 150-157, 2001.

[116] D. D. Cox and R. L. Savoy, "Functional magnetic resonance imaging (fMRI) 'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex," Neuroimage, vol. 19, no. 2, pp. 261-270, 2003.

[117] T. Naselaris, K. N. Kay, S. Nishimoto, and J. L. Gallant, "Encoding and decoding in fMRI," Neuroimage, vol. 56, no. 2, pp. 400-410, 2011.

[118] M. Chen, J. Han, X. Hu, X. Jiang, L. Guo, and T. Liu, "Survey of encoding and decoding of visual stimulus via FMRI: an image analysis perspective," Brain Imaging Behav., vol. 8, no. 1, pp. 7-23, 2014.

[119] C. Cabral, M. Silveira, and P. Figueiredo, "Decoding visual brain states from fMRI using an ensemble of classifiers," Pattern Recogn., vol. 45, no. 6, pp. 2064-2074, 2012.

[120] L. I. Kuncheva, J. J. Rodríguez, C. O. Plumpton, D. E. Linden, and S. J. Johnston, "Random subspace ensembles for fMRI classification," IEEE Trans. Med. Imaging, vol. 29, no. 2, pp. 531-542, 2010.

[121] J. A. Etzel, V. Gazzola, and C. Keysers, "An introduction to anatomical ROI-based fMRI classification analysis," Brain Res., vol. 1282, pp. 114-125, 2009.

[122] S. M. LaConte, S. J. Peltier, and X. P. Hu, "Real-time fMRI using brain-state classification," Hum. Brain Mapp., vol. 28, no. 10, pp. 1033-1044, 2007.

[123] J. B. Poline and M. Brett, "The general linear model and fMRI: does love last forever?," Neuroimage, vol. 62, no. 2, pp. 871-880, 2012.

[124] A. Mahmoudi, S. Takerkart, F. Regragui, D. Boussaoud, and A. Brovelli, "Multivoxel pattern analysis for FMRI data: a review," Comput. Math. Methods Med., vol. 2012, 2012.

[125] G. Shen, K. Dwivedi, K. Majima, T. Horikawa, and Y. Kamitani, "End-to-end deep image reconstruction from human brain activity," Front. Comput. Neurosci., vol. 21, 2019.

[126] R. Geirhos, C. R. Temme, J. Rauber, H. H. Schütt, M. Bethge, and F. A. Wichmann, "Generalisation in humans and deep neural networks," Adv. Neural Inf. Process. Syst., vol. 31, 2018.

[127] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," Adv. Neural Inf. Process. Syst., vol. 19, 2006.

[128] N. Watanabe, K. Miyoshi, K. Jimura, D. Shimane, R. Keerativittayayut, K. Nakahara, and M. Takeda, "Multi-modal deep neural decoding of visual object representation in humans," bioRxiv, 2022.

[129] J. Ashburner, G. Barnes, C. C. Chen, J. Daunizeau, G. Flandin, K. Friston, and W. Penny, "SPM12 manual," Wellcome Trust Centre for Neuroimaging, London, UK, 2464, 4, 2014.

[130] T. T. Liu, "Noise contributions to the fMRI signal: An overview," NeuroImage, vol. 143, pp. 141-151, 2016.

[131] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in Int. Workshop Artif. Neural Netw., pp. 195-201, Springer, Berlin, Heidelberg, 1995.

[132] J. W. Ryu, M. Kantardzic, and C. Walgampaya, "Ensemble classifier based on misclassified streaming data," in Proc. of the 10th IASTED Int. Conf. Artif. Intell. Appl., Austria, pp. 347-354, 2010.

[133] T. Ishida, I. Yamane, T. Sakai, G. Niu, and M. Sugiyama, "Do we need zero training loss after achieving zero training error?," arXiv preprint arXiv:2002.08709, 2020.

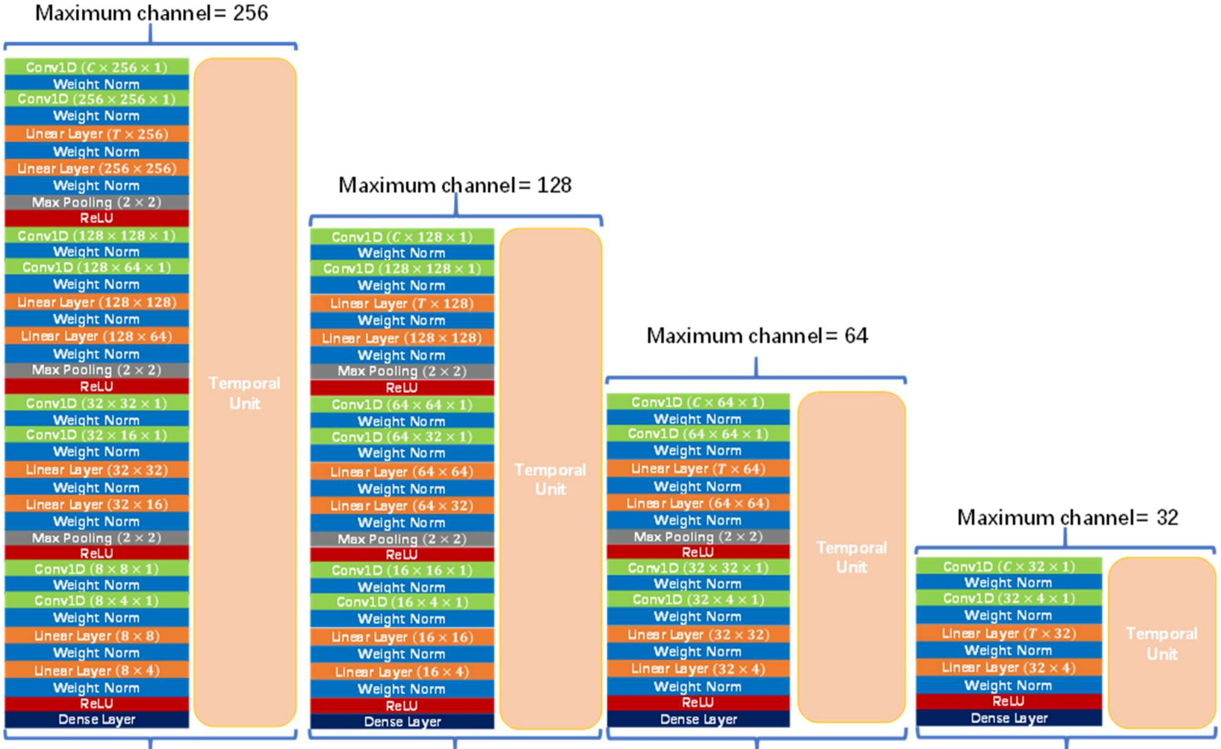
Appendix

A.1 Hyperparameter tuning & Running time – STNN.

A. 1 lists the hyperparameter tuning process of the spatial unit, where the average accuracy of STNN-3&15, 4&7, 4&8, 5&3, and 5&4 are given. We utilized the dataset-3 for training, validating, and testing models, where we performed 5-fold cross-validation on the training dataset (85 trials), and the testing dataset (100 trials) was given in Section II. V-A, V-B, T-A, and T-B are short for the validation result of subject-A, validation result of subject-B, testing result of subject-A, and testing result of subject- B. A. 2 gives the average training loss and validation loss of subjects A and B. We can see that the model performance can be improved by increasing the number of hyperparameters in the spatial unit, while the excessive increase in the hyperparameters does not significantly improve its performance. Therefore, the spatial unit with maximum channel = 128 was adopted in our P300 detection study.

A. 3 gives the hyperparameter tuning process of the global generalizers in the temporal module using the dataset-3. A. 4 shows the average training loss and validation loss of subjects A and B. We separately assembled the global generalizers with different maximum channels into STNN-3&15, 4&7, 4&8, 5&3, and 5&4. According to the average 5-fold cross-validation and testing AUC scores of these five models, we can see that the model performance can be improved using the global generalizers in the temporal modules. However, a huge amount of training parameters led to computational redundancy but obviously did not improve the model's performance. Therefore, the output channel of the temporal feature generalizer was set up to 128 in our P300 detection study.

A. 5 shows the processing time of the within-subject P300 detection and cross-subject P300 detection in the first experiment, where we list the average training and testing times of 1-10 rounds of stimuli. In the second experiment, we calculated the average training and testing times of 1-8 rounds of stimuli in the within-subject P300 detection and cross-subject P300 detection. The results in Farwell and Donchin's paradigm and the GeoSpell paradigm are given in A. 6 and A. 7, respectively. A. 8 gives the average training and testing times of two subjects using our model components and combinations in the third experiment.


Maximum channel= 256
Conv1D ($C \times 256 \times 1$)
Weight Norm
Conv1D ($256 \times 256 \times 1$)
Weight Norm
Linear Layer ($T \times 256$)
Weight Norm
Linear Layer (256×256)
Weight Norm
Max Pooling (2×2)
ReLU
Conv1D ($128 \times 128 \times 1$)
Weight Norm
Conv1D ($128 \times 64 \times 1$)
Weight Norm
Linear Layer (128×128)
Weight Norm
Linear Layer (128×64)
Weight Norm
Max Pooling (2×2)
ReLU
Conv1D ($32 \times 32 \times 1$)
Weight Norm
Conv1D ($32 \times 16 \times 1$)
Weight Norm
Linear Layer (32×32)
Weight Norm
Linear Layer (32×16)
Weight Norm
Max Pooling (2×2)
ReLU
Conv1D ($8 \times 8 \times 1$)
Weight Norm
Conv1D ($8 \times 4 \times 1$)
Weight Norm
Linear Layer (8×8)
Weight Norm
Linear Layer (8×4)
Weight Norm
ReLU
Dense Layer
Temporal Unit
Maximum channel= 128
Conv1D ($C \times 128 \times 1$)
Weight Norm
Conv1D ($128 \times 128 \times 1$)
Weight Norm
Linear Layer ($T \times 128$)
Weight Norm
Linear Layer (128×128)
Weight Norm
Max Pooling (2×2)
ReLU
Conv1D ($64 \times 64 \times 1$)
Weight Norm
Conv1D ($64 \times 32 \times 1$)
Weight Norm
Linear Layer (64×64)
Weight Norm
Linear Layer (64×32)
Weight Norm
Max Pooling (2×2)
ReLU
Conv1D ($16 \times 16 \times 1$)
Weight Norm
Conv1D ($16 \times 4 \times 1$)
Weight Norm
Linear Layer (16×16)
Weight Norm
Linear Layer (16×4)
Weight Norm
ReLU
Dense Layer
Temporal Unit
Maximum channel= 64
Conv1D ($C \times 64 \times 1$)
Weight Norm
Conv1D ($64 \times 64 \times 1$)
Weight Norm
Linear Layer ($T \times 64$)
Weight Norm
Linear Layer (64×64)
Weight Norm
Max Pooling (2×2)
ReLU
Conv1D ($32 \times 32 \times 1$)
Weight Norm
Conv1D ($32 \times 4 \times 1$)
Weight Norm
Linear Layer (32×32)
Weight Norm
Linear Layer (32×4)
Weight Norm
ReLU
Dense Layer
Temporal Unit
Maximum channel= 32
Conv1D ($C \times 32 \times 1$)
Weight Norm
Conv1D ($32 \times 4 \times 1$)
Weight Norm
Linear Layer ($T \times 32$)
Weight Norm
Linear Layer (32×4)
Weight Norm
ReLU
Dense Layer
Temporal Unit

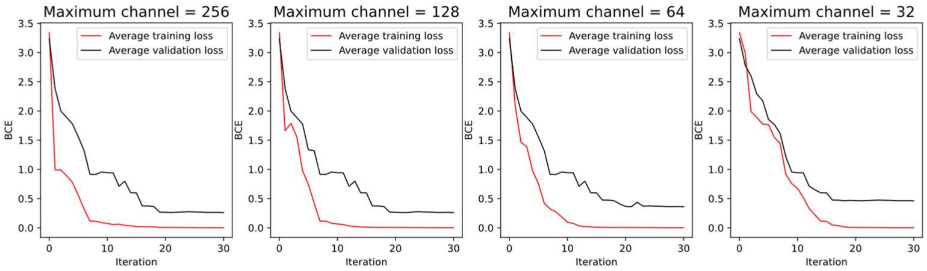
	V-A	V-B	T-A	T-B
AUC	0.864	0.889	0.862	0.886

	V-A	V-B	T-A	T-B
AUC	0.862	0.889	0.859	0.885

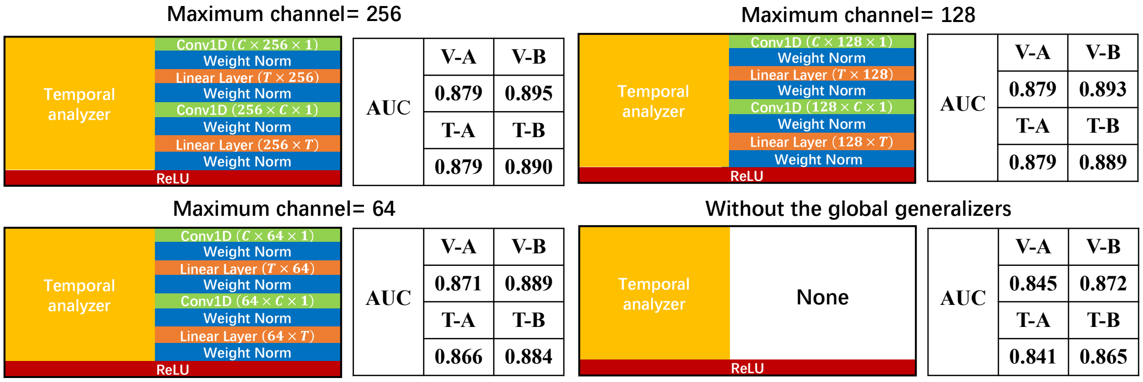
	V-A	V-B	T-A	T-B
AUC	0.851	0.877	0.849	0.871

	V-A	V-B	T-A	T-B
AUC	0.847	0.871	0.844	0.865

A. 1 Hyperparameter tuning of the spatial unit.



A. 2 Average training loss and validation loss of the spatial unit (Batch size = 32).



Maximum channel= 256

Temporal analyzer
Conv1D ($C \times 256 \times 1$)
Weight Norm
Linear Layer ($T \times 256$)
Weight Norm
Conv1D ($256 \times C \times 1$)
Weight Norm
Linear Layer ($256 \times T$)
Weight Norm
ReLU

	V-A	V-B
AUC	0.879	0.895
	T-A	T-B
	0.879	0.890

Maximum channel= 128

Temporal analyzer
Conv1D ($C \times 128 \times 1$)
Weight Norm
Linear Layer ($T \times 128$)
Weight Norm
Conv1D ($128 \times C \times 1$)
Weight Norm
Linear Layer ($128 \times T$)
Weight Norm
ReLU

	V-A	V-B
AUC	0.879	0.893
	T-A	T-B
	0.879	0.889

Maximum channel= 64

Temporal analyzer
Conv1D ($C \times 64 \times 1$)
Weight Norm
Linear Layer ($T \times 64$)
Weight Norm
Conv1D ($64 \times C \times 1$)
Weight Norm
Linear Layer ($64 \times T$)
Weight Norm
ReLU

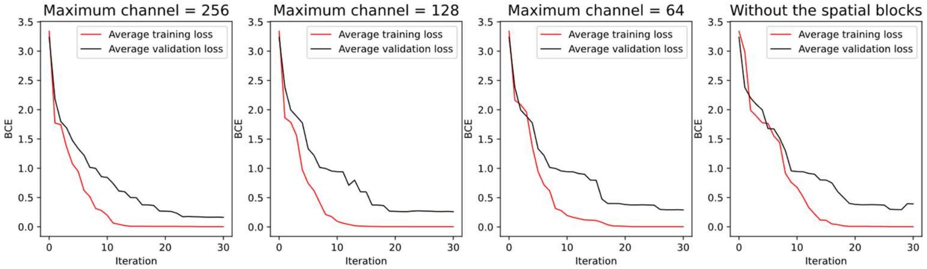
	V-A	V-B
AUC	0.871	0.889
	T-A	T-B
	0.866	0.884

Without the global generalizers

Temporal analyzer
None
ReLU

	V-A	V-B
AUC	0.845	0.872
	T-A	T-B
	0.841	0.865

A. 3 Hyperparameter tuning of the global generalizers in the temporal module.



A. 4 Average training loss and validation loss of the global generalizers in the temporal module (Batch size = 32).

A. 5 Average running time in the first experiment.

Model (# of parameters)	Training time(s)	Testing time(s)
Within-subject P300 detection, Batch size = 32		
STNN-3&15(48827)	16.15	0.026
STNN-4&7(56299)	21.05	0.027
STNN-4&8(56811)	21.27	0.030
STNN-5&3(64283)	23.69	0.038
STNN-5&4(64923)	23.93	0.037
Cross-subject P300 detection, Batch size = 32		
STNN-3&15(48827)	95.5	0.037
STNN-4&7(56299)	112.5	0.039
STNN-4&8(56811)	111.6	0.040
STNN-5&3(64283)	126.1	0.042
STNN-5&4(64923)	125.9	0.042

A. 6 Average running time in the second experiment (Farwell and Donchin's paradigm).

Model (# of parameters)	Training time(s)	Testing time(s)
Within-subject P300 detection, Batch size = 32		
STNN-3&15(70355)	9.07	0.023
STNN-4&7(72371)	10.53	0.029
STNN-4&8(74419)	10.53	0.029
STNN-5&3(76435)	11.57	0.032
STNN-5&4(78994)	11.58	0.032
Cross-subject P300 detection, Batch size = 32		
STNN-3&15(70355)	51.05	0.036
STNN-4&7(72371)	59.30	0.041
STNN-4&8(74419)	59.78	0.042
STNN-5&3(76435)	66.55	0.046
STNN-5&4(78994)	66.74	0.047

A. 7 Average running time in the second experiment (GeoSpell paradigm).

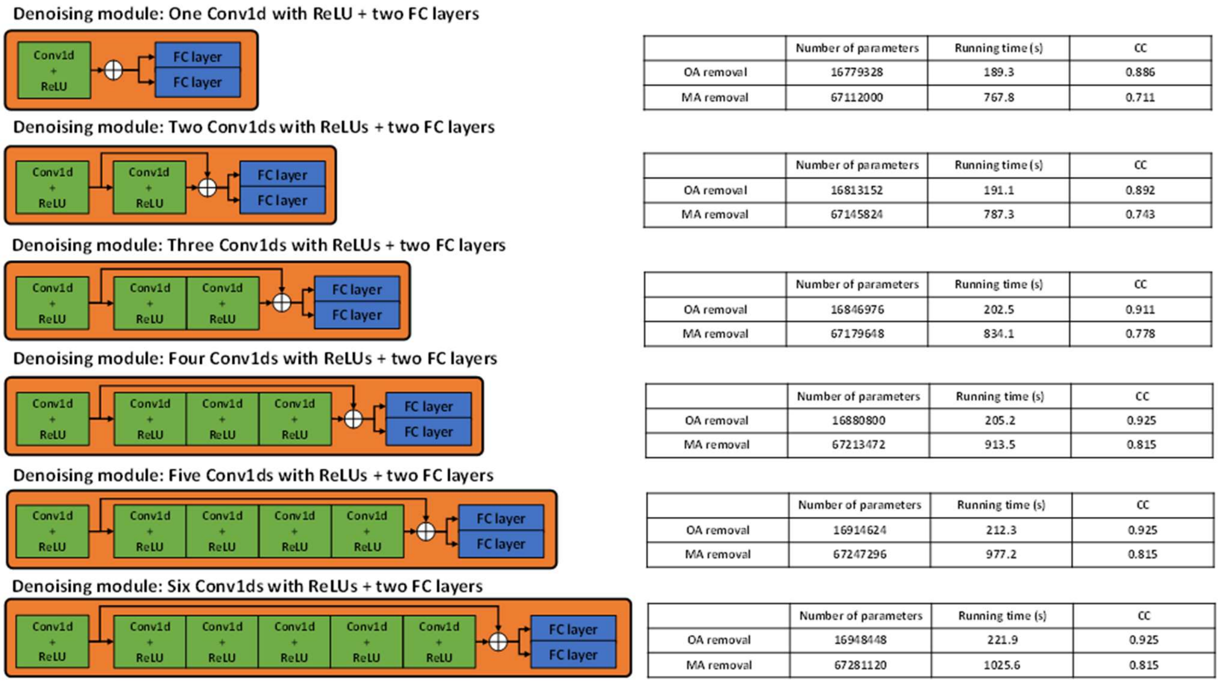
Model (# of parameters)	Training time(s)	Testing time(s)
Within-subject P300 detection, Batch size = 32		
STNN-3&15(70355)	9.22	0.025
STNN-4&7(72371)	10.31	0.027
STNN-4&8(74419)	10.52	0.028
STNN-5&3(76435)	11.44	0.030
STNN-5&4(78994)	11.46	0.030
Cross-subject P300 detection, Batch size = 32		
STNN-3&15(70355)	51.73	0.036
STNN-4&7(72371)	60.50	0.041
STNN-4&8(74419)	60.53	0.043
STNN-5&3(76435)	67.90	0.048
STNN-5&4(78994)	67.91	0.048

A. 8 Average running time in the third experiment.

Model (# of parameters)	Training time(s)	Testing time(s)
STNN-1&60(130019)	29.43	0.029
STNN-1&60-Only with the temporal unit (108737)	27.53	0.027
STNN-2&30(147171)	35.70	0.032
STNN-2&30-Only with the temporal unit (125889)	33.79	0.031
STNN-3&15(201443)	40.62	0.035
STNN-3&15-Only with the temporal unit (180161)	39.40	0.034
STNN-4&7(259331)	45.07	0.036
STNN-4&7-Only with the temporal unit (238049)	44.15	0.035
STNN-4&8(262099)	45.36	0.036
STNN-4&8-Only with the temporal unit (240817)	43.70	0.036
STNN-5&3(272987)	50.40	0.037
STNN-5&3-Only with the temporal unit (251705)	48.90	0.036
STNN-5&4(281947)	50.50	0.037
STNN-5&4-Only with the temporal unit (260665)	48.95	0.040
STNN-6&2(322563)	54.60	0.041
STNN-6&2-Only with the temporal unit (301281)	53.91	0.041
STNN - Only with the spatial unit (21282)	41.05	0.028

A.2 Denoising modules & Compared models – MMNN.

A. 9 shows the different configurations of the denoising modules, in which the number of parameters, and running time increased with the number of Conv1Ds. However, the module performance reaches its limit when using four Conv1Ds. Therefore, we configured four Conv1Ds for the denoising module in our model.



Denoising module: One Conv1d with ReLU + two FC layers

	Number of parameters	Running time (s)	CC
OA removal	16779328	189.3	0.886
MA removal	67112000	767.8	0.711

Denoising module: Two Conv1ds with ReLUs + two FC layers

	Number of parameters	Running time (s)	CC
OA removal	16813152	191.1	0.892
MA removal	67145824	787.3	0.743

Denoising module: Three Conv1ds with ReLUs + two FC layers

	Number of parameters	Running time (s)	CC
OA removal	16846976	202.5	0.911
MA removal	67179648	834.1	0.778

Denoising module: Four Conv1ds with ReLUs + two FC layers

	Number of parameters	Running time (s)	CC
OA removal	16880800	205.2	0.925
MA removal	67213472	913.5	0.815

Denoising module: Five Conv1ds with ReLUs + two FC layers

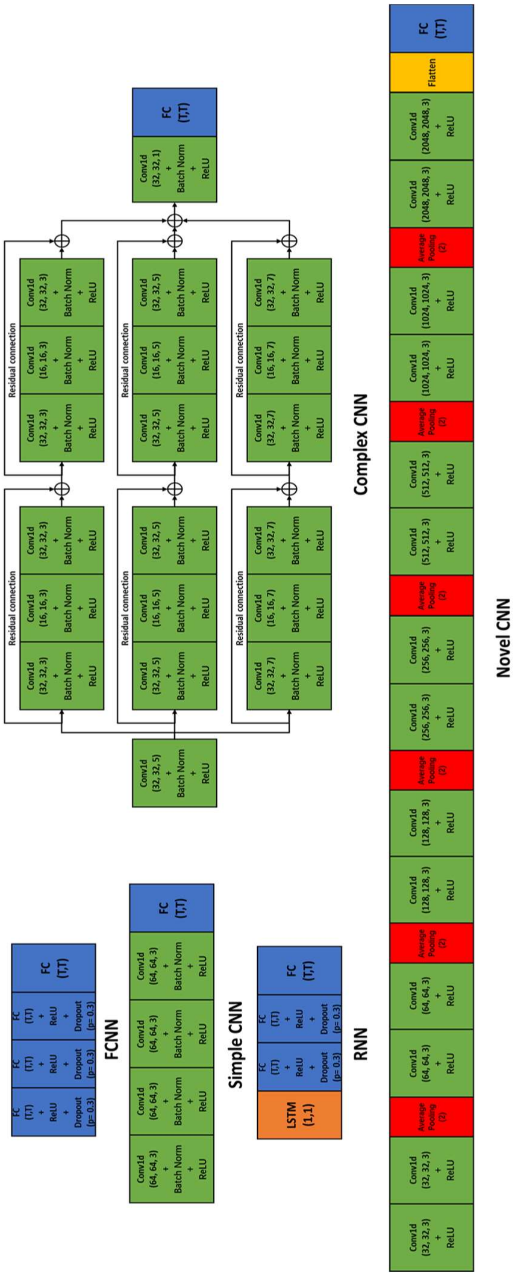
	Number of parameters	Running time (s)	CC
OA removal	16914624	212.3	0.925
MA removal	67247296	977.2	0.815

Denoising module: Six Conv1ds with ReLUs + two FC layers

	Number of parameters	Running time (s)	CC
OA removal	16948448	221.9	0.925
MA removal	67281120	1025.6	0.815

A. 9 Configuration of the denoising modules.

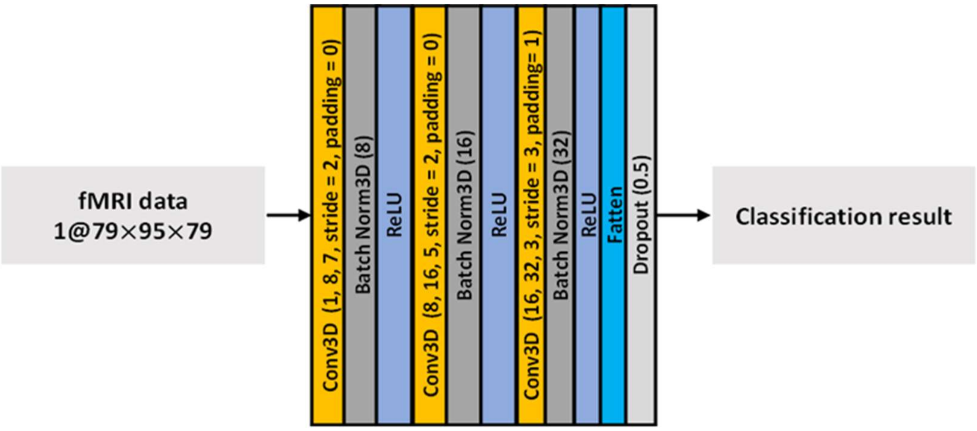
A. 10 presents the architecture of the compared models, including FCNN, Simple CNN, Complex CNN, and RNN, and Novel CNN. These models decomposed noisy EEG signals using the different combinations of Conv1D, FC, and LSTM blocks, and then reconstructed clean signals using an FC block. Hyperparameter formats of “Conv1D”, “FC”, and “LSTM” are (input channel, output channel, kernel size), (input feature, output feature), (input channel, output channel), respectively. “T” the number of discrete time points of the input EEG epoch.


FC (T,T) + ReLU + Dropout (p= 0.3)
FC (T,T) + ReLU + Dropout (p= 0.3)
FC (T,T) + ReLU + Dropout (p= 0.3)
FC (T,T)
FCNN
Conv1d (64, 64, 3) + Batch Norm + ReLU
Conv1d (64, 64, 3) + Batch Norm + ReLU
Conv1d (64, 64, 3) + Batch Norm + ReLU
Conv1d (64, 64, 3) + Batch Norm + ReLU
FC (T,T)
Simple CNN
LSTM (1,1)
FC (T,T) + ReLU + Dropout (p= 0.3)
FC (T,T) + ReLU + Dropout (p= 0.3)
FC (T,T)
RNN
Conv1d (32, 32, 5) + Batch Norm + ReLU
Residual connection
Conv1d (32, 32, 3) + Batch Norm + ReLU
Conv1d (16, 16, 3) + Batch Norm + ReLU
Conv1d (32, 32, 3) + Batch Norm + ReLU
Conv1d (32, 32, 5) + Batch Norm + ReLU
Conv1d (16, 16, 5) + Batch Norm + ReLU
Conv1d (32, 32, 5) + Batch Norm + ReLU
Conv1d (32, 32, 7) + Batch Norm + ReLU
Conv1d (16, 16, 7) + Batch Norm + ReLU
Conv1d (32, 32, 7) + Batch Norm + ReLU
Conv1d (32, 32, 1) + Batch Norm + ReLU
FC (T,T)
Complex CNN
Conv1d (32, 32, 3) + ReLU
Conv1d (32, 32, 3) + ReLU
Average Pooling (2)
Conv1d (64, 64, 3) + ReLU
Conv1d (64, 64, 3) + ReLU
Average Pooling (2)
Conv1d (128, 128, 3) + ReLU
Conv1d (128, 128, 3) + ReLU
Average Pooling (2)
Conv1d (256, 256, 3) + ReLU
Conv1d (256, 256, 3) + ReLU
Average Pooling (2)
Conv1d (512, 512, 3) + ReLU
Conv1d (512, 512, 3) + ReLU
Average Pooling (2)
Conv1d (1024, 1024, 3) + ReLU
Conv1d (1024, 1024, 3) + ReLU
Average Pooling (2)
Conv1d (2048, 2048, 3) + ReLU
Conv1d (2048, 2048, 3) + ReLU
Flatten
FC (T,T)
Novel CNN

A. 10 Architecture of the compared models for EEG denoising.

A.3 Compared models – MP3DCNN.

A. 11 presents the compared architecture for categorical (face vs. object), face sub-categorical (male face vs. female face), and object sub-categorical (natural object vs. artificial object) classifications.


fMRI data
1@79×95×79
Conv3D (1, 8, 7, stride = 2, padding = 0)
Batch Norm3D (8)
ReLU
Conv3D (8, 16, 5, stride = 2, padding = 0)
Batch Norm3D (16)
ReLU
Conv3D (16, 32, 3, stride = 3, padding= 1)
Batch Norm3D (32)
ReLU
Fatten
Dropout (0.5)
Classification result

A. 11 Architecture of the compared models for fMRI classification.

List of Publications

[1] Zhang, Z., Takeda, M., & Iwata, M. (2023). Multi-pooling 3D Convolutional Neural Network for fMRI Classification of Visual Brain States. in Proc. The IEEE Conference on Artificial Intelligence, Santa Clara, California, USA. DOI 10.1109/CAI54212.2023.00057.

[2] Zhang, Z., Yu, X., Rong, X., & Iwata, M. (2022). A Novel Multimodule Neural Network for EEG Denoising. IEEE Access, 10, 49528-49541. DOI: 10.1109/ACCESS.2022.3173261.

[3] Zhang, Z., Yu, X., Rong, X., & Iwata, M. (2021). Spatial-Temporal Neural Network for P300 Detection. IEEE Access, 9, 163441-163455. DOI: 10.1109/ACCESS.2021.3132024.

Acknowledgement

Over the past three years, I have dedicated myself to pursuing a doctoral degree at Kochi University of Technology, and this journey has been challenging, rewarding, and exciting. This period in my life is not only crucial for my academic and professional development, but it is also a time of my personal growth and self-discovery.

I would like to express my heartfelt gratitude to my thesis supervisor, Prof. Iwata, for his unwavering support, guidance, and encouragement throughout my doctoral studies. Prof. Iwata's expertise, dedication, and constructive feedback have been invaluable to my academic and personal growth, and I am grateful for his mentorship.

I would also like to express my sincere gratitude to Prof. Yu, who is my supervisor for the master's degree. I am incredibly grateful for Prof Yu's recommendation and encouragement to pursue a doctoral degree at Kochi University of Technology. Her belief in my abilities and potential has constantly motivated and inspired me. Moreover, I would like to thank Prof. Matsuzaki, Prof. Takeda, Prof. Yoshida, as well as Prof. Rong. Thanks to Prof. Matsuzaki and Prof. Yoshida for their guidance in deep learning theory, Prof. Takeda for his guidance in brain neuroscience as well as Prof. Rong for his support and advice during my doctoral studies.

Finally, I am grateful for the financial support provided by SSP (special scholarship program), which made it possible for me to pursue my doctoral studies. And I would like to express my deepest gratitude to my parents, who have always been my most potent advocates. I am deeply grateful for the values that my parents instilled in me, including the importance of hard work, dedication, and perseverance.