FINAL MACHINE LEARNING IN MEDICINE PROJECT REPORT

BACHELOR RESEARCH

EEG Motor Imagery Classification

Academic Year: 2021-2024

Machine Learning in Medicine Final

EEG Motor Imagery Classification

Presented by:
Duong Nguyen Hoang BI12 - 120
Nguyen Nguyen Cao BI12 - 332
Son Nguyen Dang BI12 - 386
Tung Nguyen Xuan BI12 - 479
Vu Nguyen Xuan Minh BI12 - 497

Contents

Abstract

Brain-Computer Interfaces (BCI) offer a communication channel between the brain and external devices, enabling control without muscle involvement. In terms of EEG (the most popular categories of BCI) Motor Imagery, classifying signals of spontaneous brain activity has always been a matter of concern. Recently, thanks to the advancement in technology, machine learning and deep learning have been widely used in this process. In this project, we utilize methods utilizing convolutional networks to extract features. The target outcome is to classify between the imaginary movements of the right hand, the left hand, and two feet. After experimenting, our two tested models have reached notable results, at 37% and 51% in accuracy in the test set. However, we can consider that the DeepEGG slightly outperformed the ShallowEGG because of the superior test accuracy, although both of them have nearly equal results in the training and validating process.

Introduction

The brain is the largest cluster of neurons in the body, managing various functions of the human body. Nowadays, Brain-Computer Interface (BCI) has become a topic that attracts many researchers' attention. BCI is defined as a technology that enables direct communication between the brain and an external device(computer or machine) without any muscular involvement. As no muscles are required in the process, it allows people to control their bodies by their brains. Therefore, researchers and companies have developed and innovated mobile assistive robots that are capable of receiving brain signals and outputting commands in order to help people with disorders or injuries related to body function. These types of inventions help the patients to live independently, reducing time and effort for their families. Furthermore, in recent years BCI has been widely used in post-stroke rehabilitation and gaming. There are two main categories of BCI: neuron discharge activity(magnetoencephalography (MEG), electroencephalography (EEG), electrocorticogram (ECoG), action potential, and local field potentials (LFPs)) and neuronal metabolic activity. EEG is considered to be the most popular one because of its non-invasiveness and high temporal resolution.

Motor imagery classification is the most common among all the EEG-based BCI paradigms. The process of EEG motor imagery classification generally includes four components: Signal acquisition, Feature extraction, Classification, and Control. Feature extraction is the most crucial part. However, manual methods designed by our knowledge and experience are often limited in terms of classification accuracy and time consumption.

In recent years, convolutional neural networks (CNN) have witnessed notable applications and have been widely used in feature extraction-related fields. The efficiency of CNN and back-propagation has raised the attention of many researchers, scholars, and scientists. The neural network has proved itself to be a proficient technique to extract features from signals, allowing stable output for the majority of cases, and reducing time and effort.

This project aims to apply and compare different state-of-the-art neural network models in an attempt to classify the given signals of spontaneous brain activity.

Data Analysis

Dataset

In this project, we use data set 2a in BCI Competition IV [3]. This data set includes EEG data from 9 subjects. Four different imaginings of movement of the left hand, the right hand, both feet and tongue. In total, there are 2 sessions with 288 trials each. All the data was recorded by twenty-two Ag/AgCl electrodes (with inter-electrode distances of 3.5 cm). The recorded signals were sampled with 250Hz and bandpass-filtered between 0.5 Hz and 100 Hz. Besides 22 EEG channels recorded, other 3 monopolar EOG channels were experimented with. They were sampled with 250Hz and bandpass filtered between 0.5Hz and 100Hz. As the EOG channels are for subsequent application of artifact processing, they are not used in the project.
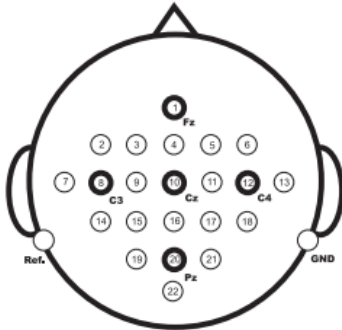

Figure 1: Electrode montage corresponding to the international 10-20 system
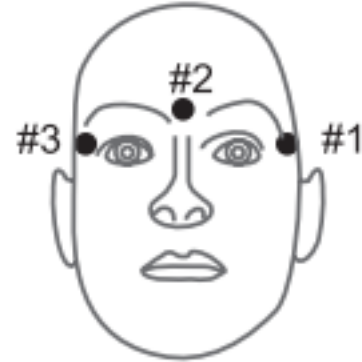

Figure 2: Electrode montage of the three monopolar EOG channels

Data Understanding

For this project, we only take the 3 different classes: Left hand, Right hand, and Both Feet, all the data in the tongue class with be removed from the dataset. As mentioned the 22 channels of EEG are used in the classification task. The data was divided into 30 3628 equal slices of length 750, each can be considered a pseudo image of size 22x750.
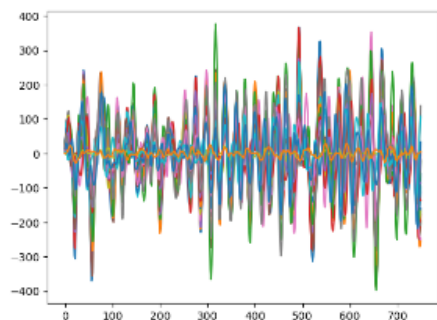
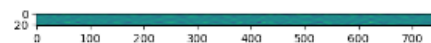Figure 3: Signal views as signal

Figure 4: Signal views as image

Methodology

ShallowEEG

ShallowEEG [1] was designed by Robin Tibor Schirrmeister et al, inspired by FBCSP(Filter Bank Common Spatial Patterns) pipeline, specified for decoding band power features. Particularly, the first layer performs temporal convolution, and the second does spatial convolution. These are similar to parts in FBSCP. The temporal convolution in ShallowEEG has a kernel size of 1x25, allowing a large range of transformations. Following the temporal and spatial convolution is a non-linearity ReLu, a max pooling layer with kernel size (1x75) and stride (1,15), and a logarithmic activation function (softmax). As this network has several pooling regions within one trial, the ShallowEEG can consider the temporal structure.
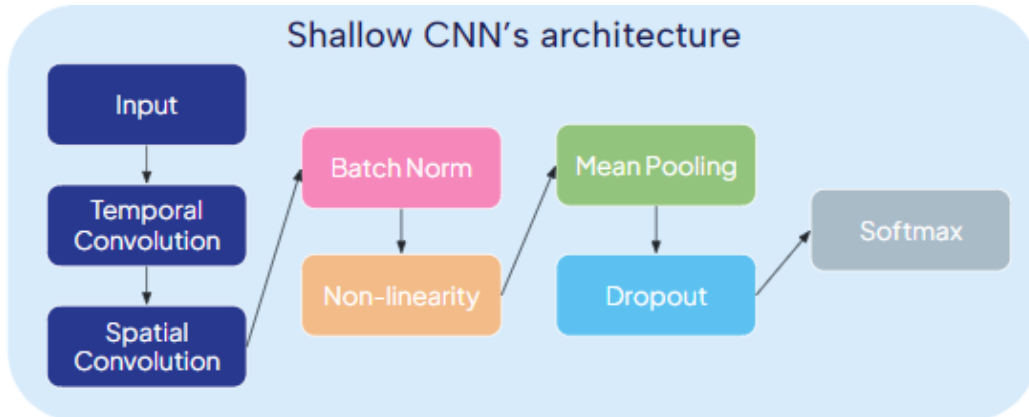

Figure 5: Shallow EEG structure

DeepEEG

Similar to ShallowEEG, DeepEEG [2] also performs the task of EEG decoding. This network is able to extract many features from EEG signals without concern about feature types. Firstly, the data goes through two convolutional layers, one for spatial convolution and the other for temporal convolution. Following the first 2 layers are Batch normalization, non-linearity ReLu, and Mean Pooling. The next part is three convolution blocks including dropout 0.5, batch normalization, and ReLu. The final softmax layers classify the outputted feature maps.
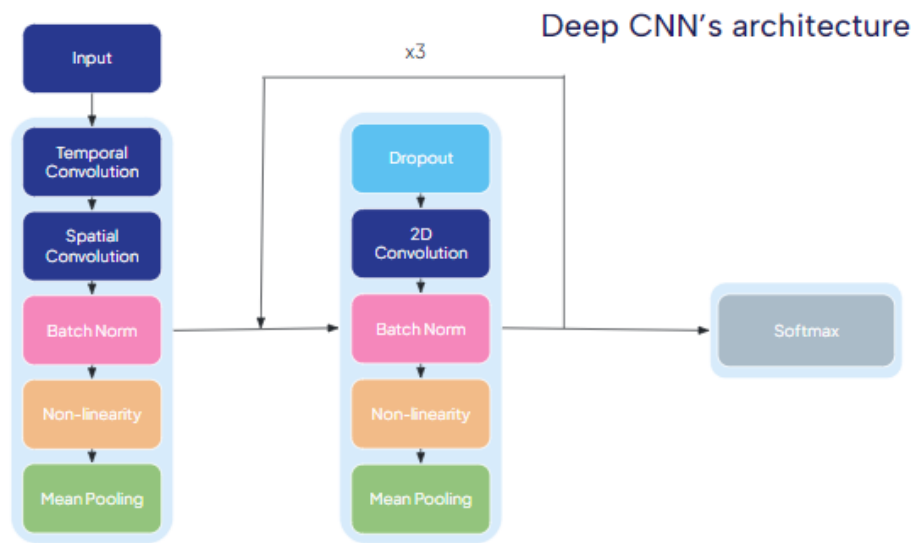
Figure 6: DeepEEG's architecture

Experiment

Data-Preprocessing

Euclidean Alignment

Euclidean Alignment [4] is proposed to align the dataset so that the train and the test set are closer in distribution in Euclidean space, which is a geometric space where distances are determined by the Euclidean distance metric. This method is used because of its simplicity and efficiency. The formula of Euclidean Alignment is defined as:

$$\bar{R} = \frac{1}{n}(\sum_{i=1}^{n} X_i X_i^T) \tag{1}$$

$$\tilde{X}_i = \bar{R}^{-\frac{1}{2}} X_i \tag{2}$$

where $\bar{R}$ accounts for the arithmetic mean of all covariance matrices from a subject, n is the number of trials and $\tilde{X}_i$ is the Observation.
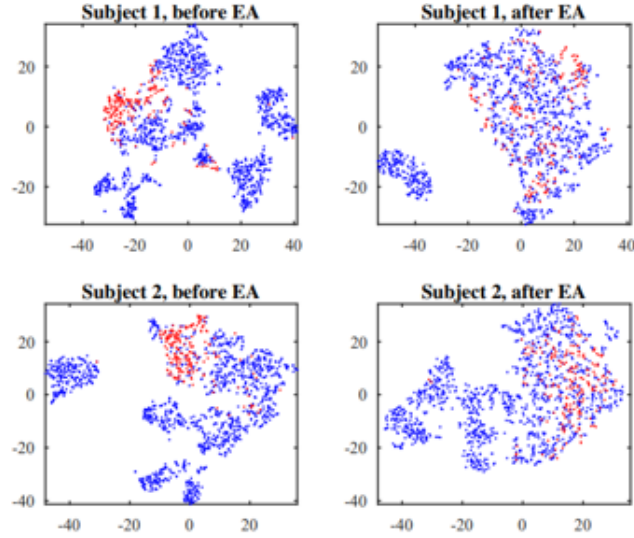

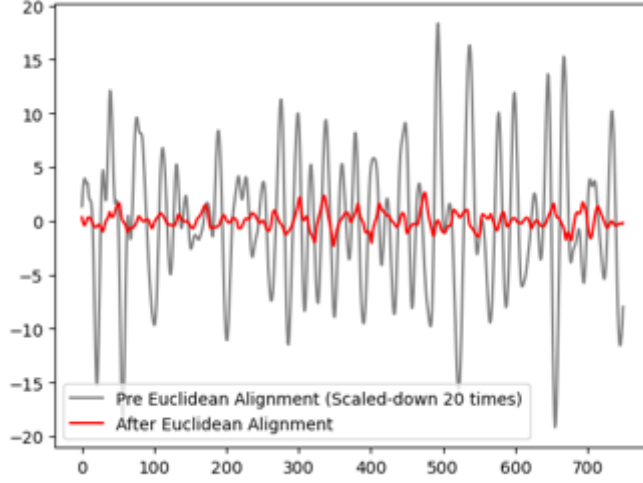Figure 7: Subject 1 and 2 before and after EA (Scaled down to 2-d)

Figure 8: Euclidean Alignment before and after EA

Mean/STD Normalization

Mean/STD normalization, also known as z-score normalization or standardization, is a technique used to scale features in a dataset such that they have a mean of zero and a standard deviation of one. This process is commonly performed in machine learning and data analysis to ensure that features with different scales and units contribute equally to the analysis. Mean/STD Normalization is calculated by:

$$\tilde{X} = \frac{X - \bar{X}}{STDX} \tag{3}$$

Experiment Pipeline

The experimenting process is identical to a normal deep-learning procedure. We implement the proposed methods using the Pytorch framework. After taking the 22-channel data from 9 subjects, 8 subjects are taken for training tasks while 1 subject is for testing. Specifically, there 3052 slices are used for training and 576 slices for the test process. In order to prevent the model from over-fitting, the training data is separated into 90% and 10% for training and validating consequently. The input is then pre-processed, by using 2 mentioned techniques in both training, testing, and validating data. The input testing data are feed-forward through the model, learning the patterns. The outputted model will be tested on the test data to evaluate finally.
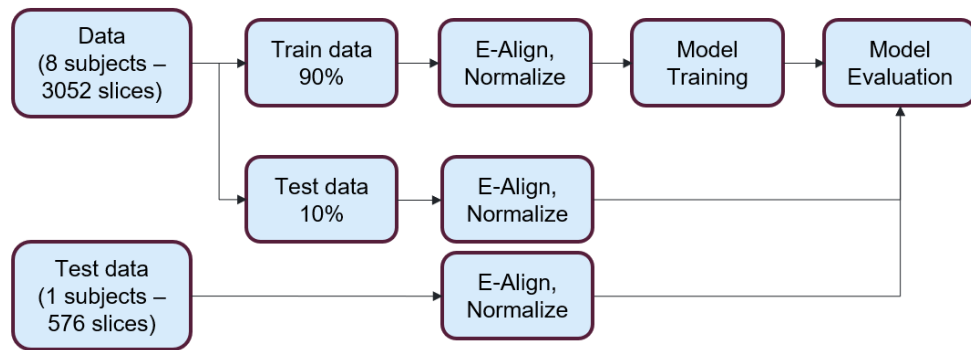
Figure 9: Experimenting Pipeline
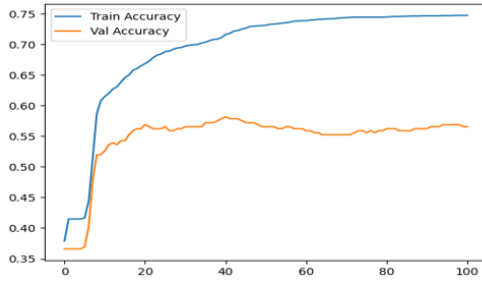
Result and Evaluation

ShallowEEG

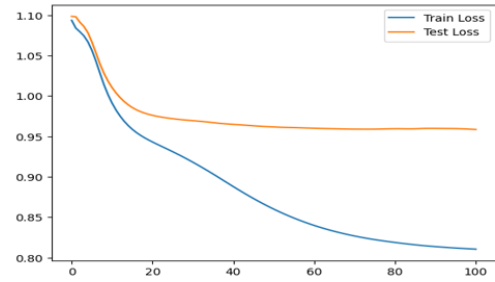Figure 10: Training and Validating accuracy

Figure 11: Training and Validating loss

Overall, it can be seen that the ShallowEEG performs well with the given input data. It drops significantly in the first ten epochs. After that, while the loss for the train set continues to decrease, that for the validate set seems to undergo no notable changes, indicating the model may be over-fitted from epoch 20.

Figure 12: Confusion Matrix

Class	Precision (%)	Recall (%)
1	54	35
2	60	30
3	56	98

Figure 13: Precision and Recall

In terms of testing, the model used ShallowEEG performs outstandingly in detecting the third class, with 98% in recall and 56% in precision. The first class results in 54% in precision but only 35% in recall. The second get 60% in precision and 30% in recall.

9

DeepEEG

Unlike ShallowEEG, the accuracy and loss for training and validating fluctuate more. They start to converge after around 60 epochs, which implies over-fitness after that.
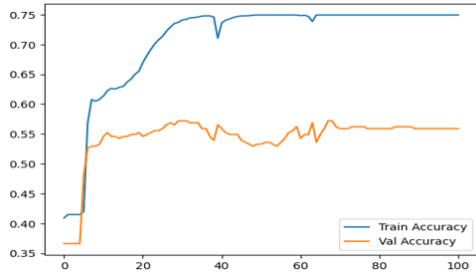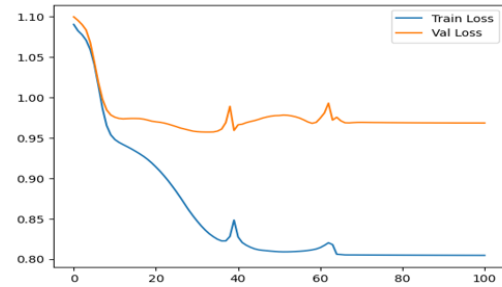

Figure 14: Training and Validating accuracy

Figure 15: Training and Validating loss

Figure 16: Confusion Matrix

Class	Precision (%)	Recall (%)
1	55	35
2	58	31
3	56	94

Figure 17: Precision and Recall

Similar to ShallowEEG, the DeepEEG also achieves good results in the third class, with 94% in recall and 55% in precision. 55% and 35% account for the precision and recall for the first class, meanwhile the second gets 58% in precision and 31% in recall.

Comparision

As trained on the same data set and tested on the same test set, we can easily compare the two models. Overall, both models perform well on both train and test data. However, the test accuracy of DeepEGG is 51%, much higher than that of ShallowEEG. This somehow shows that the DeepEGG performs better with unseen data, here is the test data with one person not included in the train set.

Epoch 100	ShallowEEG	DeepEEG
Train Loss	0.81267	0.81070
Train Accuracy	74.982%	74.690%
Valid Loss	0.97564	0.958514
Valid Accuracy	55.882%	57.516%
Test Loss	1.17124	1.05144
Test Accuracy	37.326%	51%

Figure 18: Comparison of ShallowEEG and DeepEEG

Conclusion and Future Work

In this work, we have investigated and implemented some deep learning models for classification, specifically for labeling the body movements (left hand, right hand, and two feet) using brain signals. The two models that show the best results are ShallowEEG and DeepEEG. The performances of these two models are quite similar and both predicted well on the third class, while poorer detecting first and second classes. To achieve these results, we have applied some methods to handle raw EEG data: Euclidean Alignment and Mean/STD Normalization to make them usable.

However, the evaluation is not as good as expected due to the lack of data. Therefore, in the future, we will apply some data augmentation methods and research for better classification algorithms and models. On the other hand, we will try to implement it in real-time, integrate it with different modalities, do practical applications, and personalize it.

Bibliography

[1] Schirrmeister, R.T.; Springenberg, J.T.; Fiederer, L.D.; Glasstetter, M.; Eggensperger, K.; Tangermann, M.; Hutter, F.; Burgard, W.; Ball, T. *Deep learning with convolutional neural networks for EEG decoding and visualization.* Hum. Brain Mapp. 2017, 38, 5391–5420.

[2] Cooney, C.; Korik, A.; Folli, R.; Coyle, D. *Evaluation of Hyperparameter Optimization in Machine and Deep Learning Methods for Decoding Imagined Speech EEG.* Sensors, 20 (16) (2020), p. 4629

[3] Brunner, C.; Leeb, R.; M¨uller-Putz, G.R; Schl¨ogl, A.; Pfurtscheller, G. *BCI Competition 2008 – Graz data set A.*

[4] He, H.; Wu, D.. *Transfer Learning for Brain-Computer Interfaces: A Euclidean Space Data Alignment Approach.* arXiv:1808.05464 [cs.LG]