The Effects of Subject Transfer on Transformer-Based EEG Classification of Finger Movement

Zach Snow
University of Kentucky
Lexington, Kentucky 40506
zsn222@uky.edu

Adham Atyabi
University of Colorado, Colorado Springs
1420 Austin Bluffs Parkway
Colorado Spring, Colorado 80918
aatyabi@uccs.edu

Abstract

Brain-Computer Interfaces (BCIs) employing Electroencephalogram (EEG) signals are powerful mechanisms for controlling prostheses without physical manipulation. Current methods of processing EEG signals are either limited in their accuracy or in their flexibility, both of which hinder their broad use and application. EEG signals have exceptional temporal resolution but are comparatively lacking in spatial resolution, which can make deep features of their signals difficult to interpret. Vision Transformers (ViTs) specialize in extracting patterns and features in images, and categorizing them by detecting the relative importance of image regions. This functionality makes ViTs strong candidates for translating EEG signals into specific motor functions. Transfer learning is the act of using a model trained on a similar task as a baseline for the intended task to reduce the amount of data required and improve classification accuracy. An implementation of transfer learning in the medical field is subject transfer, a form of transfer learning in which a model is trained on other subjects before being tested or fine-tuned on a target subject. This paper proposes the use of transformers and subject transfer to more effectively classify EEG signals as movements of individual fingers.

1 Introduction

An Electroencephalogram (EEG) is a brain signal recording method that uses multiple electrodes placed on the scalp to take measurements of signals in different areas of the brain. EEGs are often used to evaluate and diagnose neurological conditions such as epilepsy, sleep disorders, and other encephalopathies. Brain-Computer Interfaces (BCIs) are systems that take signals from the brain, analyze them, and relay them as commands to an output device. One increasingly prevalent application of BCIs is in the field of prosthetics, as BCIs have the capacity to convert brain signals from imagined motion into physical motion of a controlled device.

EEGs have high temporal resolution, typically taking samples of brain activity at a sampling rate of between 200 and 1000 times per second. The drawback of EEGs in prosthetics is their poor spatial resolution (Ieracitano et al. 2021). Current EEG systems consist of between 4 and 256 electrodes with the most common layouts having between 20 and 64 electrodes in their placement (Towle et al. 1993). The international standard 10-20 EEG system, in which the T3, C3, Cz, C4, and T4 electrodes are placed at intervals of 10%, 20%, 20%, 20%, 20%, and 10% across the center of the head respectively, is pictured in Figure 1. Even with 256 electrodes, it is impossible to measure brain activity with any reasonable degree of spatial precision. Since the number of neurons in a human brain is 10 orders of magnitude higher than the number of electrodes in a typical EEG, EEGs can only detect the activity of large clusters of neurons which leads to results that can be challenging to interpret.

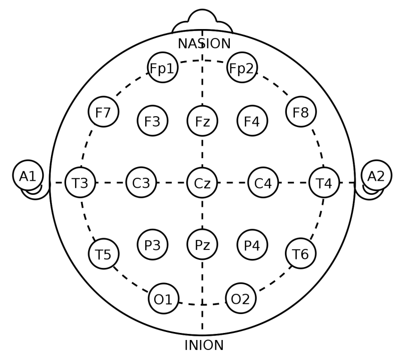

Figure 1: Electrode locations of the 10-20 system (Towle et al. 1993).

Due to the relatively poor spatial resolution of EEGs, it can be difficult to discern between movements of small, specific body parts such as individual fingers. Additionally, EEG results for the same motion can vary drastically between different test subjects, and can even vary for the same subject at different times. In spite of this, having the ability to accurately discern between different finger movements is critical to the future development of more functional prostheses. This paper aims to improve upon current models used for EEG classification using transformers and subject transfer, which has not been used for the task of finger classification.

2 Related Works

In 2017, Google Brain and Google Research published the paper "Attention is All you Need," which introduced a novel neural network architecture known as a transformer

(Vaswani et al. 2017). Transformers employ "self-attention" which allows them to identify the importance of a given element in the input in relation to other elements in the input.

Transformers are the new preferred architecture for Natural Language Processing (NLP) tasks due to their ability to learn how words in a sentence relate to other words. Furthermore, transformers are rapidly becoming one of the preferred methods of classification in Computer Vision (CV) due to their ability to extract information about how certain pixels of an image relate to other pixels. This could prove useful in EEG and Electrocorticogram (ECoG) classification because regions in those signals are not independent of other regions.

Convolutional Neural Networks (CNNs) are similar to traditional dense artificial neural networks in their functionality. Both serve to extract deep features in a data set to classify new data, but due to the structure of CNNs, CNNs tend to use significantly fewer parameters while maintaining comparable accuracy with two- and three-dimensional data such as images. Instead of connections between every individual node, CNNs utilize "kernels." A kernel can be thought of as a small filter that passes over the image and accentuates desirable features.

CNNs excel in mapping convoluted patterns in images to an output class. Many of the top performing neural networks designed to work with images today are CNNs. Consequently, CNNs are a method of identifying deep features of EEG data and classifying them as movement of a specific body part (Sadiq et al. 2022).

In 2021, Ieracitano et al. (2021) showed strong results in using Convolutional Neural Networks (CNNs) to discriminate between open-hand/rest and closed-hand/rest positions using EEG signals. Their work was able to classify pre-hand closed versus rest and pre-hand open versus rest with an accuracy of approximately 90%. Furthermore, they were able to discriminate between preparation of different sub-movements with a precision of approximately 62%.

Kim et al. (2021) discovered that employing Sequential Transfer Learning has the potential to further increase classification accuracy of Motor Imagery (MI) of hand, foot, and tongue movements. Their pretrained model correctly identified the body part being moved with an accuracy of 63.8% (with transfer learning) compared to a baseline of 61.6% (without transfer learning). Consequently, the performance of their MI-BCI showed improvement when compared to previous CNN-based approaches. In cases with smaller datasets, it can be beneficial to train on similar data prior to fine-tuning on the intended dataset.

In their 2022 paper, Khademi, Ebrahimi, and Kordy (2022) constructed a hybrid CNN and Long-Short Term Memory (LSTM) deep learning model to classify motor imagery signals from the BCI Competition IV Dataset 2a, a dataset with 4 motor imagery classes (left hand, right hand, feet, and tongue). They achieved a classification accuracy of approximately 90% with their hybrid model of ResNet-50 (a pretrained CNN) and an LSTM.

The electrodes in an ECoG, are placed directly on the cerebral cortex as, opposed to electrodes in an EEG, which are placed on the scalp. Consequently, ECoG data can be more reliable for predicting fine motor movement due to its improved spatial and temporal resolutions as compared to EEG data. As a result, research by Xie, Schwartz, and Prasad (2018) showed promise in being able to distinguish finger movement using a Long-Short Term Memory (LSTM) architecture.

3 Methods

A combination of pre-existing pre-processing methods, ensemble learning, and transfer learning were employed to improve accuracy of classification.

3.1 Obtaining Data

General Overview: Data used was obtained from a large, publicly available EEG motor imagery dataset for brain-computer interfaces. This dataset contains approximately 60 hours of EEG recordings spread over 75 recording sessions of 13 participants (8 of which participated in the 5 finger, or 5F experiment). It contains more than 60000 examples of motor imageries in 4 interaction paradigms (one of which is the 5F paradigm) and is one of the largest publicly-available EEG BCI datasets currently published (Kaya et al. 2018). For more details about this dataset, please see Table 1.

Stimuli and Experimental Design: Participants in the 5F (five finger) experiment were first seated in front of a screen. The experiment proceeded as follows: "At the beginning of each trial, an action signal appeared (represented by a number from 1 to 5) directly above the finger whose movement imagery was to be implemented. The action signal remained on for 1s, during which time the participants implemented the corresponding imagery once. The imageries were invoked as a flexion of the corresponding fingers up or down, per the preference of the participant. There was no passive state in this paradigm – each action signal required a response. Single imagery was implemented per action signal. After executing the imagery, participants remained passive until the next action signal presentation." (Kaya et al. 2018)

3.2 Preprocessing

Overview: Preprocessing was conducted using MATLAB's PREP Pipeline for standardized EEG preprocessing (Bigdely-Shamlo et al. 2015). The PREP Pipeline allows for automatic selection and rejection of epochs (lengths of time that a stimulus is presented), automatic rejection of channels, removal of line noise, and detrending of the data. Beyond this, electrodes Cz, C3, C4, T3, T4, Fz, F3, and F4 are being utilized due to their close proximity to the motor cortex, the area in the brain that primarily controls motor movement.

Process: All EEG data was imported directly into MATLAB. Channels Cz, C3, C4, T3, T4, Fz, F3, and F4 were extracted from the dataset prior to preprocessing in order to reduce computational demand. These channels were selected due to their proximity to the motor cortex, which is the region of the brain controlling motor function. Data was detrended and line noise was reduced using the standard PREP Pipeline settings. Finally, data was referenced using

Dataset	Type of Data	Classes	Sample Rate (Hz)	Channels	Subjects	Epoch Length	Task
BCI 5F Dataset	EEG	5	200 and 1000	22	8	1 second	Finger Movement

Table 1: Dataset Information

a RANSAC approach. Data was saved as text files and then converted to MATLAB files to be transformed via Short-time Fourier transform.

3.3 Overlapping Methods

One preliminary issue that reduced classification accuracy was a lack of data. There were approximately 130 examples of each motor task per subject, which was not enough to effectively train a vision transformer. In order to improve classification accuracy, the number of images being fed into the vision transformer was increased. To do this, a process known as "overlapping" was performed in which multiple offset windows are taken from each epoch.

Standard Overlapping: In standard overlapping, extracted windows are overlapped by a set amount across the full epoch. The typical overlapping percentage ranges from 10% to 90%. For the purpose of this paper, overlapping percentages of 25%, 50%, 75%, and 90% were tested with a window of 0.5 seconds. 75% overlapping showed the most promising results compared to the computational demand and was thus used for further testing.

Triangular Overlapping: Triangular overlapping is an overlapping method proposed by Atyabi, Fitzgibbon, and Powers (2012) in which desirable sections of the EEG data are overlapped at a higher percentage than other, less desirable sections. For example, in a 1-second long epoch, the beginning of the signal is not useful due to the participants reaction time, and the end of the signal isn't as useful due to mental fatigue. Consequently, overlapping the center of the signal more densely will lead to the model being trained on higher-quality data. A modified triangular overlapping approach was taken to ensure that there were no duplicate samples which would lead to challenges in splitting of test/train/validation data prior to training. Unfortunately, this method was infeasible for application to all subjects due to its computational demand.

3.4 Short-Time Fourier Transform

Once the EEG data is preprocessed, it would be ready to be classified if the intended classification model was built to handle sequential data, such as a Long Short-Term Memory model (Wang et al. 2018) or a Recurrent Neural Network (Ma et al. 2018). However, sequential data is not useful for a vision-based classifier such as a CNN or a Vision Transformer. Consequently, the data was passed through a script that applied a Short-time Fourier transform (STFT) to each individual channel of every epoch of every trial. STFT converts a periodic function into a spectrogram by determining the sinusoidal frequency and phase content of the function over time. In an STFT, a short window function is slid across the signal over the time axis, and the resulting image is a construction of the individual, short-time dot products of the window function and the signal. An STFT-generated spectrogram of one epoch of EEG data can be seen in Figure 2.

$$\mathbf{STFT}\{x(t)\}(\tau, \omega) = \int_{-\infty}^{\infty} x(t)\omega(t-\tau)e^{-i\omega t}dt \qquad (1)$$

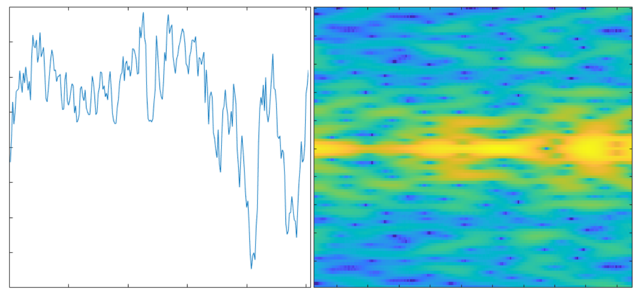

Figure 2: Comparison of raw signal (left) and transformed spectrogram (right).

3.5 Ensemble Learning

Another remedy to the lack of data was the utilization of several channels of the EEG. To train the model on multiple channels, ensemble learning was implemented. Ensemble learning requires the simultaneous training of multiple models. Each model was responsible for the classification of a specific channel. To decide the classification of an input, each model in the ensemble "voted" for the class that it decided the input falls into. This method improves classification accuracy by spreading the data amongst multiple models, reducing the required information storage of each individual model. The drawback of this approach is that it takes significantly longer to train than a single-transformer model. A visualization of this model can be seen in Figure 3.

3.6 Implementing a Transformer

The transformers used to classify the electroencephalogram data are BERT-like transformer encoder models that have been pretrained on ImageNet-21k and ImageNet 2012. Pretrained models generally require less data to train and achieve solid performance in less time than comparable untrained models (Kim et al. 2021). Training with limited data is critical in this case as access to large, publicly available EEG datasets is limited.

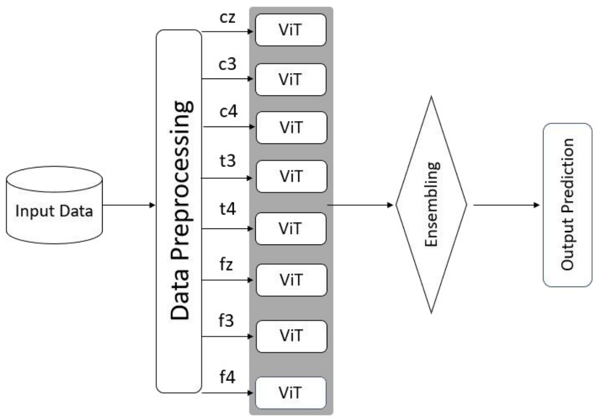

Figure 3: Visualization of Transformer Ensemble

3.7 Subject Transfer

Transfer learning is the act of using a model trained on one task as a baseline for another task. In this case, the model was trained on seven of the eight subjects before being used to classify the final subject. Two methods of subject transfer were implemented. The first method involved training the transformer ensemble on seven subjects and testing on the last subject with no fine-tuning. The other method involved training the ensemble on seven subjects followed by fine tuning on a randomly selected half of the samples from the final subject.

4 Results

4.1 Single Subject: No Fine Tuning

Subject C, Trial 1. Results were obtained by training a separate transformer on channels Cz, C3, C4, T3, T4, Fz, F3, and F4 from subjects A, B, E, F, G, H, and I before testing on subject C without any fine-tuning. Training was performed with a batch size of 32, a learning rate of 2E-5, and 10 epochs. During testing, each transformer was responsible for predicting the class of one channel from the input. Class was chosen via majority voting. Ties were broken by selecting the lowest tied value. For these results, see Table 2. Precision, recall, and F1 score are an average across all classes. For specific precision, recall, and F1 scores by class, see the appendix.

Subjects Pretrained On	Precision	Recall	F1
A	.195	.199	.182
A, B	.206	.201	.187
A, B, E	.206	.203	.189
A, B, E, F	.212	.220	.202
A, B, E, F, G	.194	.196	.185
A, B, E, F, G, H	.210	.213	.191
A, B, E, F, G, H, I	.198	.196	.167

Table 2: Single Subject with No Fine-Tuning: Results

4.2 Single Subject: With Fine-Tuning

Subject C, Trial 1. Results were obtained by training a separate transformer on channels Cz, C3, C4, T3, T4, Fz, F3, and F4 from subjects A, B, E, F, G, H, and I. This training was followed by randomly splitting Subject C, Trial 1 into 50% train, 50% test. Training was performed with a batch size of 32, a learning rate of 2E-5, and 10 epochs for both pretraining and fine-tuning. During testing, each transformer predicted the class based on the input channel. Class was chosen via majority voting. Ties were broken by selecting the lowest tied value. For these results, see Table 4. Precision, recall, and F1 score are an average across all classes. For specific precision, recall, and F1 scores by class, see the appendix.

Subjects Pretrained On	Precision	Recall	F1
A, B, E, F, G, H, I	.238	.240	.234

Table 3: Single Subject with Fine Tuning: Results

4.3 Single Subject: No Subject Transfer

Subject C, Trial 1. Results were obtained by training a separate transformer on channels Cz, C3, C4, T3, T4, Fz, F3, and F4 from 50% of Subject C. This training was followed by testing on the remaining 50% of Subject C. Training was performed with a batch size of 32, a learning rate of 2E-5, and 10 epochs. During testing, each transformer predicted the class based on the input channel. Class was chosen via majority voting. Ties were broken by selecting the lowest tied value. For these results, see Table 3. Precision, recall, and F1 score are an average across all classes. For specific precision, recall, and F1 scores by class, see the appendix.

Subjects Pretrained On	Precision	Recall	F1
None	.201	.215	.200

Table 4: Single Subject with no Subject Transfer: Results

5 Future Work

5.1 Weighted Voting

In the current transformer ensemble, each transformer gets a “vote” on which class it believes the input falls into. All votes are weighted equally and the class that gets the majority of votes is the class that the ensemble assigns the input. In reality, some transformers classify their channel more effectively than other transformers in the ensemble. Consequently, weighting votes in favor of more accurate transformers should lead to increased classification accuracy.

5.2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have been the state-of-the-art for image classification for years. Consequently, over 70% of deep-learning approaches to EEG classification employ CNNs (Al-Saegh, Dawwd, and Abdul-Jabbar 2021).

In the future, a CNN-based classification model will be implemented to compare results with the more recently introduced transformer-based classification model.

6 Conclusion

In order to create more capable prostheses, it is paramount that more work is done to be able to accurately discriminate between small, precise motor movements. As of now, there is much improvement to be done on current models that distinguish between individual finger movements. The goal of this project is to build upon current models and strive toward more accurate CNN- and transformer-based EEG classification in order to further advance the field of prosthetics.

7 Acknowledgement

The work reported in this paper is supported by the National Science Foundation under Grant No. 2050919. Any opinions, findings and conclusions or recommendations expressed in this work are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Al-Saegh, A.; Dawwd, S. A.; and Abdul-Jabbar, J. M. 2021. Deep learning for motor imagery eeg-based classification: A review. *Biomedical Signal Processing and Control* 63:102172.

Atyabi, A.; Fitzgibbon, S.; and Powers, D. 2012. Multiplication of eeg samples through replicating, biasing, and overlapping. volume 7670, 1–13.

Bigdely-Shamlo, N.; Mullen, T.; Kothe, C.; Su, K.-M.; and Robbins, K. A. 2015. The prep pipeline: standardized preprocessing for large-scale eeg analysis. *Frontiers in neuroinformatics* 9:16.

Ieracitano, C.; Mammone, N.; Hussain, A.; and Morabito, F. C. 2021. A novel explainable machine learning approach for eeg-based brain-computer interface systems. *Neural Computing and Applications* 1–14.

Kaya, M.; Binli, M. K.; Ozbay, E.; Yanar, H.; and Mishchenko, Y. 2018. A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces. *Scientific data* 5(1):1–16.

Khademi, Z.; Ebrahimi, F.; and Kordy, H. M. 2022. A transfer learning-based cnn and lstm hybrid deep learning model to classify motor imagery eeg signals. *Computers in Biology and Medicine* 143:105288.

Kim, D.-K.; Kim, Y.-T.; Jung, H.-R.; Kim, H.; and Kim, D.-J. 2021. Sequential transfer learning via segment after cue enhances the motor imagery-based brain-computer interface. In *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*, 1–5.

Ma, X.; Qiu, S.; Du, C.; Xing, J.; and He, H. 2018. Improving eeg-based motor imagery classification via spatial and temporal recurrent neural networks. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1903–1906. IEEE.

Sadiq, M. T.; Aziz, M. Z.; Almogren, A.; Yousaf, A.; Siuly, S.; and Rehman, A. U. 2022. Exploiting pretrained cnn models for the development of an eeg-based robust bci framework. *Computers in Biology and Medicine* 105242.

Towle, V. L.; Bolaños, J.; Suarez, D.; Tan, K.; Grzeszczuk, R.; Levin, D. N.; Cakmur, R.; Frank, S. A.; and Spire, J.-P. 1993. The spatial location of eeg electrodes: locating the best-fitting sphere relative to cortical anatomy. *Electroencephalography and clinical neurophysiology* 86(1):1–6.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems* 30.

Wang, P.; Jiang, A.; Liu, X.; Shang, J.; and Zhang, L. 2018. Lstm-based eeg classification in motor imagery tasks. *IEEE transactions on neural systems and rehabilitation engineering* 26(11):2086–2095.

Xie, Z.; Schwartz, O.; and Prasad, A. 2018. Decoding of finger trajectory from ecog using deep learning. *Journal of neural engineering* 15(3):036009.

8 Appendix

No Fine-Tuning	**A**					**A,B**					**A,B,E**				
Class	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Precision	.177	.205	.204	.214	.175	.184	.173	.202	.241	.230	.178	.245	.172	.236	.198
Recall	.396	.112	.204	.142	.135	.398	.114	.245	.124	.126	.399	.131	.166	.205	.113
F1 Score	.245	.145	.199	.171	.152	.252	.137	.222	.163	.163	.246	.171	.169	.219	.143

No Fine-Tuning	**A,B,E,F**					**A,B,E,F,G**					**A,B,E,F,G,H**				
Class	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Precision	.165	.218	.220	.218	.238	.184	.230	.187	.191	.178	.182	.225	.224	.209	.211
Recall	.316	.135	.175	.234	.238	.118	.280	.328	.169	.083	.340	.139	.398	.113	.078
F1 Score	.217	.167	.195	.225	.206	.144	.252	.238	.180	.113	.237	.172	.287	.147	.114

No Fine-Tuning	**A,B,E,F,G,H,I**				
Class	1	2	3	4	5
Precision	.166	.170	.233	.254	.167
Recall	.502	.134	.175	.130	.039
F1 Score	.249	.149	.200	.172	.063

With Fine-Tuning	**A,B,E,F,G,H,I,C**				
Class	1	2	3	4	5
Precision	.190	.232	.230	.264	.275
Recall	.257	.348	.154	.224	.219
F1 Score	.219	.279	.184	.242	.244

No Subject Transfer					
Class	1	2	3	4	5
Precision	.202	.235	.140	.168	.264
Recall	.144	.439	.106	.172	.214
F1 Score	.168	.306	.120	.170	.236