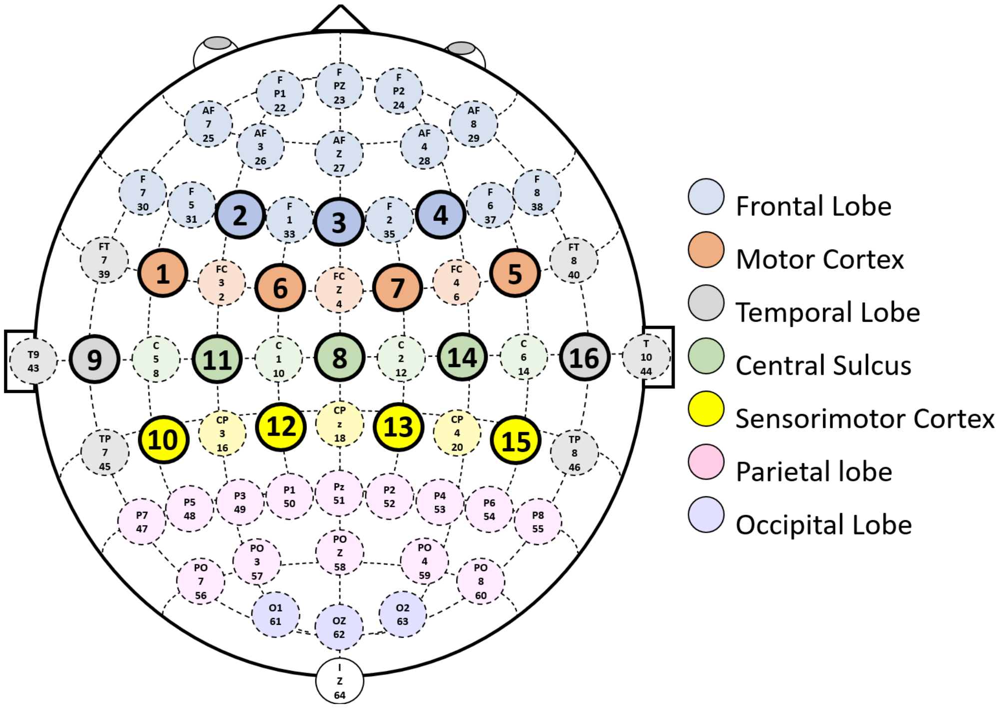
<u>Classification of motor imagery and motor execution based on EEG signals</u>

Context:-If I were to tell you to imagine about doing a task like have dorsal Inflexion(bend ur feet backwards) and then I tell you to actually do it in real time,like in the picture

and during the whole process if I put an electrode cap with designated electrodes at the 10-10 configuration



On each electrode we will be noticing some changes in the electrode potentials and there will be a spike in those potentials whenever you were successful to build that thought or action and that is our raw data for data.

Topic Explanation:-This project tries to classify those data into imagery and action by using the methods discussed in the class.There were 8 tasks namely

The tasks are described below:
- Recording a Baseline with Eyes Open (BEO) without any task command: only once at the beginning of each run.
- Closing Left Hand (CLH): .
- Closing Right Hand (CRH): .
- Dorsal flexion of Left Foot (DLF): .
- Plantar flexion of Left Foot (PLF): .
- Dorsal flexion of Right Foot (DRF): .
- Plantar flexion of Right Foot (PRF):
- Resting in between tasks (Rest):

This project used 6 tasks as resting and BEO would trigger another dimension for classification.

Chosen dataset -MILimbEEG: An EEG Signals Dataset based on Upper and Lower Limb Task
During the Execution of Motor and Motorimagery
Taskshttps://data.mendeley.com/datasets/x8psbz3f6x/2

Problem statement and motivation- Classification between motor imagery and motor action Motivation for choosing this dataset/problem-The MILimbEEG dataset offers a comprehensive collection of over 8,680 four-second EEG recordings from 60 volunteers, capturing both motor execution and motor imagery tasks involving various limb movements with roughly 1.8 million data points which covers the problem of having enough data.Classification between motor Imagery and motor action would help in aiding brain-computer-machine interfaces which can help privileged people have better machines aiding them eg. bionic hands,bionic legs etc.The project aims to evaluate the classification task with the topics discussed in class.

Method

Dataset Consolidation:

Merged all columns into one master table with Subject ID and Task Type(Target feature)
Resulted in18 columns (16 electrodes and One Subject ID and Task Type)and 1.8 million rows

	count	mean	std	min	25%	50%	75%	max
FC5	179950	0.05338	111.188	-187058	-5.8022	-0.0319	5.77338	11235.2
F3	179950	0.005181	139.0794	-78819.8	-6.30272	-0.0677	6.18401	38595.3
Fz	179950	0.083454	129.9186	-31737.1	-5.97712	-0.08754	5.845963	64908
F4	1799500	0.048843	108.5498	-12059.3	-6.36803	-6.40E-23	6.280631	5543.64
FC6	1799500	0.070627	112.0368	-13414.4	-6.44659	-0.00912	6.421152	22445.07
C5	1799500	0.053149	115.2273	-26786.4	-5.96786	6.34E-24	5.968652	13140.15
C3	1799500	0.079129	113.5174	-9380.59	-5.91625	-0.01912	5.891199	13662.29

Cz	1799500	0.06828	108.7371	-6217.39	-5.3509	6.34E-24	5.351015	5543.67
C4	1799500	0.027221	119.7316	-35019.6	-5.63134	3.65E-05	5.644809	12413.88
C6	1799500	0.008226	113.8248	-25435.8	-5.59258	-0.04422	5.551193	6500.451
CP5	1799500	0.017264	114.5309	-28655.9	-6.28849	-0.00759	6.295945	7351.358
CP3	1799500	0.03589	113.4842	-7658.63	-5.73298	6.34E-24	5.703262	7944.876
CP1	1799500	0.050714	113.2431	-7659.09	-5.87039	-0.02754	5.821088	7536.232
CP2	1799500	0.031215	115.6334	-20400.6	-6.02751	-0.00579	6.039154	6501.645
CP4	1799500	0.009923	114.2252	-29681.6	-5.94776	-0.02814	5.932172	6499.033
CP6	1799500	0.035999	137.2146	-35707	-6.1025	1.06E-22	6.130665	29869.36

Outlier Handling:

After running the statistics:
Identified high, unrealistic and inconsistent values.
Observed the stable values within the interquartile range.
Applied the Interquartile Range (IQR) method to filter outliers.
Reconstructed the cleaned dataset with Subject ID and Task Type.
Ran the statistical measure again and now it was consistent

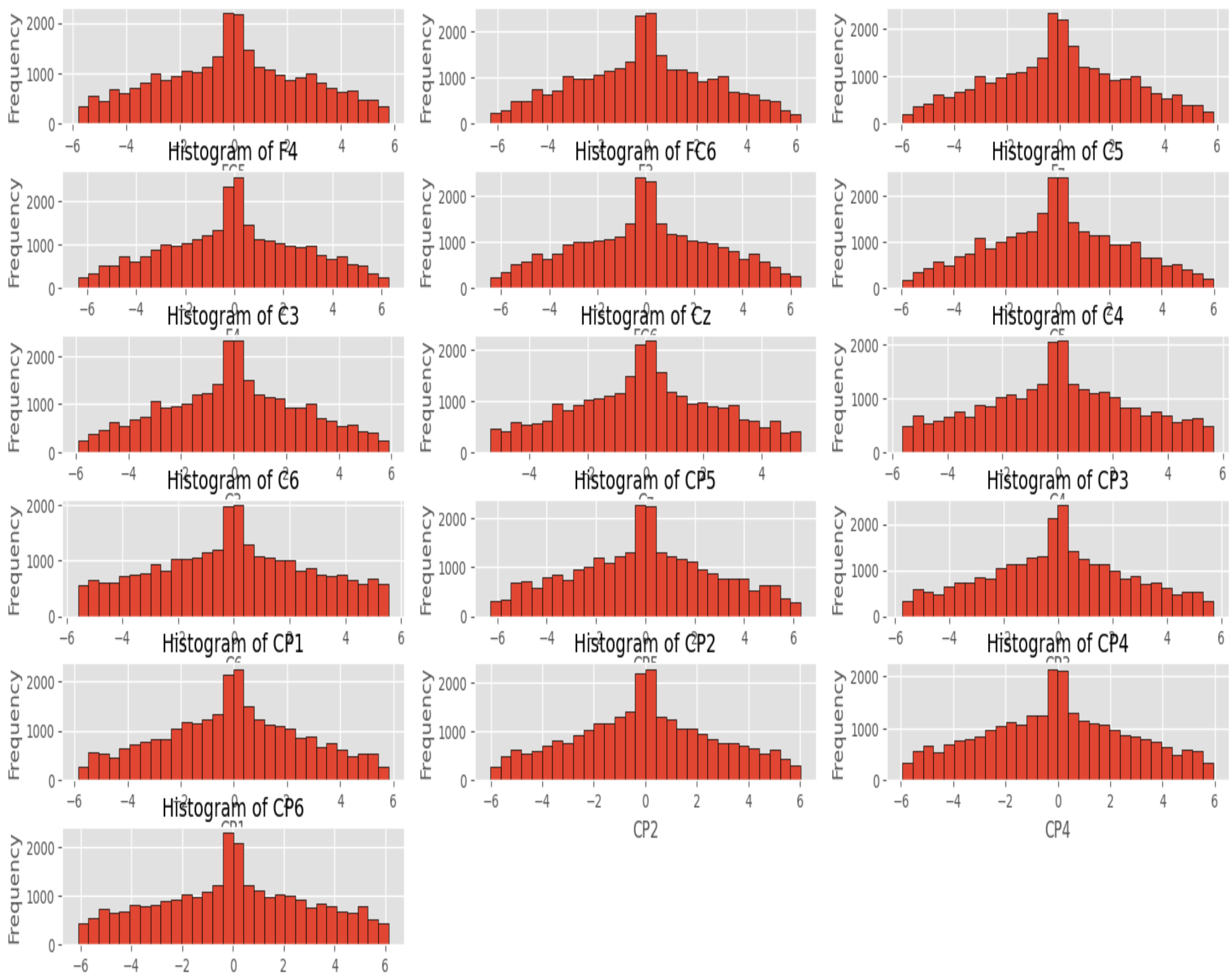
	count	mean	std	min	25%	50%	75%	max
FC5	27525	-0.01734	2.732473	-5.80213	-1.98949	-0.0102	1.948798	5.772496
F3	27525	-0.06697	2.764306	-6.3011	-2.03453	-0.04406	1.868423	6.183946
Fz	27525	-0.05576	2.637821	-5.9757	-1.90068	-0.04744	1.779793	5.841752
F4	27525	-0.01143	2.833461	-6.367	-1.97816	-0.00176	1.976922	6.278633
FC6	27525	-0.02405	2.902855	-6.44519	-2.11075	-0.0203	2.028628	6.419731
C5	27525	-0.04585	2.571867	-5.96775	-1.8197	-0.04959	1.733263	5.968126
C3	27525	-0.02497	2.633964	-5.91505	-1.87278	-0.01637	1.810051	5.888282
Cz	27525	-0.01871	2.500444	-5.35083	-1.78541	-0.0066	1.749215	5.349948
C4	27525	0.012272	2.757749	-5.63132	-1.95572	0.019896	1.966904	5.644806
C6	27525	-0.0191	2.783912	-5.59257	-2.0603	-0.02064	2.013624	5.550914
CP5	27525	-0.05892	2.905007	-6.28849	-2.1075	-0.03683	1.950291	6.29291
CP3	27525	-0.03338	2.638134	-5.73223	-1.85957	-2.18E-22	1.77721	5.698298
CP1	27525	-0.02543	2.695197	-5.86435	-1.90014	-0.00467	1.841827	5.819343
CP2	27525	-0.0321	2.78406	-6.02433	-1.96321	-0.0499	1.888904	6.032268
CP4	27525	-0.068	2.82123	-5.94411	-2.11051	-0.06855	1.932492	5.93114
CP6	27525	0.024267	3.001047	-6.10216	-2.15573	0.01008	2.213329	6.130176

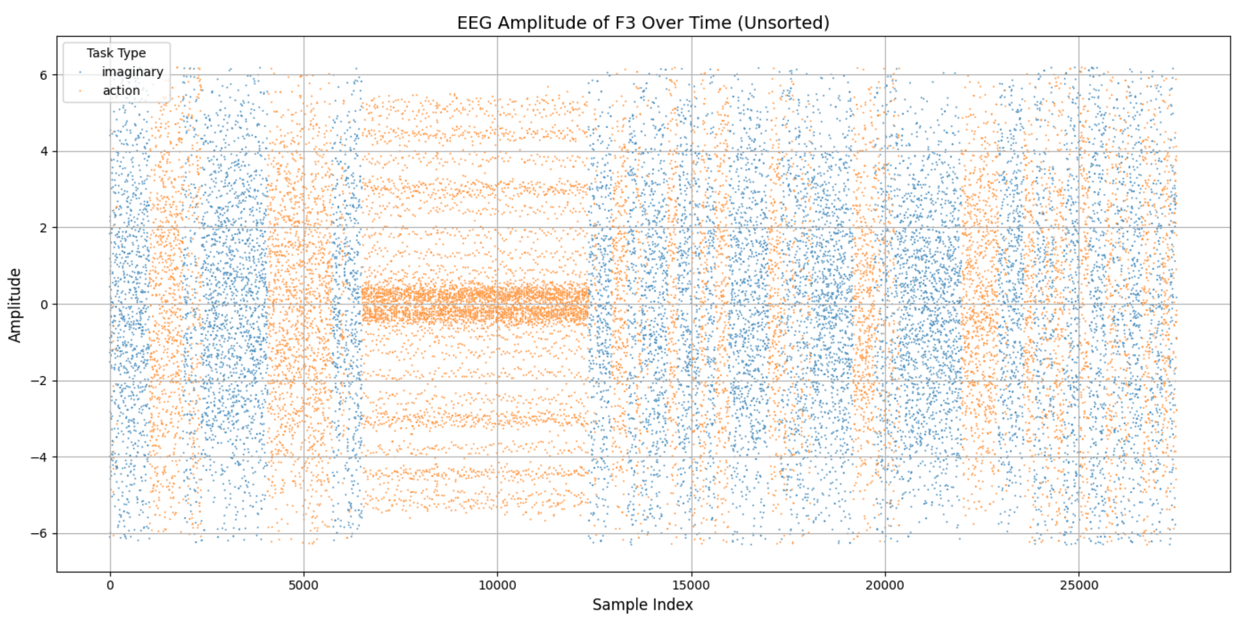
Statistical Summary & Visualization:

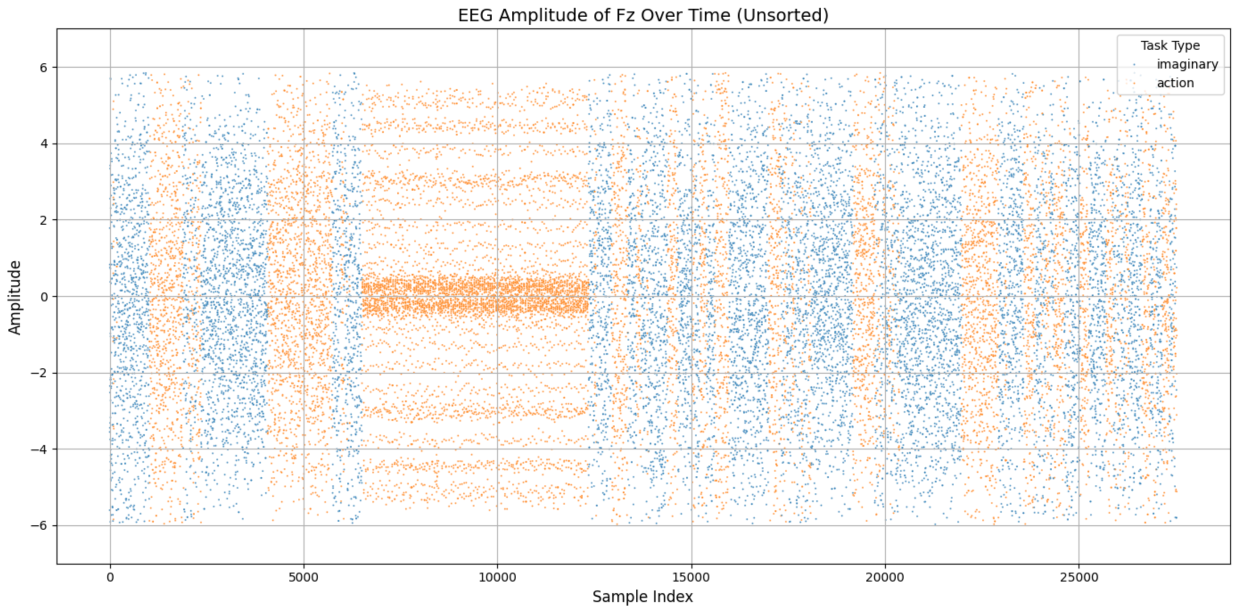
Recalculated descriptive statistics of the columns post-cleaning.Confirmed consistent values

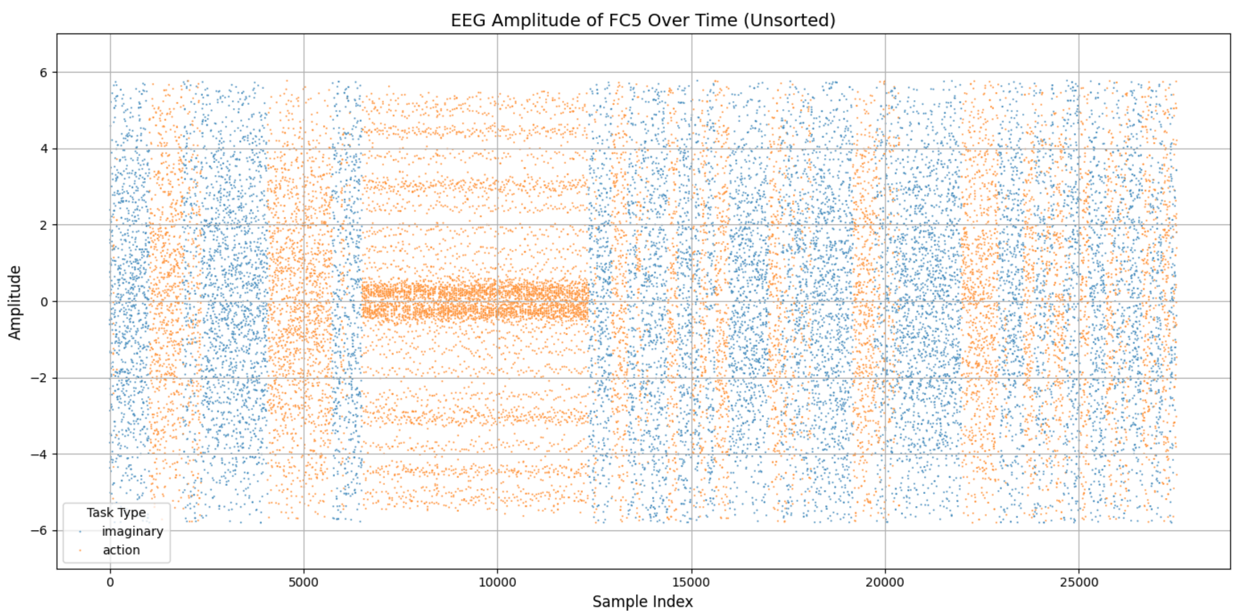
Used:

Histograms: To explore distribution shapes.


Histogram of F4
Histogram of FC6
Histogram of C5
Histogram of C3
Histogram of Cz
Histogram of C4
Histogram of C6
Histogram of CP5
Histogram of CP3
Histogram of CP1
Histogram of CP2
Histogram of CP4
Histogram of CP6
Frequency
CP2
CP4

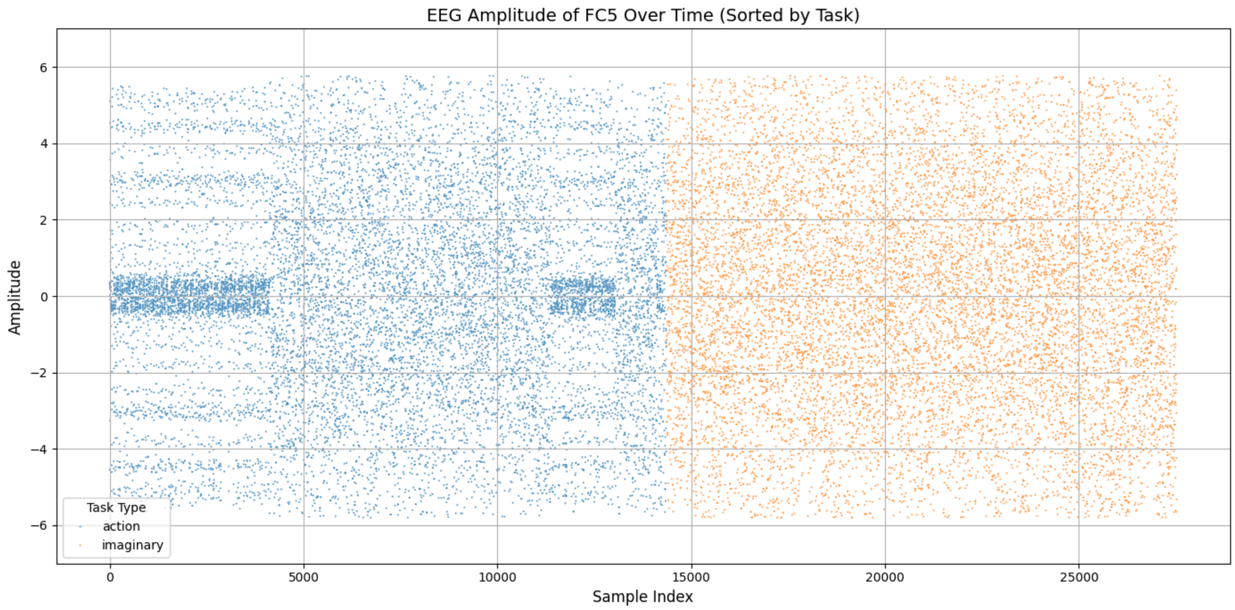

EEG Amplitude of F3 Over Time (Unsorted)
Task Type
imaginary
action
Amplitude
Sample Index
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000

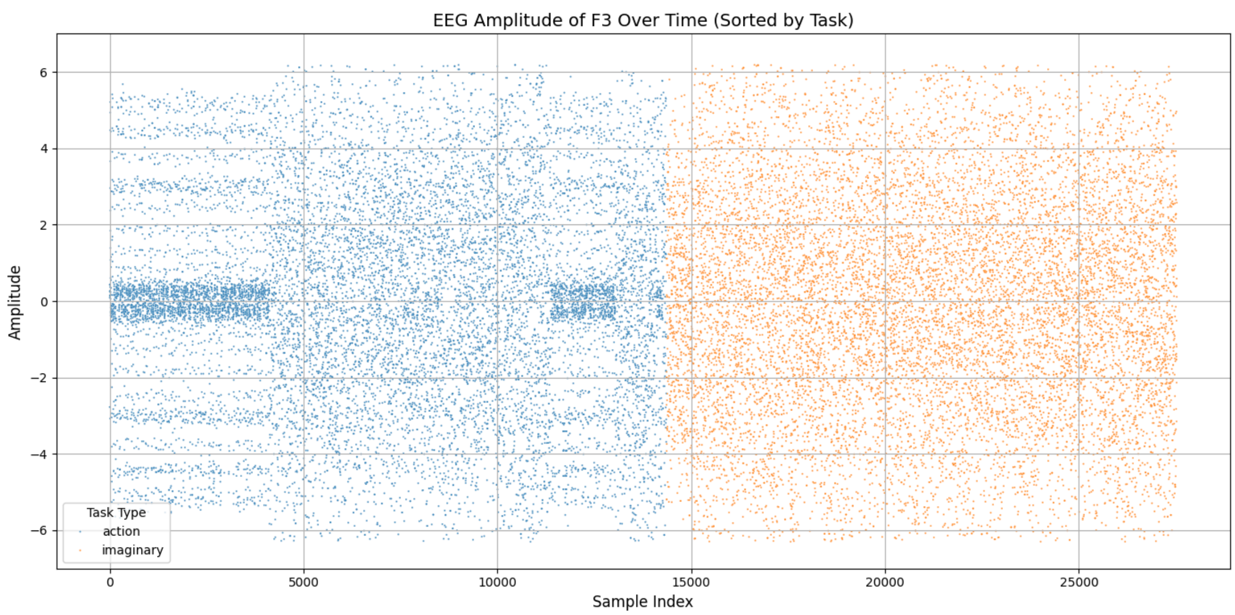

EEG Amplitude of Fz Over Time (Unsorted)
Task Type
imaginary
action
Amplitude
Sample Index
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000

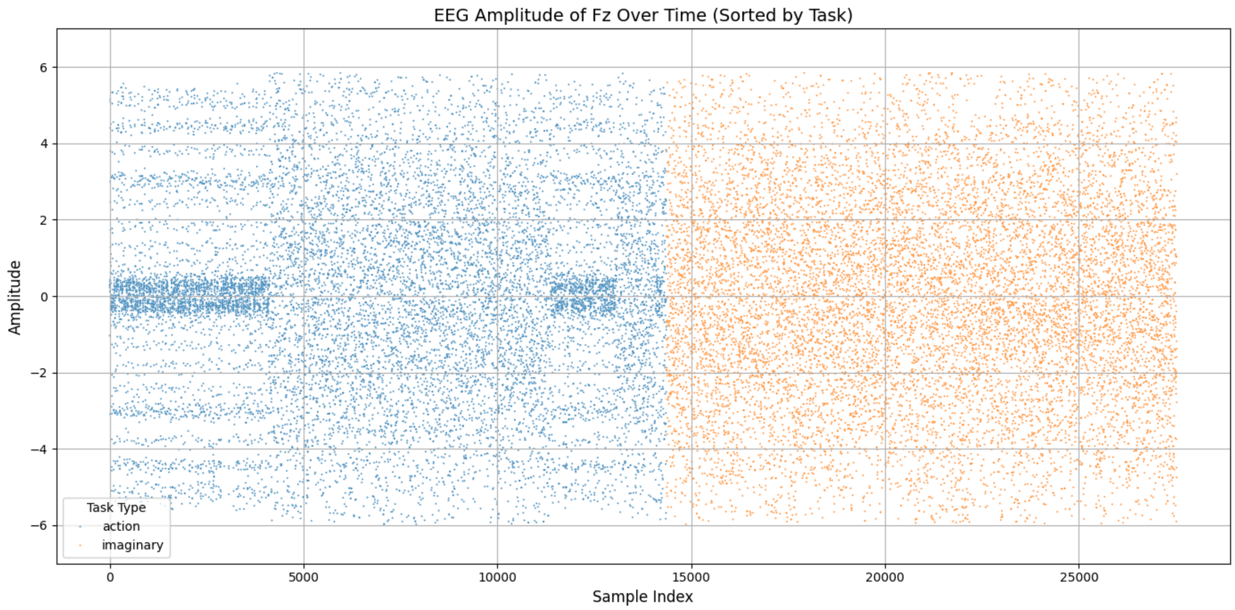

EEG Amplitude of FC5 Over Time (Unsorted)
Task Type
imaginary
action
Amplitude
Sample Index
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000

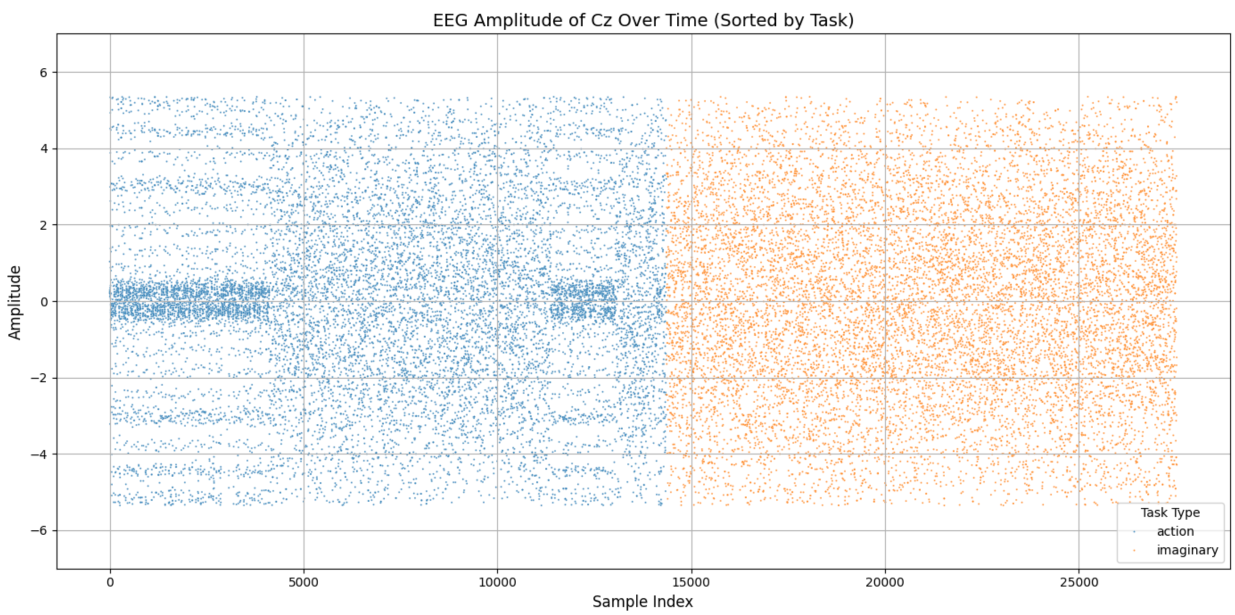
Scatter plots: Saw alternating patterns — possibly cyclic or task-dependent behaviors.

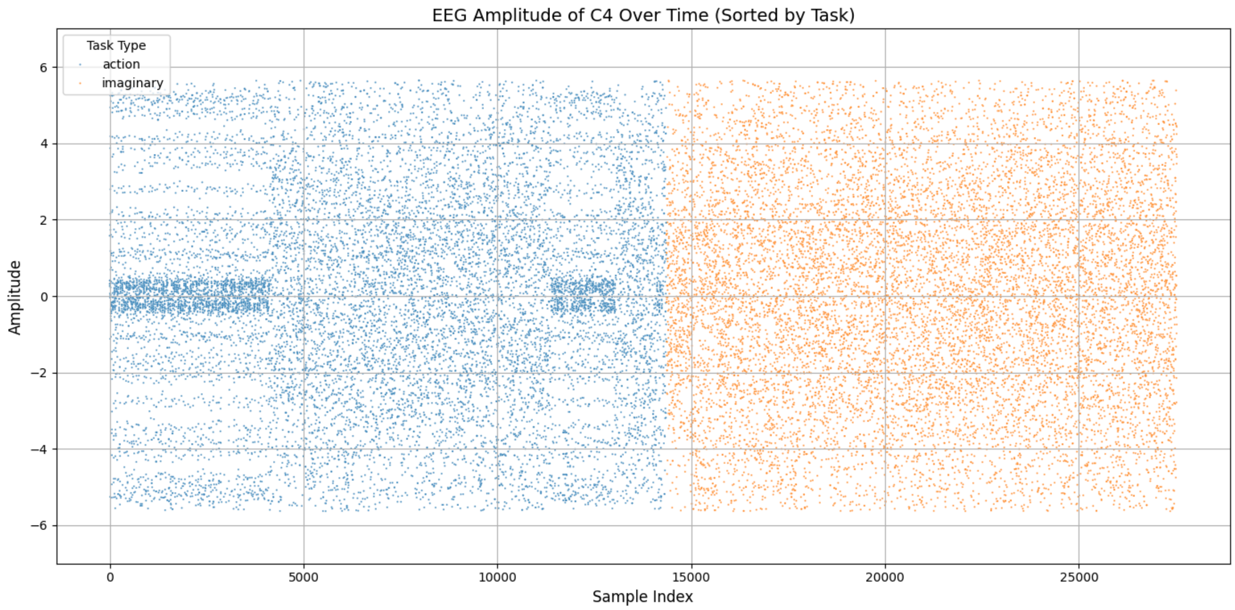
Sorted based on Action Type — possibly to understand sequential behavior or prepare for modeling.

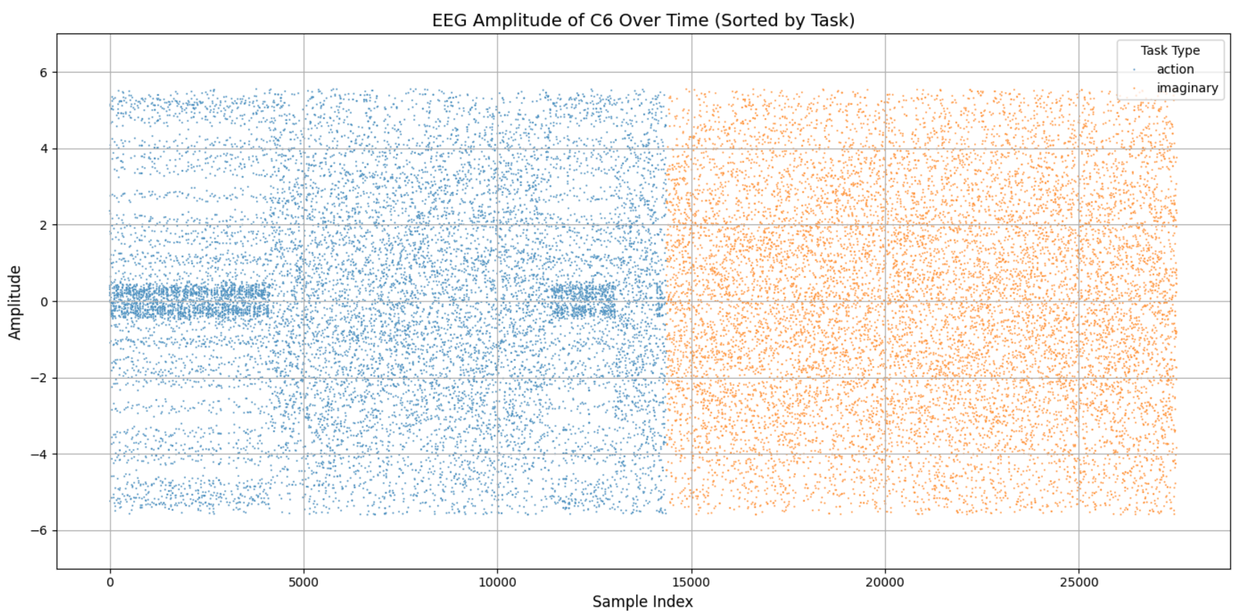

EEG Amplitude of FC5 Over Time (Sorted by Task)
Amplitude
Sample Index
Task Type
action
imaginary
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000

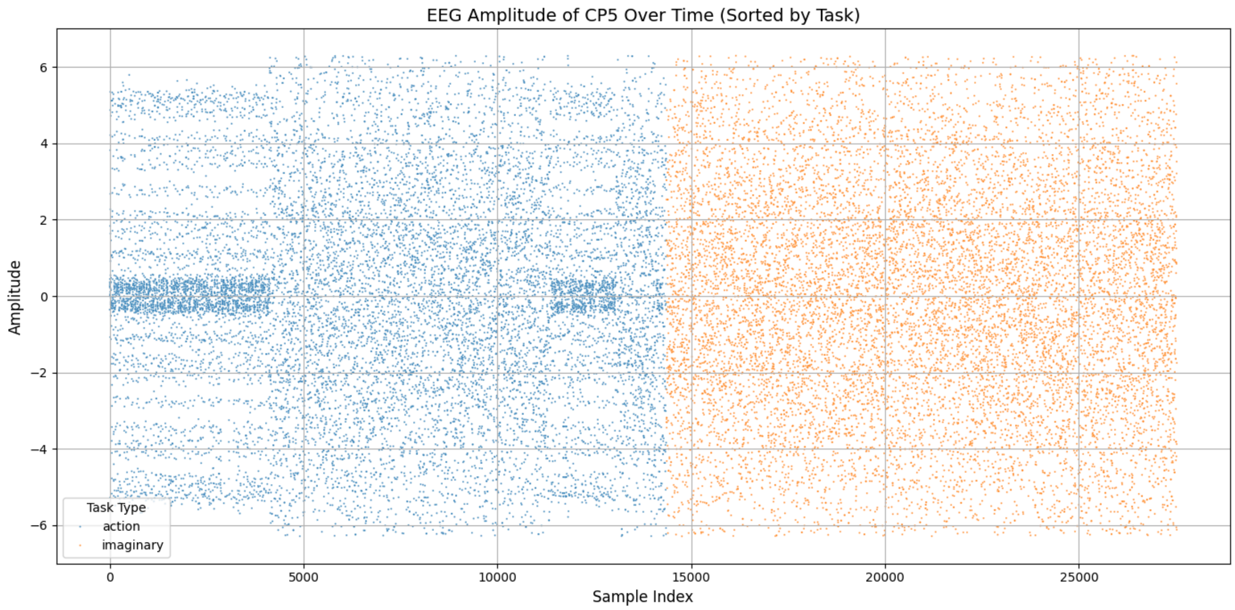

EEG Amplitude of F3 Over Time (Sorted by Task)
Amplitude
Sample Index
Task Type
action
imaginary
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000

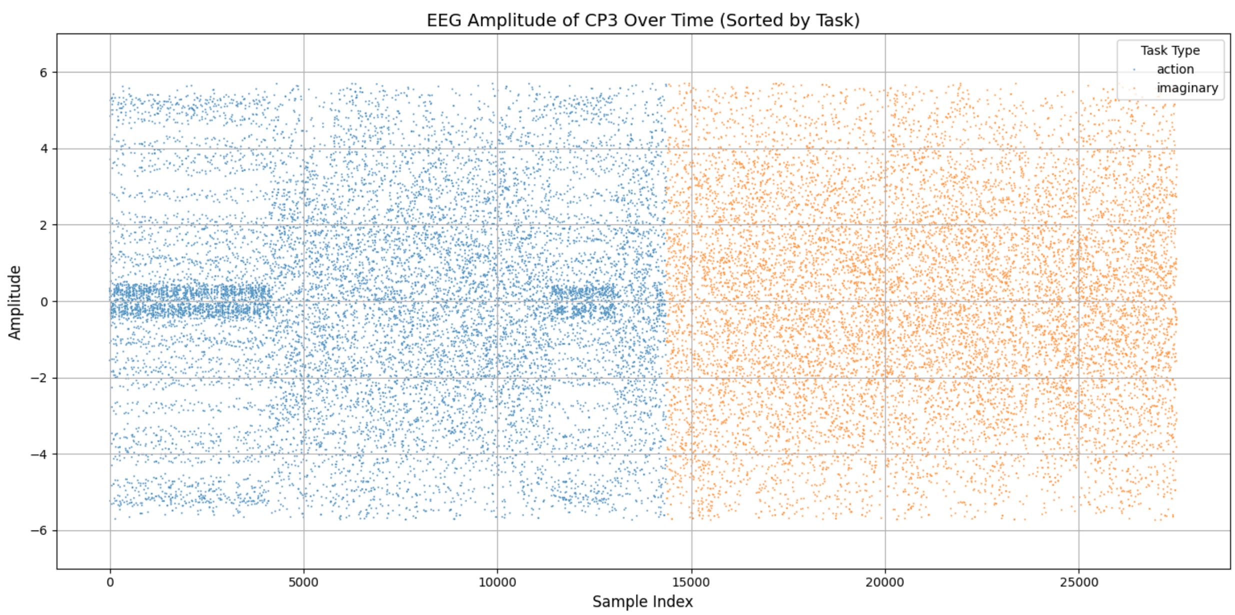

EEG Amplitude of Fz Over Time (Sorted by Task)
Amplitude
Sample Index
Task Type
action
imaginary
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000

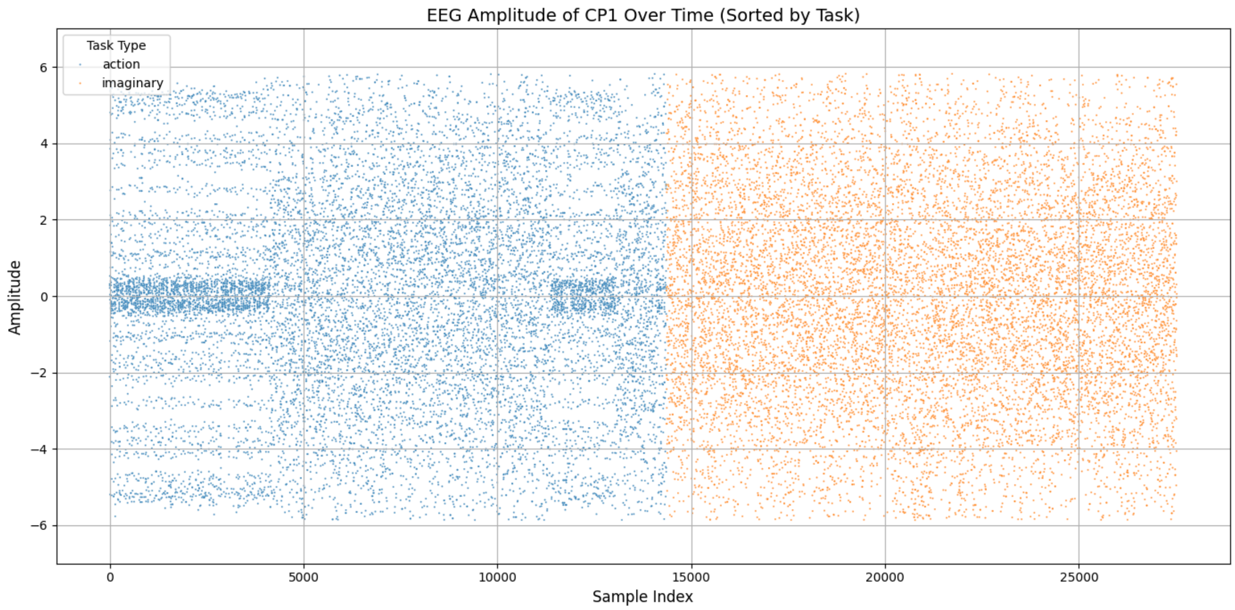

EEG Amplitude of Cz Over Time (Sorted by Task)
Amplitude
Sample Index
Task Type
action
imaginary
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000

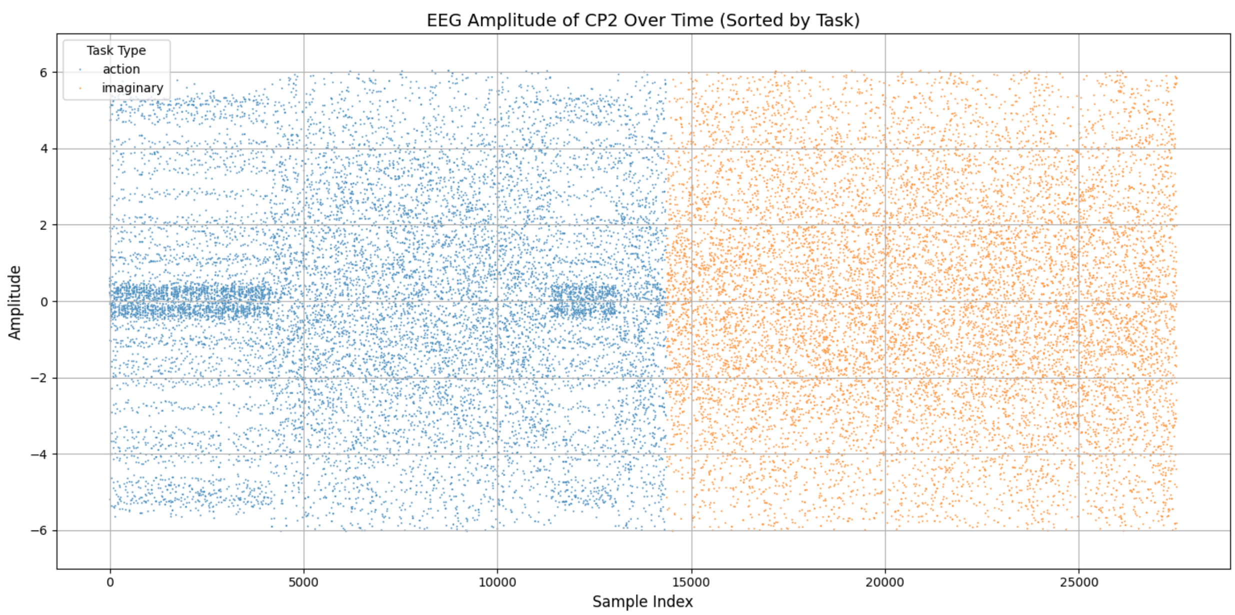

EEG Amplitude of C4 Over Time (Sorted by Task)
Task Type
action
imaginary
Amplitude
Sample Index
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000

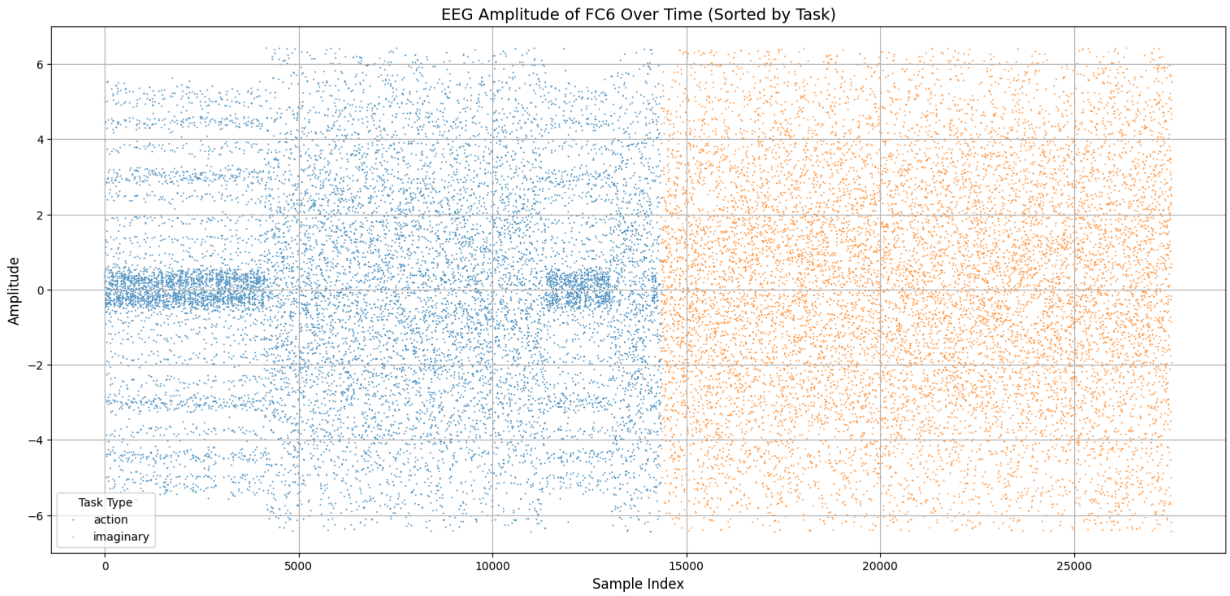

EEG Amplitude of C6 Over Time (Sorted by Task)
Task Type
action
imaginary
Amplitude
Sample Index
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000

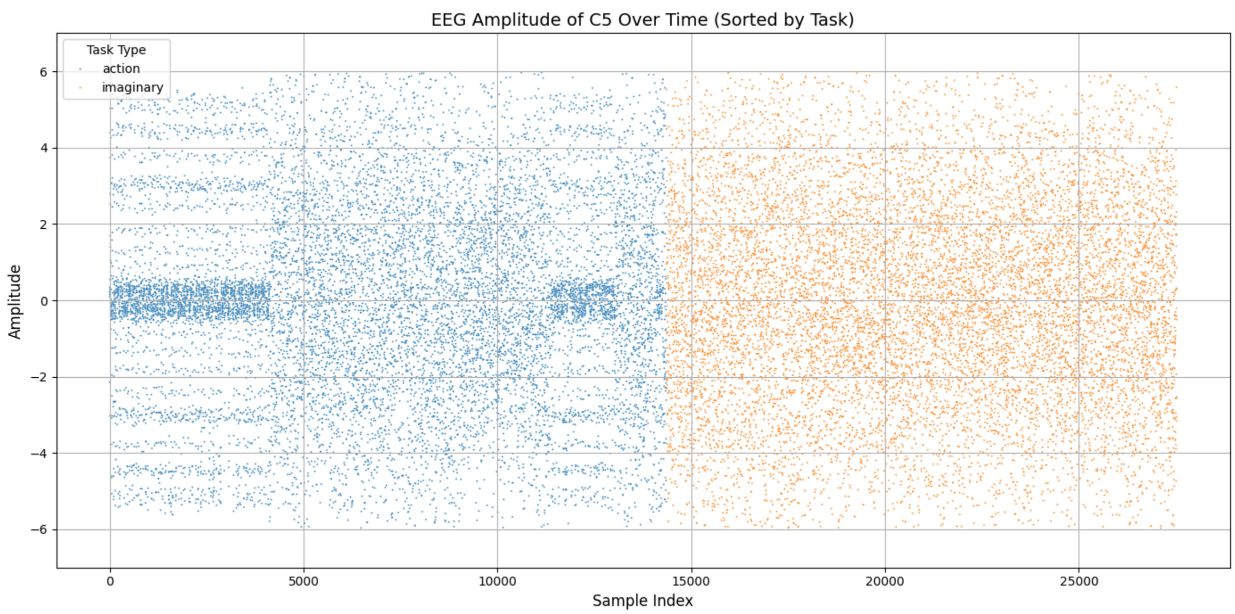

EEG Amplitude of CP5 Over Time (Sorted by Task)
Amplitude
Sample Index
Task Type
action
imaginary
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000

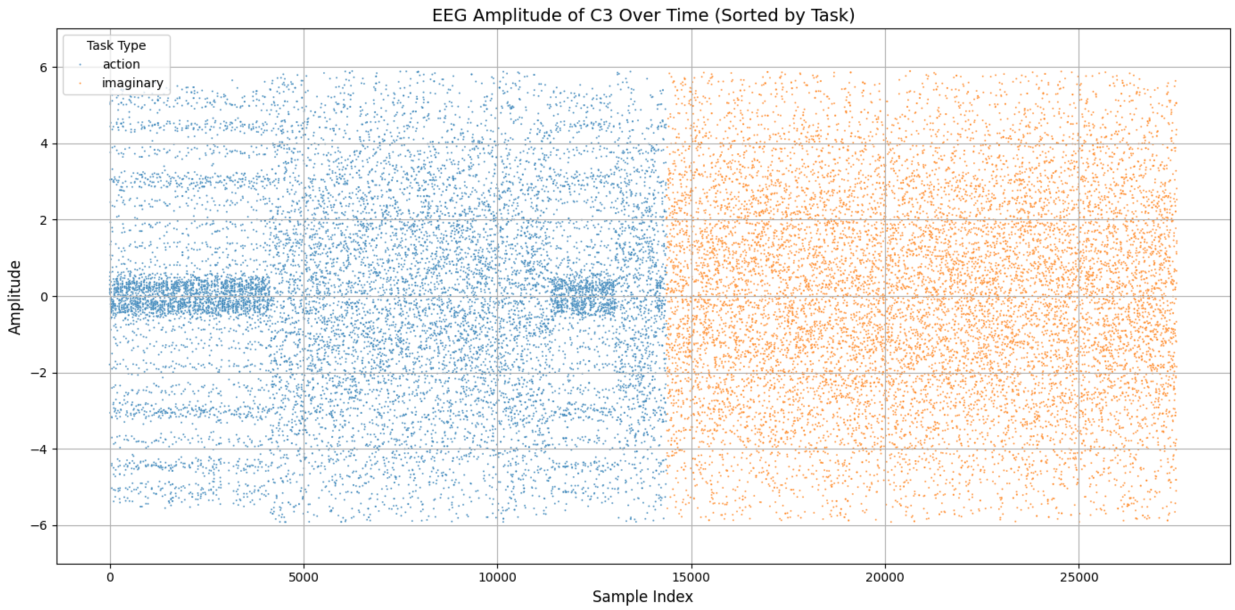

EEG Amplitude of CP3 Over Time (Sorted by Task)
Amplitude
Sample Index
Task Type
action
imaginary
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000

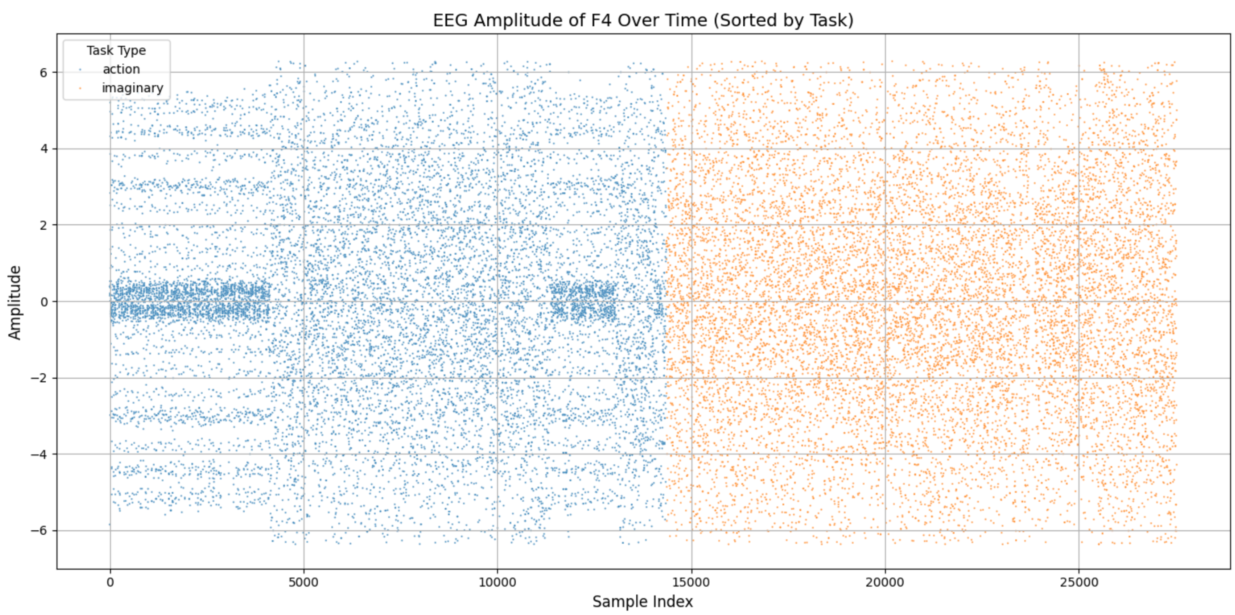

EEG Amplitude of CP1 Over Time (Sorted by Task)
Task Type
action
imaginary
Amplitude
Sample Index
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000


EEG Amplitude of CP2 Over Time (Sorted by Task)
Task Type
action
imaginary
Amplitude
Sample Index
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000


EEG Amplitude of FC6 Over Time (Sorted by Task)
Amplitude
Sample Index
Task Type
action
imaginary
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000


EEG Amplitude of C5 Over Time (Sorted by Task)
Amplitude
Sample Index
Task Type
action
imaginary
6
4
2
0
−2
−4
−6
0
5000
10000
15000
20000
25000


EEG Amplitude of C3 Over Time (Sorted by Task)
Task Type
action
imaginary
Amplitude
Sample Index
0
5000
10000
15000
20000
25000
6
4
2
0
−2
−4
−6


EEG Amplitude of F4 Over Time (Sorted by Task)
Task Type
action
imaginary
Amplitude
Sample Index
0
5000
10000
15000
20000
25000
6
4
2
0
−2
−4
−6

For Base Classifier

Modeling –
Load the EEG dataset from a CSV file called "eeg_sorted_by_task.csv" into a pandas DataFrame.

Shuffle the dataset randomly to remove any order bias, using a fixed random seed (random_state=42) for reproducibility.

Split features and labels:
X: the first 16 columns as EEG channel data
y: the "task_type" column, which represents the target labels i.e imaginary and action
(Intentionally dropped subject id column to avoid overfitting)

Encode the labels (y) into numeric values using LabelEncoder so they can be used by the classifier.

```
le = LabelEncoder()
y_encoded = le.fit_transform(y)
```

Split the data into training and testing sets:
80% for training, 20% for testing.
Stratified splitting ensures that each class is proportionally represented in both training and testing sets.

Train a Linear Discriminant Analysis (LDA) classifier using the training data (X_train, y_train).
Make predictions on the test set (X_test) using the trained LDA model

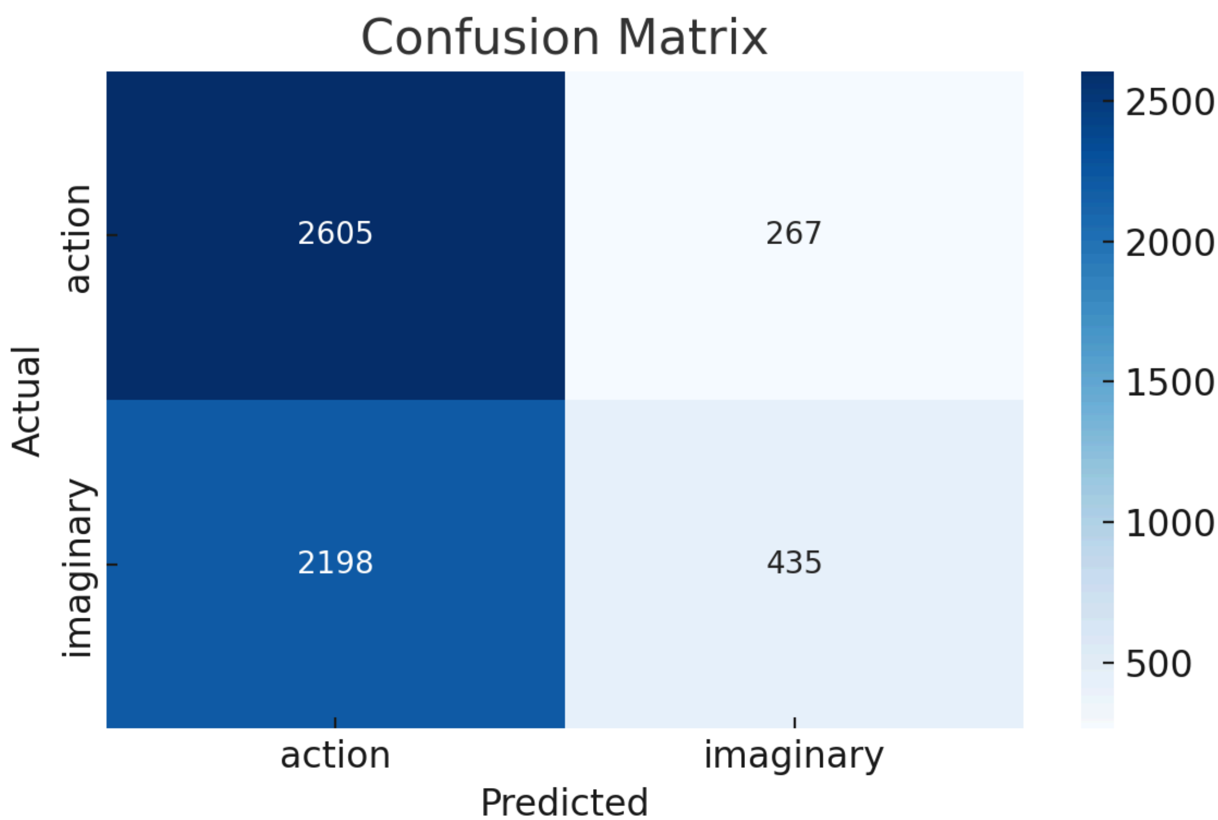
Results

Evaluated with:
To make sure that I had enough data,I made changes in the training and test size
Accuracy-with changes in training and test size variation
[LDA Accuracy for training 0.7 and test size 0.3--0.5500
LDA Accuracy for training 0.8 and test size 0.2--0.5522
LDA Accuracy for training 0.9 and test size 0.1--0.5481]

After changing train/test splits — little accuracy change.

	precision	recall	f1-score
action	0.54	0.91	0.68
Imagery	0.62	0.17	0.26

Confusion Matrix



Normalization with Train/Test Variations:

Tried with normalized dataset along with train and test size variations,there was no change to the accuracy,precision and f1 score.

For Advanced Classifier

Generated feature median,min and max,and then checked the statistics again

	count	mean	std	min	25%	50%	75%	max
FC5	27525	-0.01734	2.732473	-5.80213	-1.98949	-0.0102	1.948798	5.772496
F3	27525	-0.06697	2.764306	-6.3011	-2.03453	-0.04406	1.868423	6.183946
Fz	27525	-0.05576	2.637821	-5.9757	-1.90068	-0.04744	1.779793	5.841752
F4	27525	-0.01143	2.833461	-6.367	-1.97816	-0.00176	1.976922	6.278633
FC6	27525	-0.02405	2.902855	-6.44519	-2.11075	-0.0203	2.028628	6.419731
C5	27525	-0.04585	2.571867	-5.96775	-1.8197	-0.04959	1.733263	5.968126
C3	27525	-0.02497	2.633964	-5.91505	-1.87278	-0.01637	1.810051	5.888282
Cz	27525	-0.01871	2.500444	-5.35083	-1.78541	-0.0066	1.749215	5.349948
C4	27525	0.012272	2.757749	-5.63132	-1.95572	0.019896	1.966904	5.644806
C6	27525	-0.0191	2.783912	-5.59257	-2.0603	-0.02064	2.013624	5.550914
CP5	27525	-0.05892	2.905007	-6.28849	-2.1075	-0.03683	1.950291	6.29291
CP3	27525	-0.03338	2.638134	-5.73223	-1.85957	-2.18E-22	1.77721	5.698298
CP1	27525	-0.02543	2.695197	-5.86435	-1.90014	-0.00467	1.841827	5.819343
CP2	27525	-0.0321	2.78406	-6.02433	-1.96321	-0.0499	1.888904	6.032268
CP4	27525	-0.068	2.82123	-5.94411	-2.11051	-0.06855	1.932492	5.93114

CP6	27525	0.024267	3.001047	-6.10216	-2.15573	0.01008	2.213329	6.130176
median_value	27525	-0.02691	1.743369	-5.536	-1.02044	-0.0098	0.988719	5.523323
min_value	27525	-3.52179	2.313333	-6.44519	-5.25306	-4.28463	-2.46737	5.297291
max_value	27525	3.461603	2.335739	-5.35043	2.35239	4.23529	5.222165	6.419731

Dataset Splitting
The dataset was partitioned into three subsets:
Training set: 60%
Validation set: 20%
Test set: 20%

Feature Selection using Forward-Backward Selection
Base Model: Random Forest Classifier
Evaluation Metric: Class-wise Average Accuracy
Initial Feature Set: ['max_value', 'Cz', 'Fz', 'CP6', 'FC6', 'CP5']
The selection process involved:
Iteratively adding or removing features.

Evaluating performance on the validation set.
Choosing the feature set that resulted in the highest improvement in class-wise average accuracy.

Based on best feature set found:
Feature set achieving the best validation performance with Class-wise Average Accuracy:0.71395

Hyperparameter Tuning
Method: Discrete Exhaustive Search with Cross-Validation
Objective: Optimize model hyperparameters for the selected feature set
Each combination was evaluated using cross-validation to ensure generalizability and stability.

Evaluated on test data with two different sets of parameter and the results displayed in a comparison manner

Metric	First Run	Second Run	Observation
Parameter Grid Size	324 candidates, 972 fits	432 candidates, 1296 fits	Second run searched more thoroughly
Best Parameters	`n_estimators=50`	`n_estimators=100`	More estimators gave a slight boost
	`max_depth=15`, `max_features='sqrt'`	Same	
	`min_samples_split=2`, `min_samples_leaf=1`	Same	
Best CV Classwise Accuracy	**0.7108**	**0.7156**	Small improvement (+0.5%)
Test Classwise Accuracy	**0.7059**	**0.7093**	Slight improvement (+0.34%)

Class	Metric	First Run	Second Run	Interpretation
Action	**Precision**	**0.87**	**0.89**	**Slight improvement in predicting it correctly**
	Recall	**0.49**	**0.48**	**No improvement — still misses over half**
	F1-Score	**0.63**	**0.63**	**No gain overall**
Imagery	**Precision**	**0.62**	**0.62**	**Stable**
	Recall	**0.92**	**0.94**	**Slightly better**
	F1-Score	**0.74**	**0.75**	**Slightly better**

Model performance:

- **~71% classwise accuracy on CV and test in both cases.**
- **Test accuracy stays around 70%.**

The bias toward the 'Imagery' class remains unchanged:

- **"Imagery" class recall is very high (> 0.92), even increasing.**
- **"Action" class recall is consistently low (~0.48–0.49), indicating it is under-represented or harder to detect.**

The small improvements in the second run (from more trees and a larger grid search) yield incremental performance gains, not major breakthroughs.

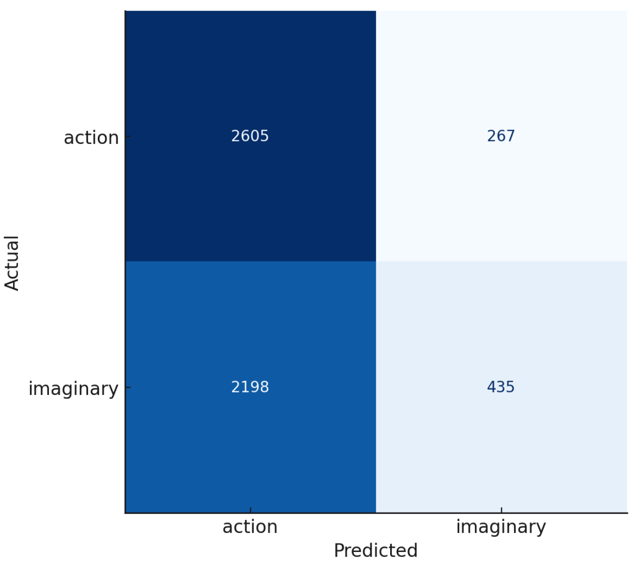
The results of comparison between LDA and Rf with the best hyperparameters(scope of the project)

	precision	recall	f1score	—(RF)
action	0.89	0.48	0.62	
Imagery	0.62	0.94	0.75	

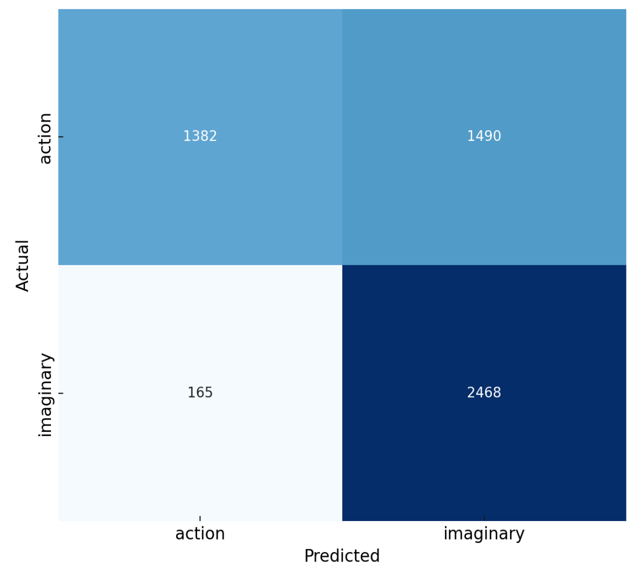
action	0.54	0.91	0.68	—(LDA)
Imagery	0.62	0.17	0.27	

Confusion matrix for LDA and Random Forest respectively

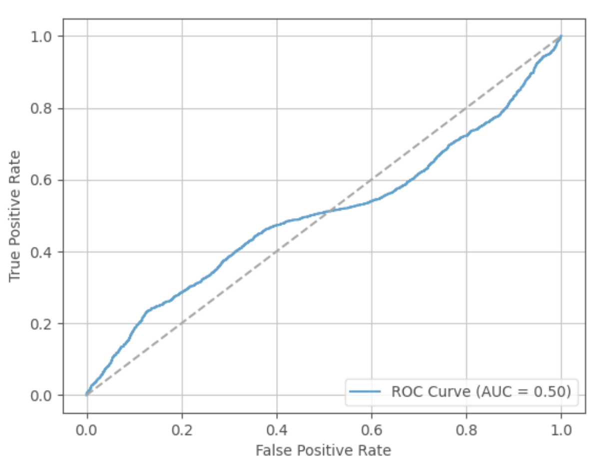
LDA



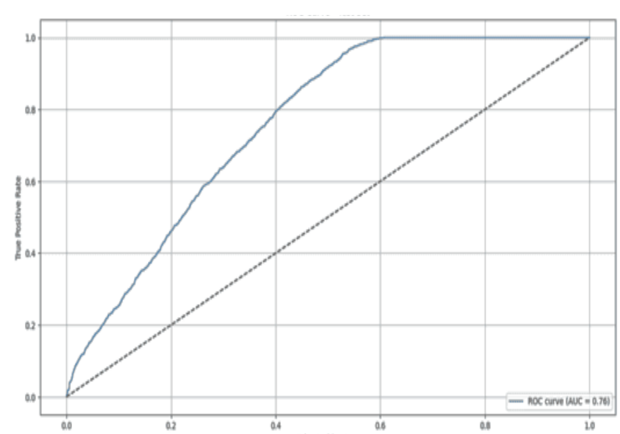
RF



LDA



Random Forest Classifier



Conclusion: For predicting task types from our EEG dataset, the best model is a Random Forest classifier with the following parameter and hyperparameter settings:

- `n_estimators = 100`
- `max_depth = 15`
- `max_features = 'sqrt'`
- `min_samples_split = 2`
- `min_samples_leaf = 1`

This model was trained using the selected EEG feature set (16 channels) and achieved optimal performance in classifying mental tasks.

Discussion

Limitations:
Analyzing and filtering EEG data typically requires domain-specific knowledge, including techniques such as bandpass filtering, artifact removal (e.g., eye blinks or muscle noise), and feature extraction (e.g., power spectral density, frequency bands). These steps were not deeply explored in this project and may impact model performance.
Lack of GPU Acceleration:
GPU resources were not utilized in this project due to limitations in the libraries used. Leveraging GPU acceleration could have enabled more parallel computation, potentially allowing the use of more complex models or faster training — particularly beneficial for large EEG datasets.
No Evaluation with Neural Networks:
This study did not include neural networks, which are commonly applied in EEG classification tasks due to their ability to learn complex temporal and spatial features. Future work could explore CNNs, RNNs, or transformers for potentially improved performance.

Learning(Non-technical):
Bias Understanding:
Initially, I assumed the dataset was free of time-series characteristics because the data appeared in a discretized format, without any explicit frequency-dependent columns or timestamps. Based on that, I treated it as a typical tabular classification task.

However, after working more deeply with the data, I came to realize that this assumption was flawed. EEG signals are inherently time-series in nature, even if presented in a simplified or summarized form. The underlying brain activity involves temporal dynamics, which may still be reflected in patterns across samples
However,it can be dealt with taking the actual spikes within the data,but it requires domain knowledge.

Clarity of what to do and more specifically what not to do
Always Check Dataset Statistics First
Understanding basic statistics (mean, variance, class distribution, missing values) provides essential insight before modeling.

Randomize the Data
Shuffling the dataset ensures that the model doesn't learn from artificial ordering, which can lead to misleading performance.

Remove Identifiers That Cause Overfitting
Features like subject ID can cause the model to "cheat" by learning subject-specific patterns instead of generalizable EEG features. These must be removed.

Be Cautious with Outliers
Unless you are absolutely certain an outlier is an error, it's often best to leave it in — removing legitimate outliers can introduce bias.

Adjust Train/Test Size Thoughtfully
Changing the ratio of train to test can help ensure the model has enough data to learn, especially in smaller datasets.

Visualizing Data is Invaluable
Good plots(dependent on data) eg. scatter plotting box plot,histograms etc can reveal issues and insights much earlier than raw numbers alone.

Use Separate Validation and Test Sets
In real-world scenarios, having a dedicated validation set and a test set (for final evaluation) is essential for reliable results

Emotional Involvement:
Personally, I felt frustrated and discouraged at times during the classification phase. Despite applying multiple techniques, exploring different preprocessing steps, and even trying variations of the dataset, I saw little to no improvement in performance. It was disheartening to put in consistent effort without immediate results.However,recognizing when a method isn't working and having the mindset to pivot is often a sign of growth, not failure. As said by Prof Markus Mayer,

"*Sometimes the "success" is us noticing that we have to try another approach*".-Prof. Dr.-Ing. Markus Mayer

Note: I've added the GitHub link in case anyone would like to contribute or share suggestions – https://github.com/aiaI127/EEG_classification

Disclaimer: The terms *imagery* and *imaginary* have been used interchangeably throughout the project.