NEUROCHECK

THE BIG DATA PROJECT PRESENTATION

MOHAMMED FURQAAN KHAN - 1319132

BAHADURI PRACHITI JAGDISH - 1317686

TANVI DIPAN PATEL 1316221

PRATIKSHA GURUDEV KANDE-1309129

PANKTI BHATT - 1287645

INTRODUCTION

The Brainwave-Powered Student State Detector is a project that aims to accurately detect the brain states of students during learning tasks. Here are some key points about the project:

- **Purpose**: The project aims to develop a model that accurately detects the brain states of students based on their levels of relaxation, focus, and neutrality during learning tasks.

- **Techniques:** The model will be developed using big data analytics techniques such as machine learning algorithms and data visualization tools.

- **Data:** The model will be trained on a large dataset of brainwave data collected from students in different learning environments.

- **Benefits**: The NeuroCheck project has the potential to revolutionize the field of education by providing real-time feedback to teachers and students about the cognitive state of learners. This feedback can be used to optimize learning environments and improve learning outcomes.

- **Future implications:** The project may also have implications beyond education, including in fields such as healthcare, where brainwave data can be used for diagnosis and treatment of various conditions.

WHY WE THINK IT'S IMPORTANT

- **Improved understanding of student emotions**: By analyzing brainwave data, you can gain insights into the brain states associated with different emotions. This can help educators and students better understand the relationship between emotions and cognitive function, which can in turn improve student well-being and academic performance.

- **Personalized learning:** By detecting the brain states of individual students, it may be possible to personalize learning experiences based on their unique needs. For example, if a student is in a relaxed state, it may be more effective to teach them new material, while if they are in a more focused state, it may be more effective to have them engage in problem-solving activities.

- **Early detection of learning difficulties:** Certain brainwave patterns may be associated with learning difficulties or cognitive disorders. By analyzing brainwave data, educators may be able to detect these issues early on and provide targeted interventions to support struggling students.

- **Advancements in neurotechnology:** The field of neurotechnology is rapidly advancing, and by working with brainwave data, you can contribute to the development of new technologies that can help individuals better understand and regulate their own brain activity. This has potential applications in a variety of fields, including mental health, education, and sports performance.

OUR DATA

- We used a dataset from kaggle named: **EEG brainwave dataset: mental state**.
- The data was collected **from four people (2 male, 2 female) for 60 seconds per state - relaxed, concentrating, neutral.**
- It used a **Muse EEG headband** which recorded the TP9, AF7, AF8 and TP10 EEG placements via dry electrodes.

OUR OBJECTIVE

- **Techniques:** The model will be designed using big data analytics techniques such as machine learning algorithms and data visualization tools.

- **Data:** The model will be trained on a large dataset of brainwave data collected from people in various environments.

- **Primary objective:** The project aims to provide educators with a tool to better understand and respond to individual student needs, improve teaching strategies, and enhance learning outcomes.

- **Real-time feedback:** The model's real-time feedback on student cognitive and emotional states can help educators tailor their teaching strategies to better meet the needs of individual students.

- **Contributing to the understanding of the relationship between brain states and learning:** Another objective of the project is to contribute to the understanding of the relationship between brain states and learning. The insights gained from the data collected and analyzed through the model could lead to new discoveries about how to optimize learning environments and practices.

- **Mental Health :** The project could also have implications for mental health and understanding of brain states beyond the context of education. The accurate detection of brain states using the model may contribute to the development of tools and strategies for the diagnosis, treatment, and management of mental health conditions that are associated with specific brainwave patterns

UNDERSTANDING THE DATA

- **Dataset :** (2479,989)
- **After variance and correlation technique:** (2479,374)
- **After z-score:** (819,100)
- **After PCA :** (819,20)

The high-dimensional datasets, why they are challenging to analyze, and the importance of feature selection and Dimensionality reduction techniques in machine learning?

- As the number of features increases, the likelihood of finding spurious correlations or noise in the data also increases. This can lead to overfitting.
- High-dimensional datasets are computationally expensive to analyze.

PREPROCESSING, FEATURE SELECTION AND DIMENSIONALITY REDUCTION

feature selection and dimensionality reduction techniques are essential for analyzing high-dimensional datasets, improving classification performance, and gaining insights into the underlying patterns and mechanisms of the data.

What we used:

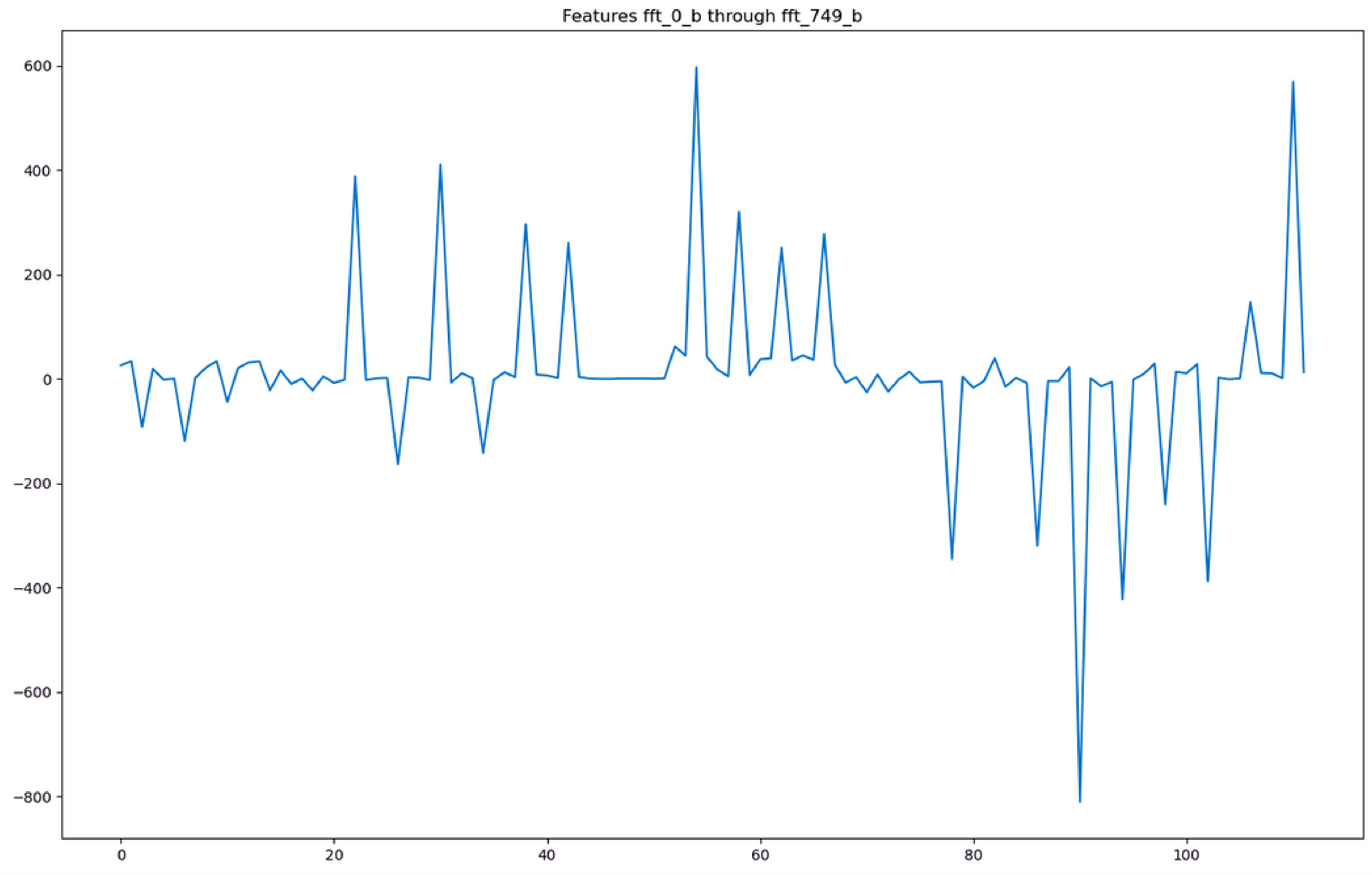
- **Variance-based feature selection:** This technique is useful for identifying features with low variability, which may not be informative for classification. Variance-based feature selection is a simple and efficient technique that does not require any parameter tuning.

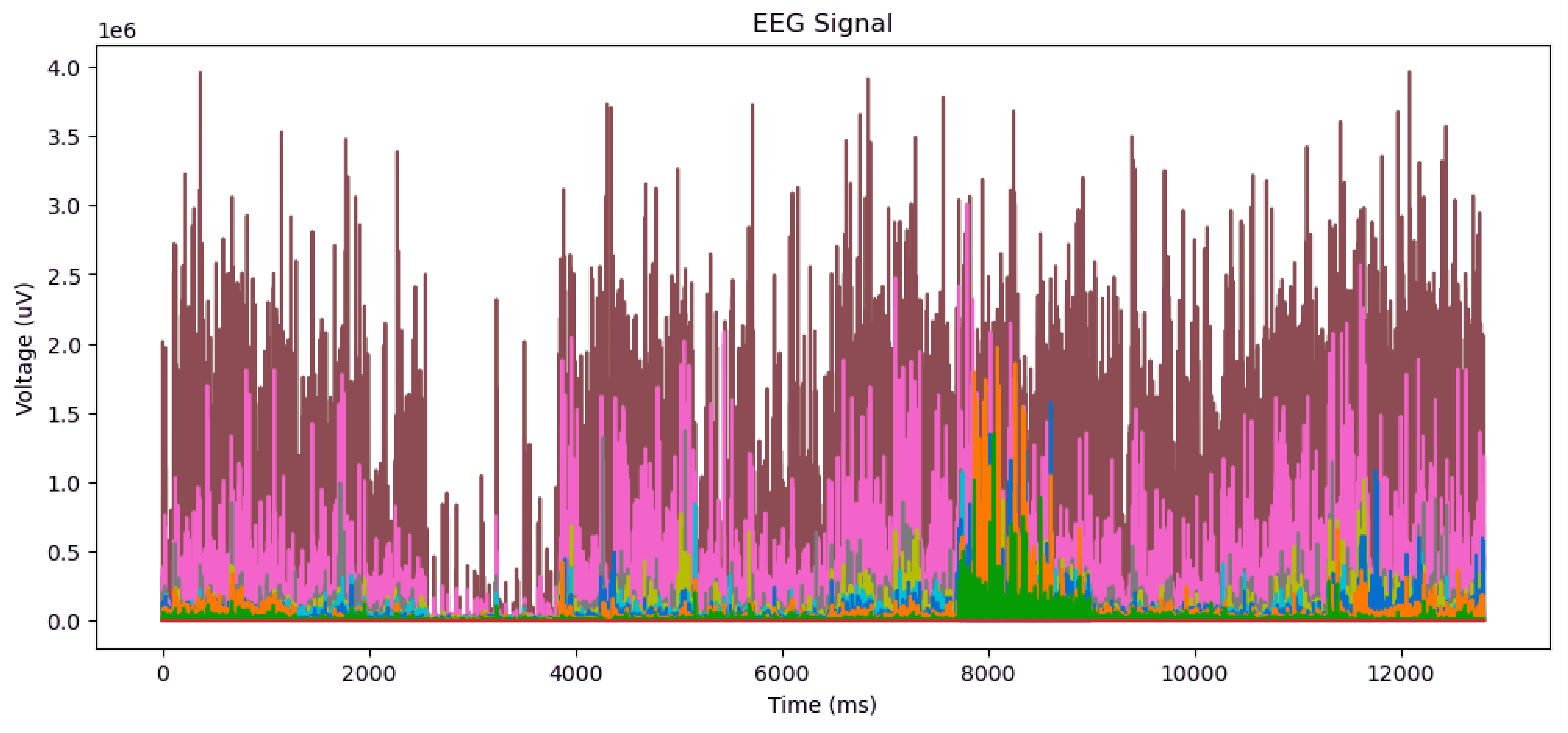
- **Correlation-based feature selection:** This technique is useful for identifying pairs of highly correlated features, which can be redundant and provide no additional information for classification. This is a widely used technique that can handle linear dependencies between features

- **Z-score normalization:** This technique allows us to compare the features on the same scale and reduce the impact of outliers or extreme values.

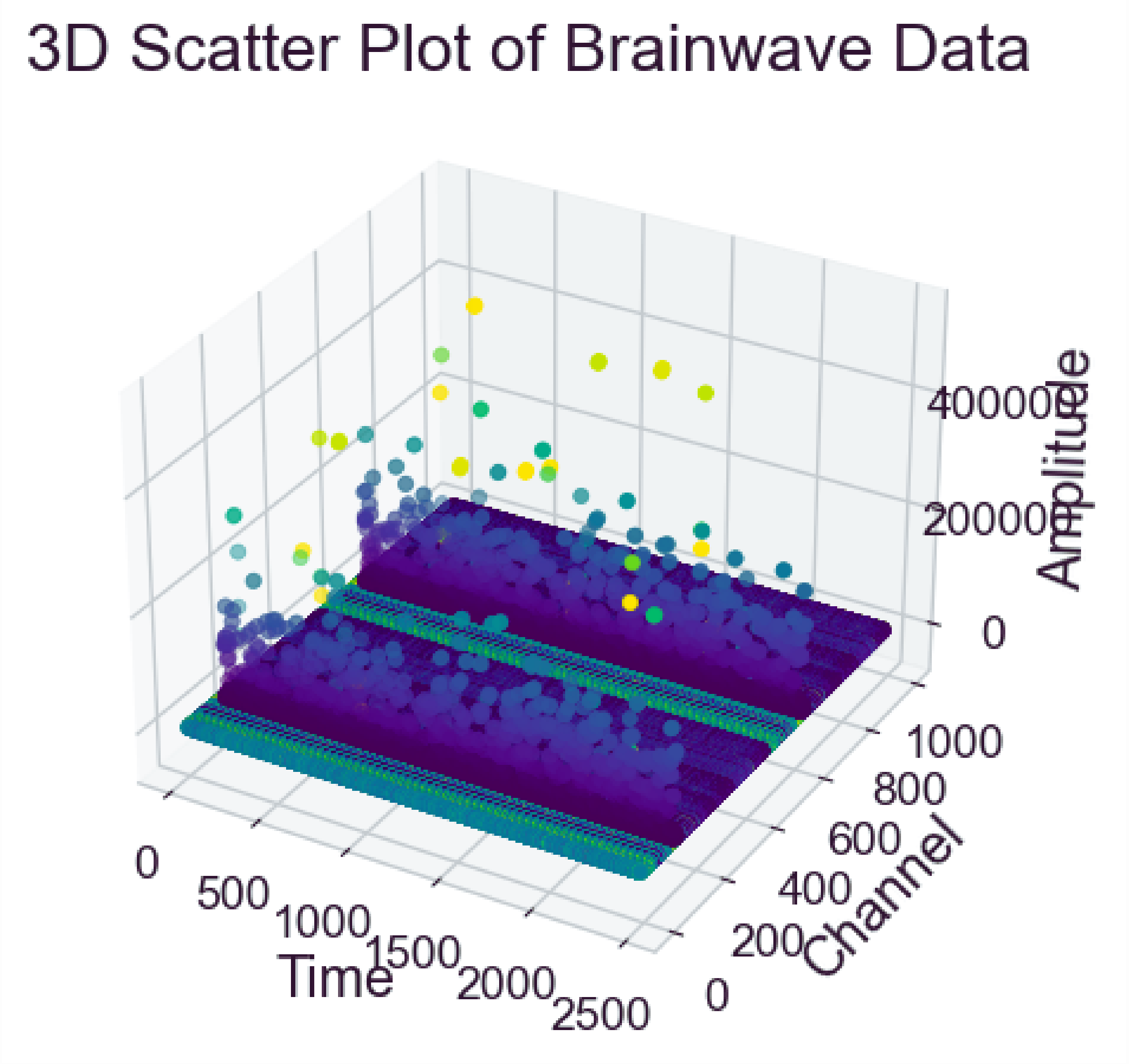
- **PCA-based feature selection:** This selection technique is useful for reducing the dimensionality of the data by projecting it onto a lower-dimensional space defined by the principal components.

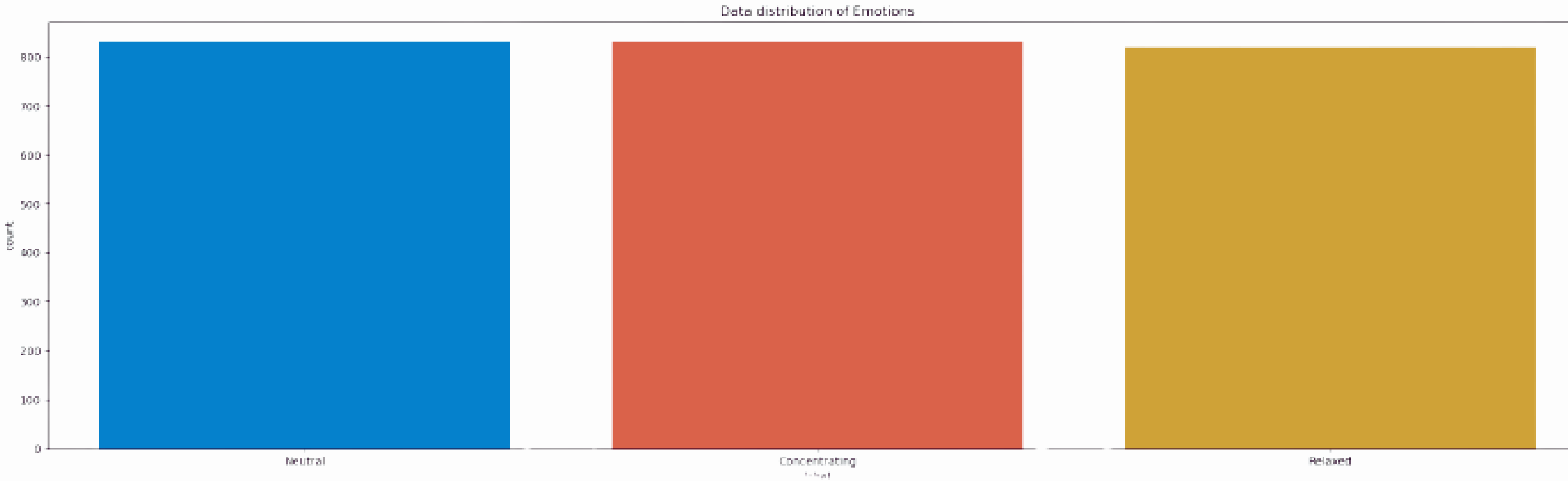
VISUALIZING OUR DATA



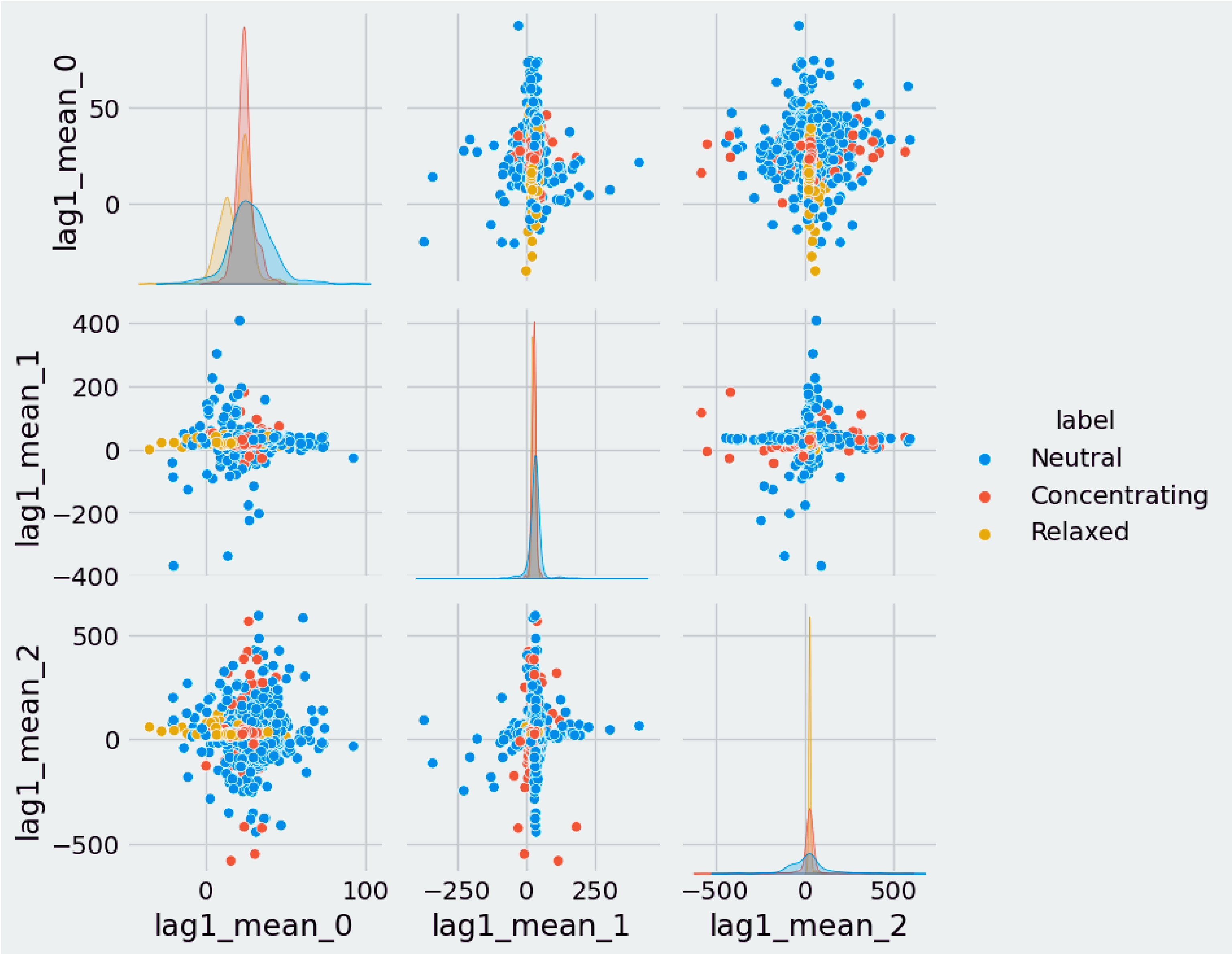


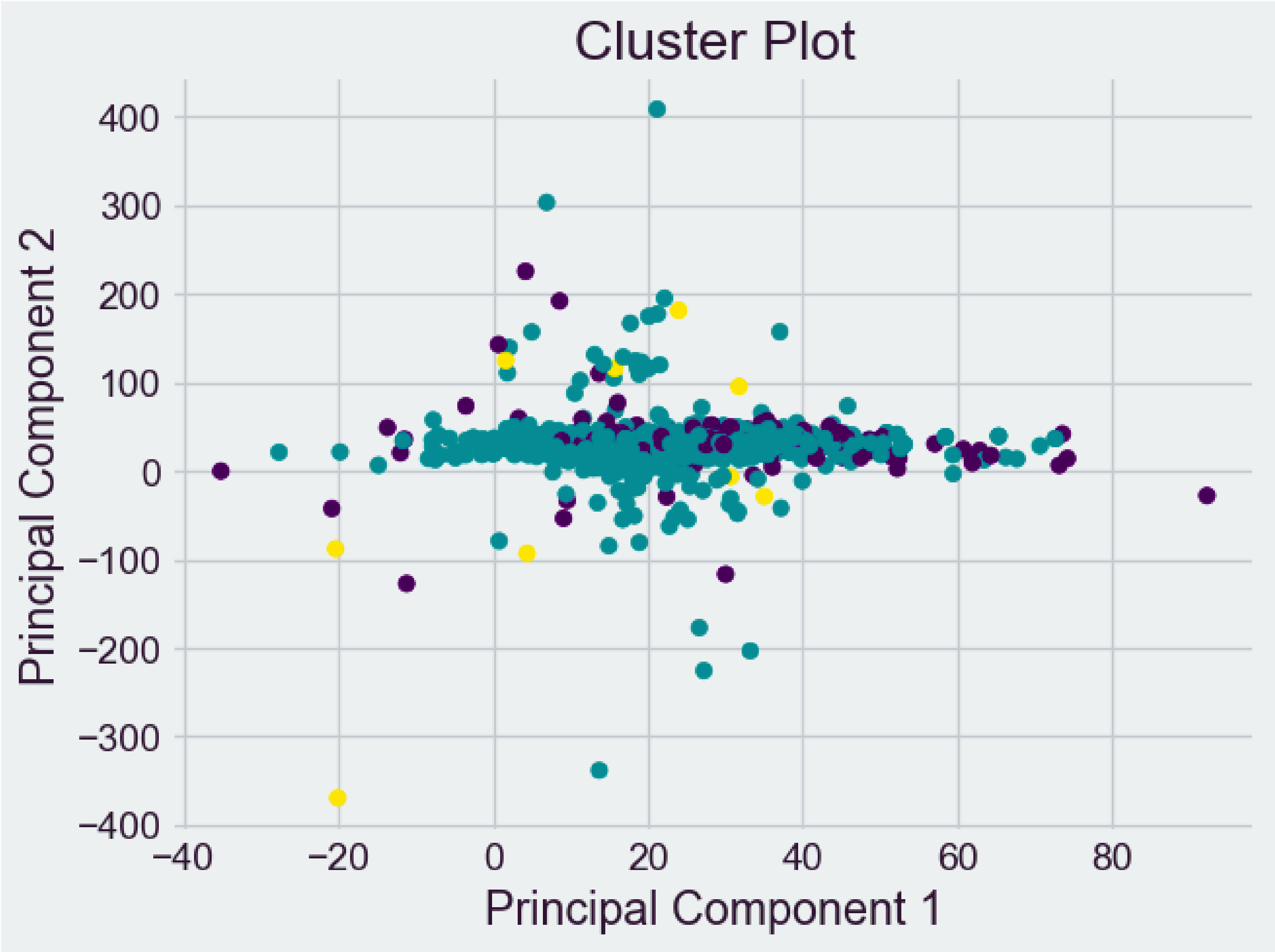
VISUALIZING OUR DATA



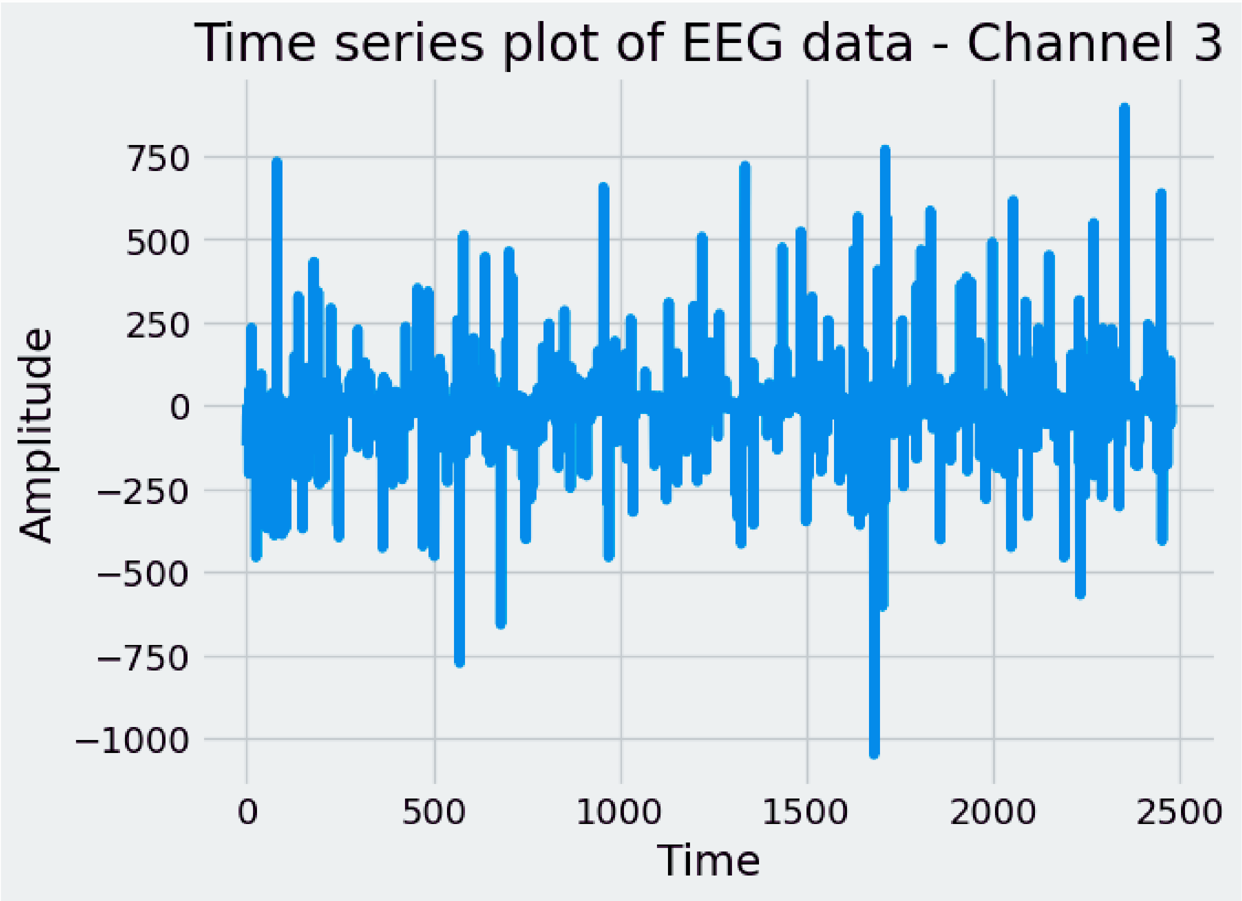


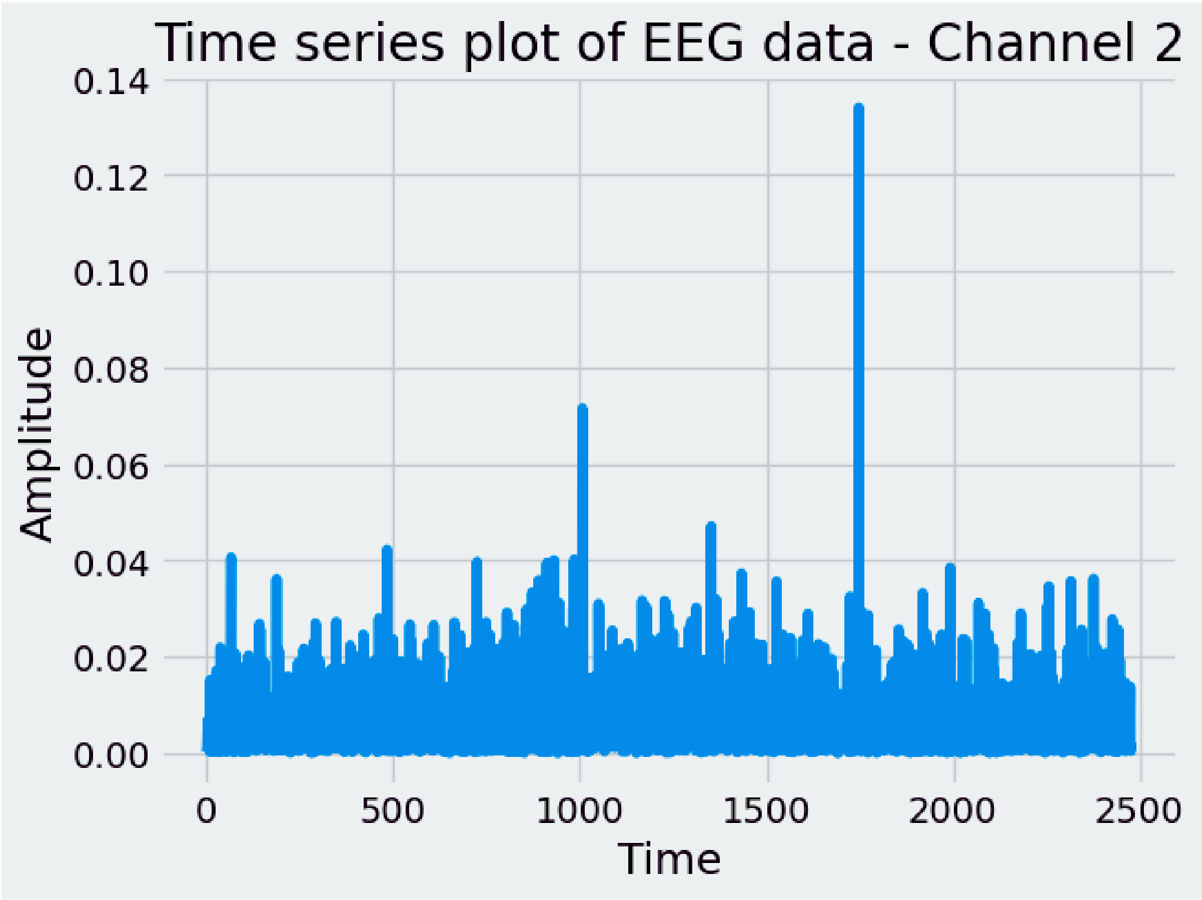
VISUALIZING OUR DATA

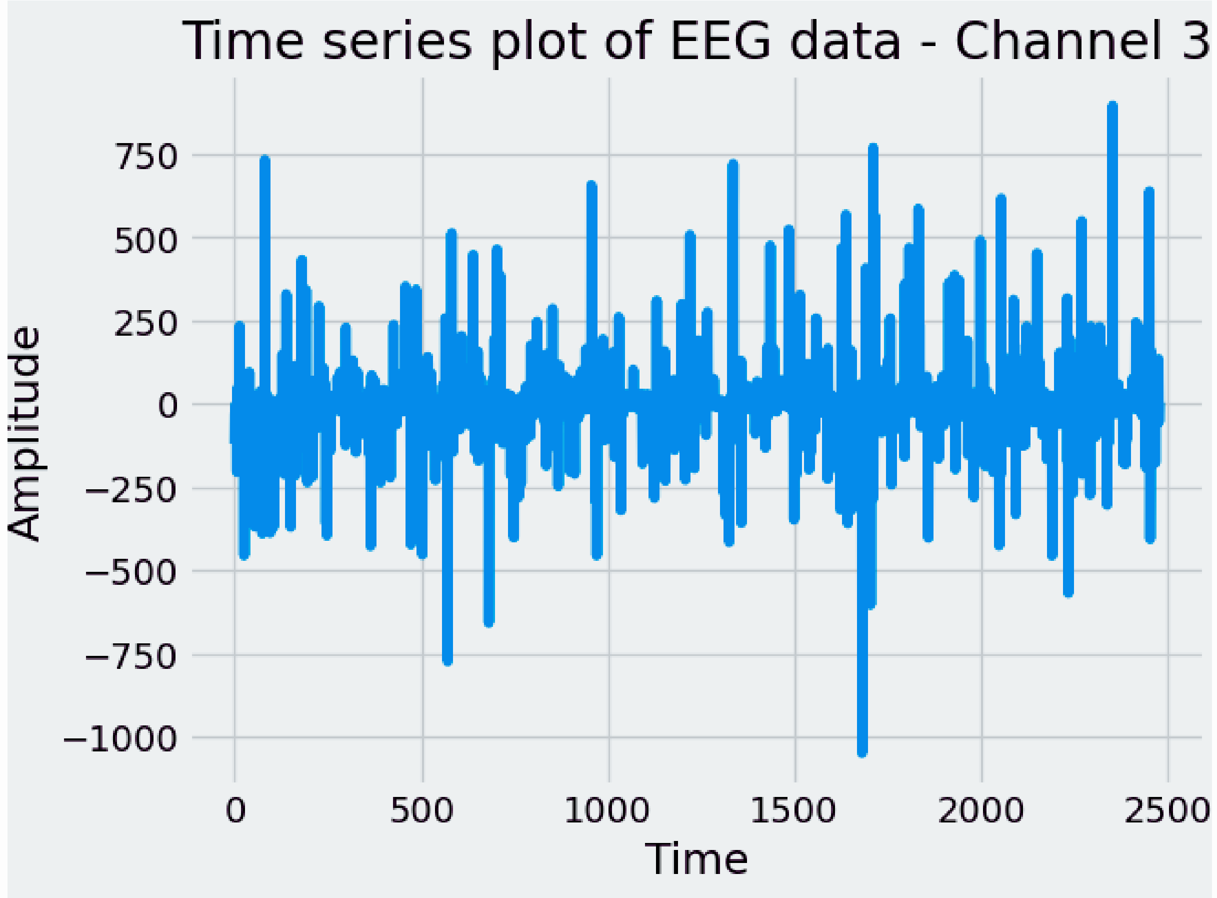




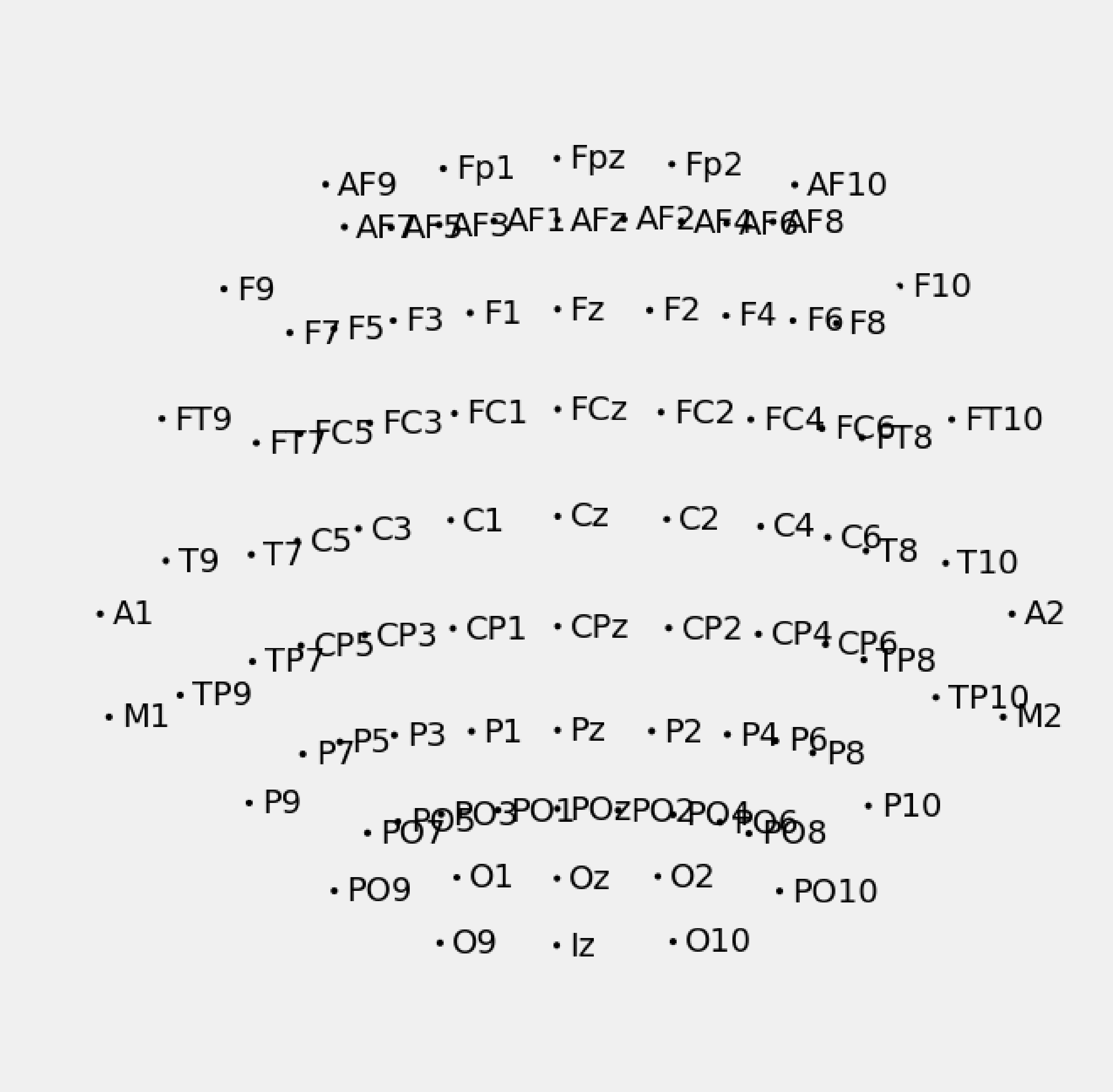
VISUALIZING OUR DATA

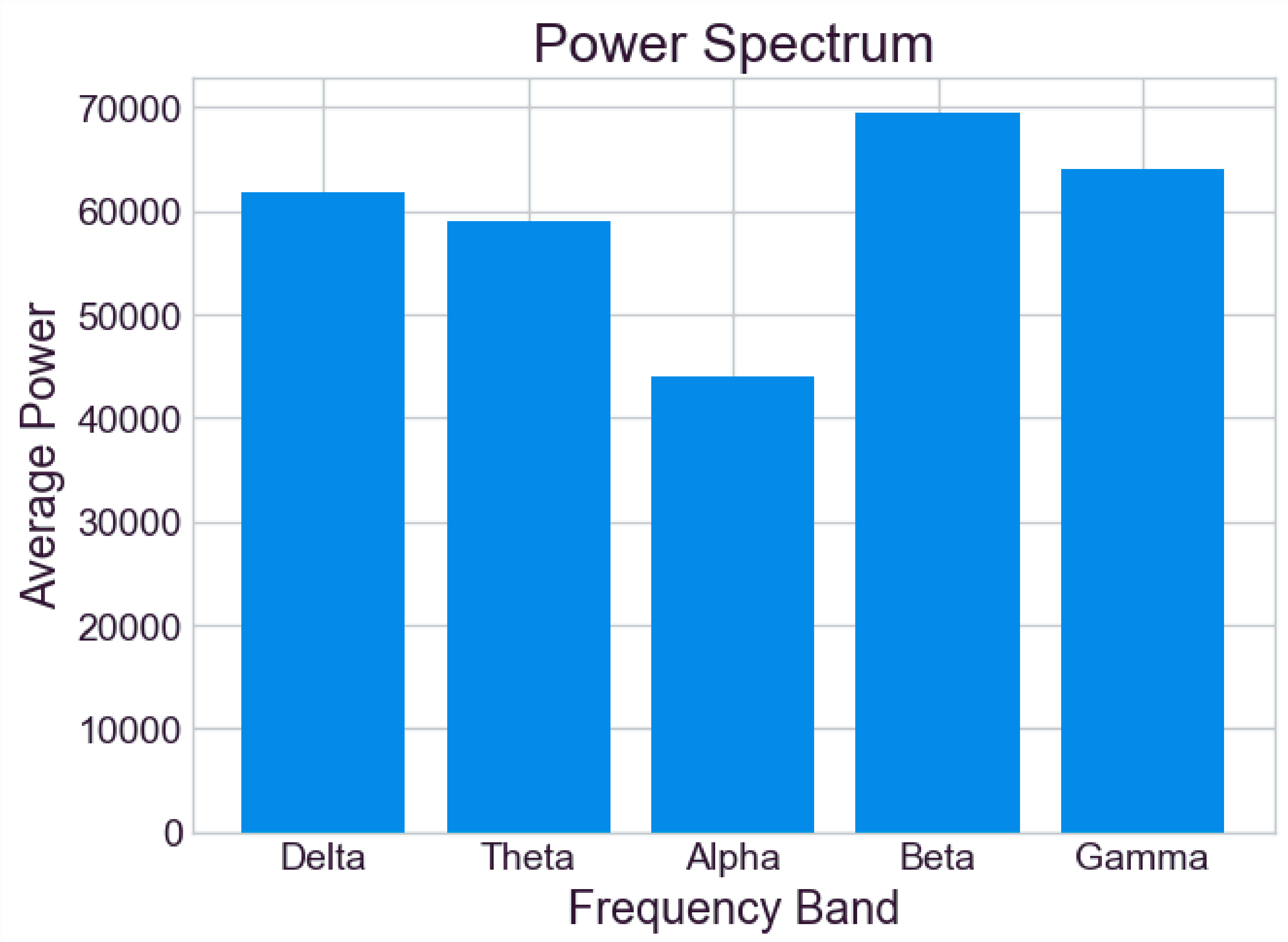




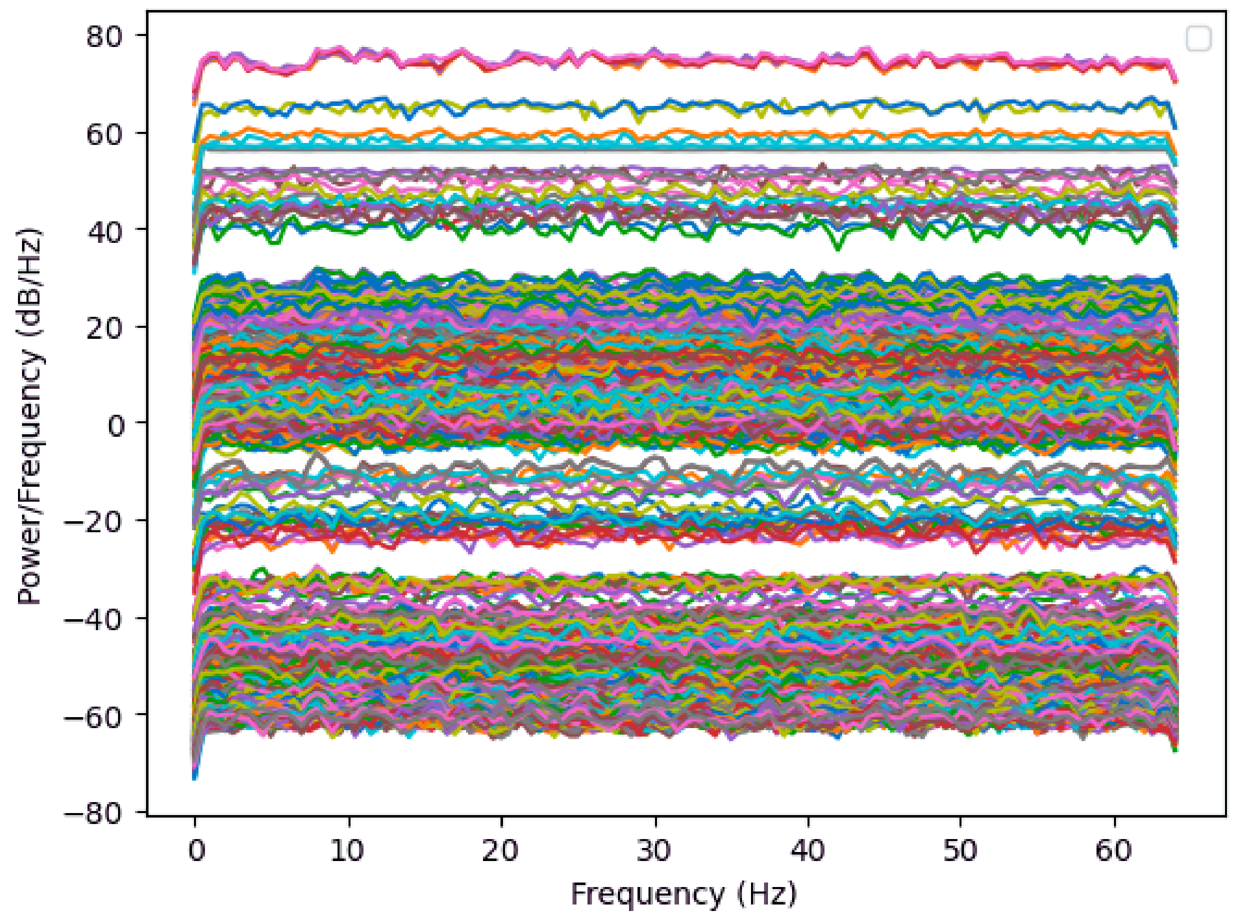


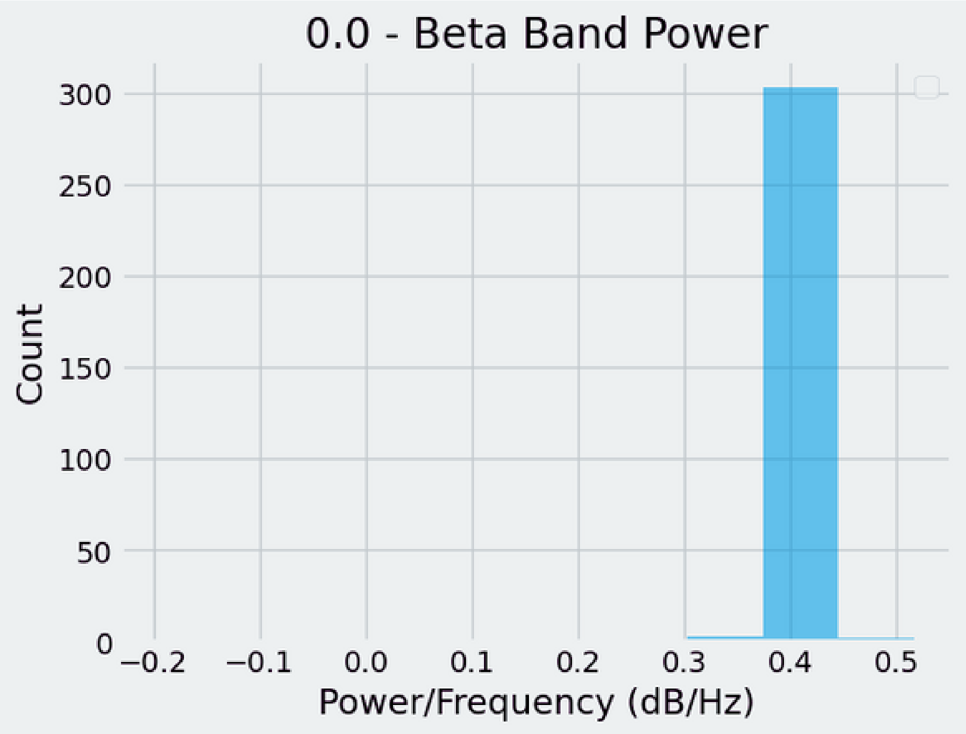
VISUALIZING OUR DATA

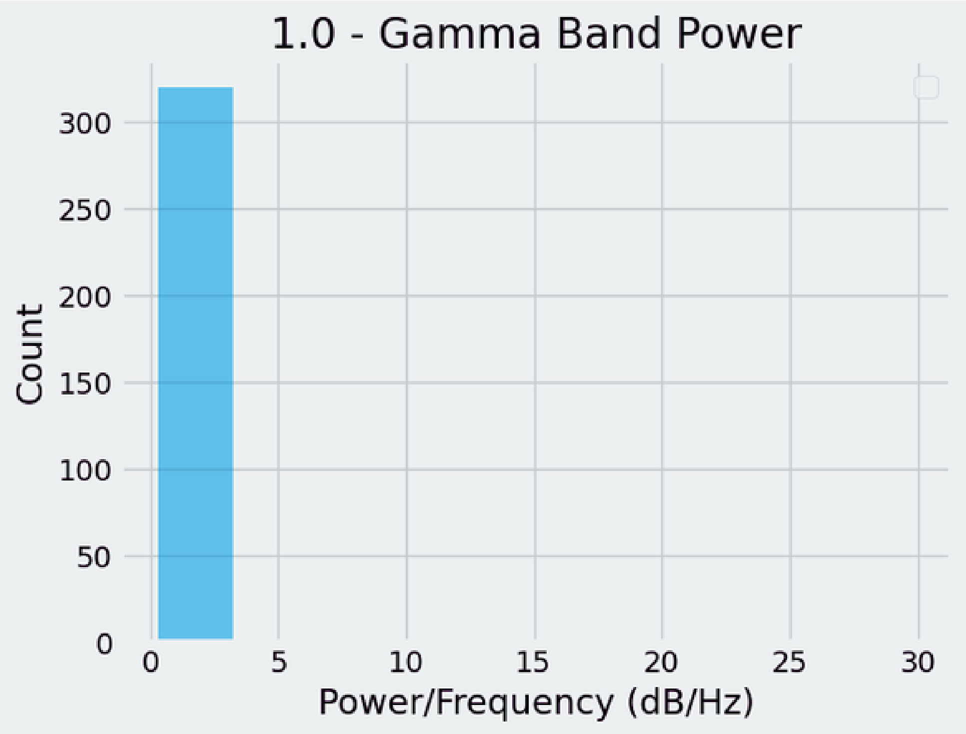


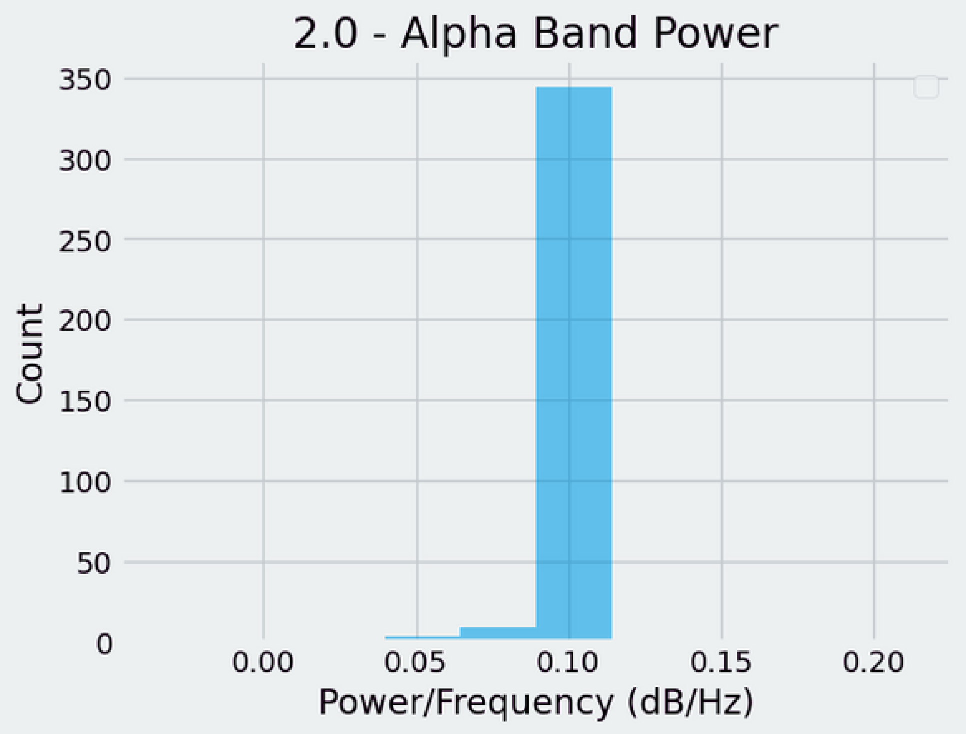


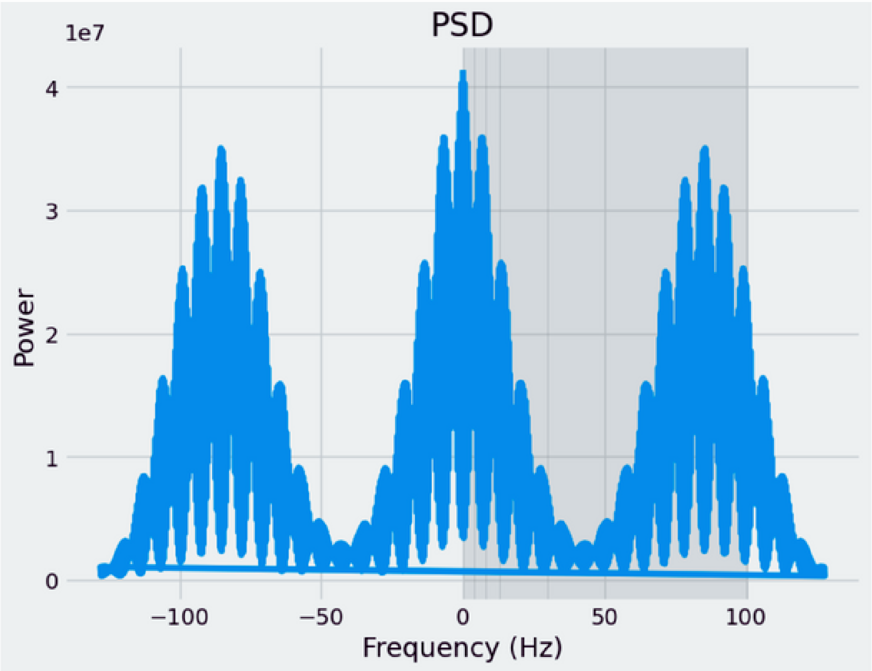
VISUALIZING OUR DATA



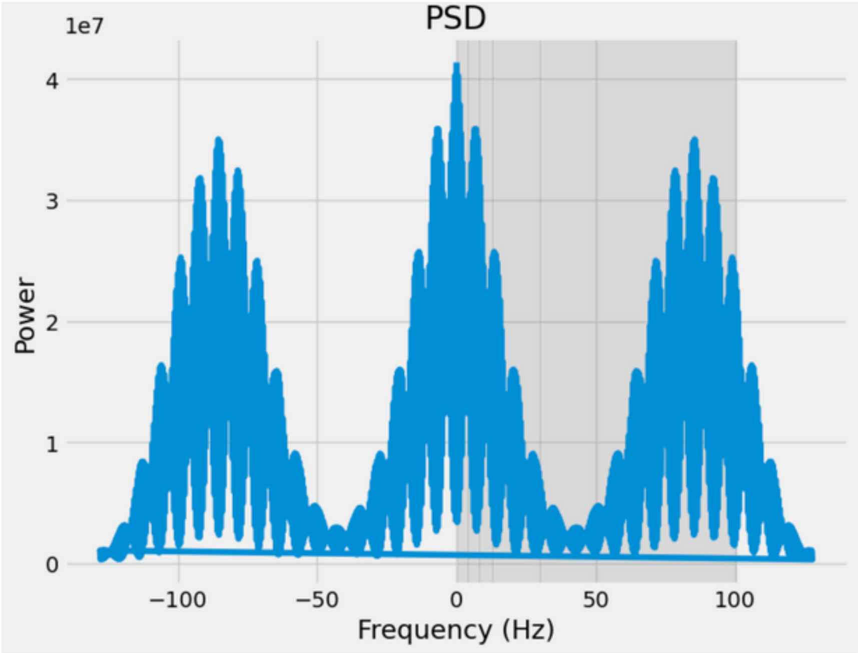






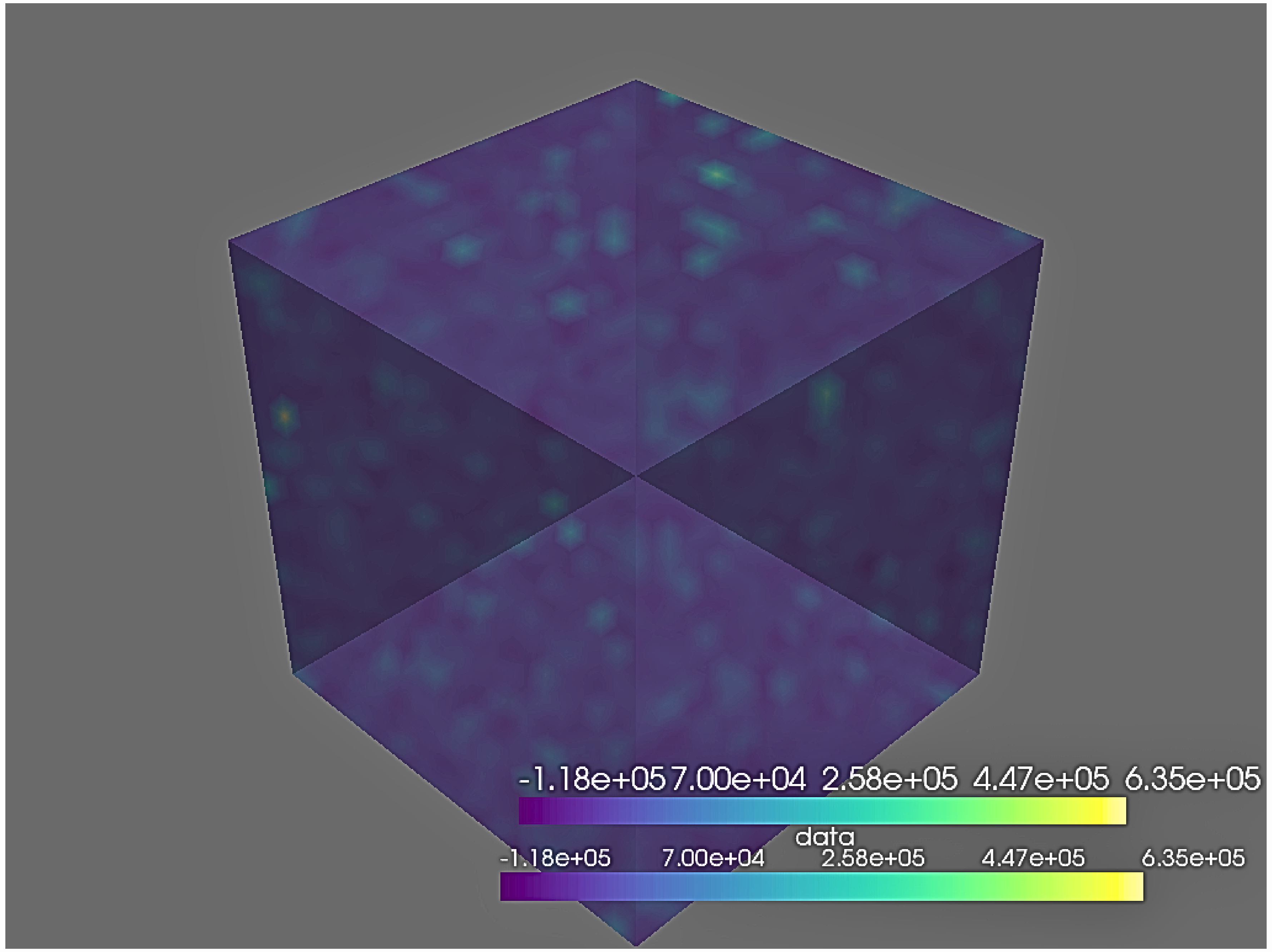


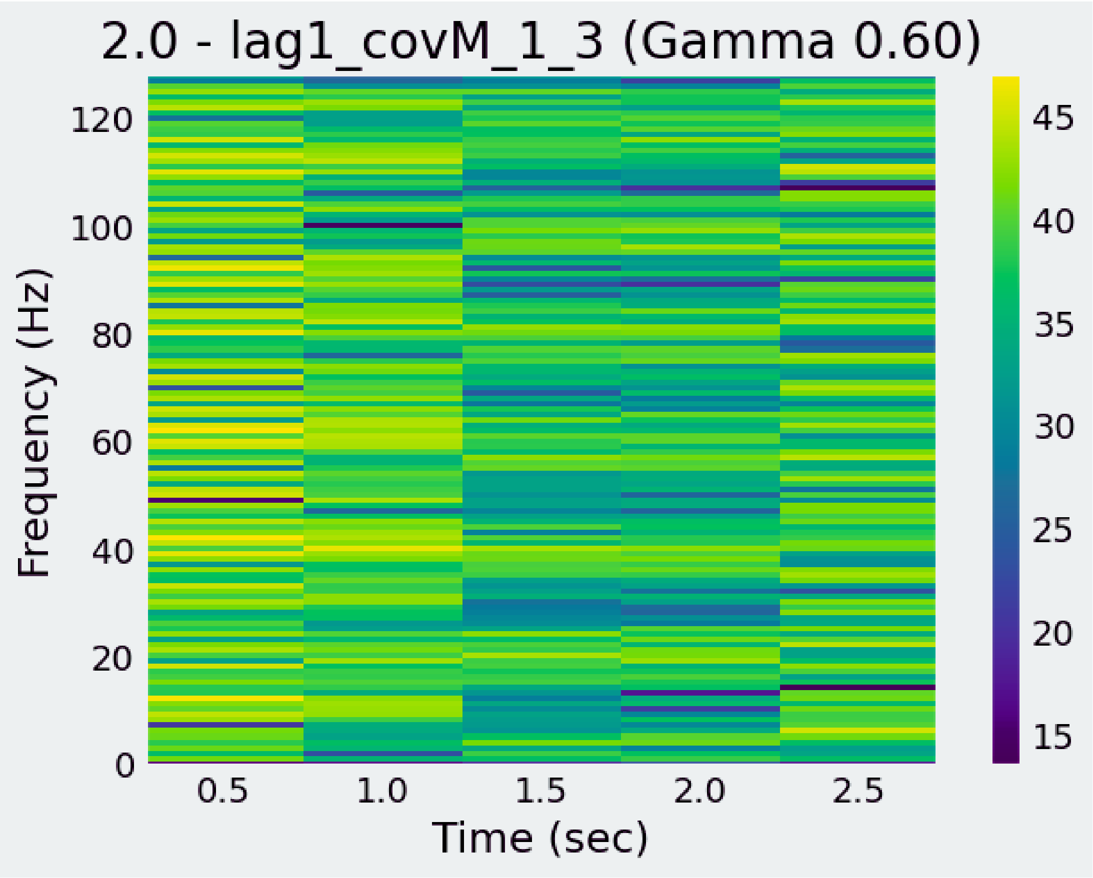

```
Average power in delta band: 22516814.28
Average power in theta band: 19320217.95
Average power in alpha band: 17373162.41
Average power in beta band: 8397487.59
Average power in gamma band: 9635000.76
```

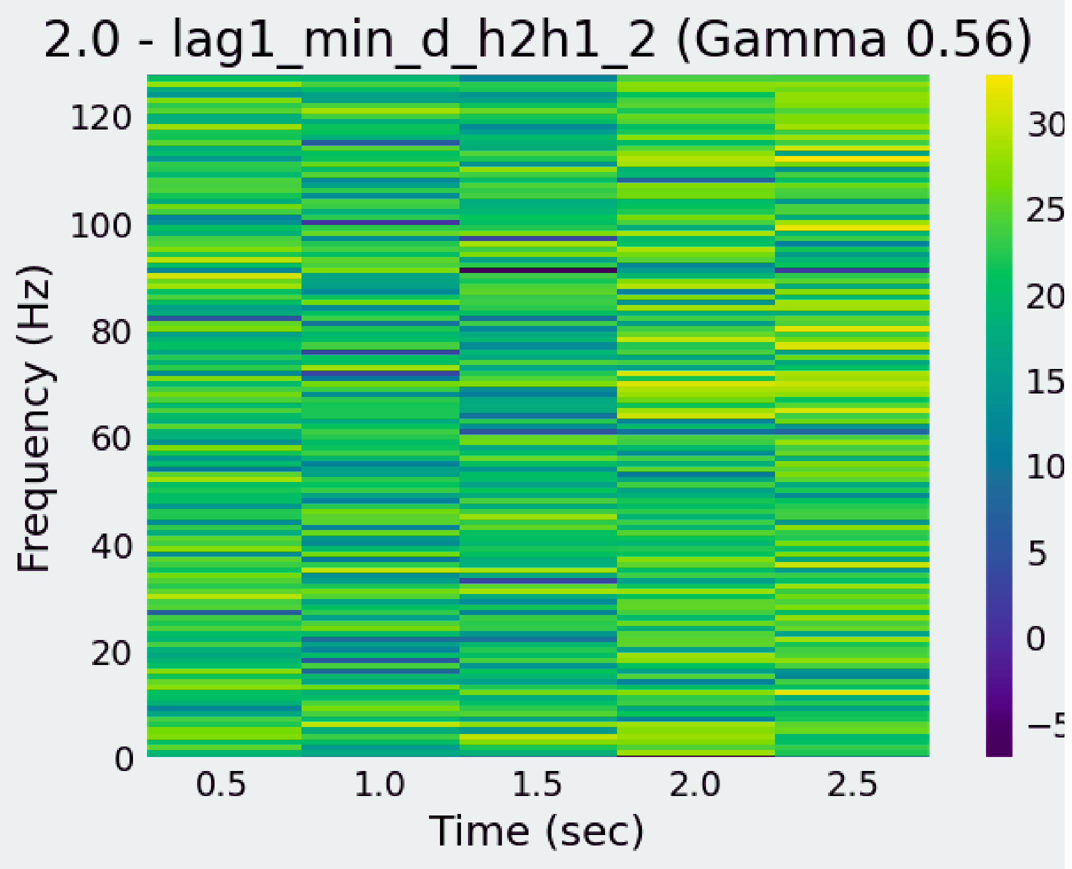



```
Frequency bands for 2.0:
delta: 52324436.68
theta: 44824800.60
alpha: 40099950.75
beta: 18725588.37
gamma: 22218417.09
Frequency bands for 1.0:
delta: 13431284.57
theta: 11831570.63
alpha: 11337968.39
beta: 6141821.52
gamma: 6403262.73
Frequency bands for 0.0:
delta: 1516402.89
theta: 1062310.35
alpha: 457383.06
beta: 216631.90
gamma: 157719.94
```

VISUALIZING OUR DATA







MODEL SELECTION AND EVALUATION

Model Selection: Model selection is the process of choosing the best algorithm or model for a given task or problem. It involves comparing the performance of different models using a set of evaluation metrics and selecting the best model.

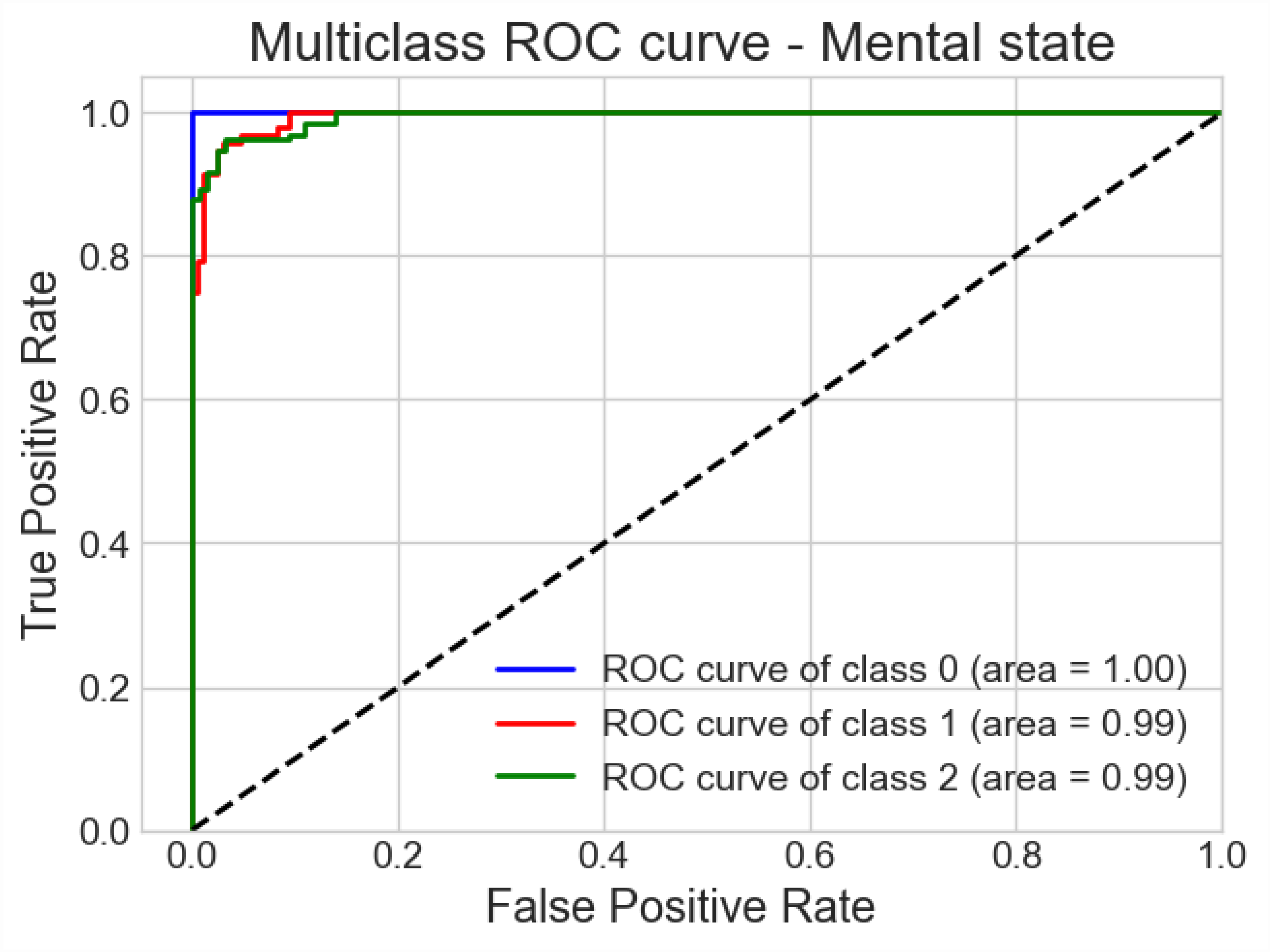
To select the appropriate machine Learning model, we go through the following steps:

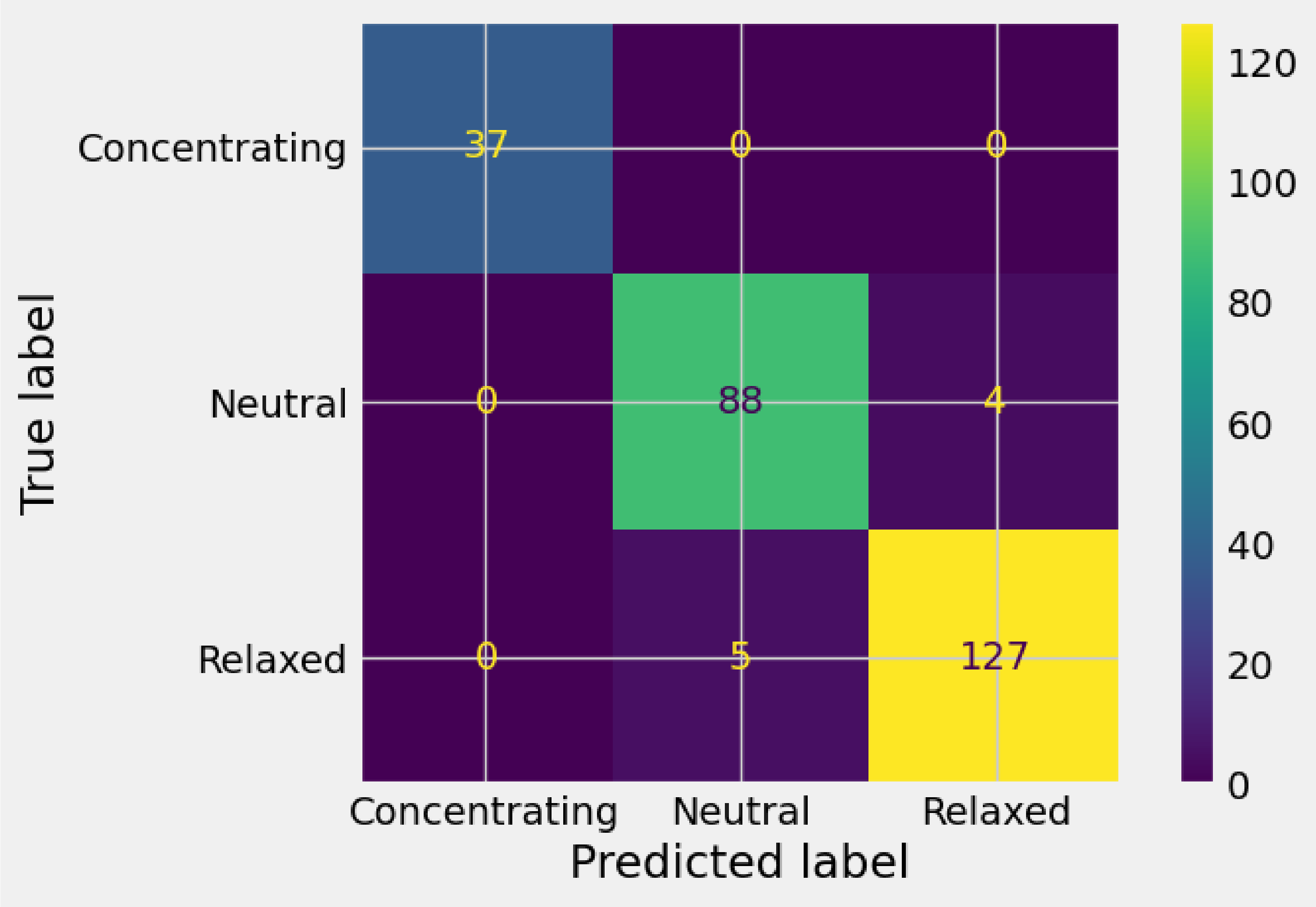
- Define the problem and task: It is a multi-class classification problem
- Select a set of models: LogisticRegression, DecisionTreeClassifier, RandomForestClassifier, SVC
- Training and finding the best model: Performed 5-fold cross-validation across the different models to determine the model that best fits the data.
- Performance metrics: Evaluated the accuracy, sensitivity, specificity, and F1 score for each class with the help of the confusion matrix generated.
- Select the best model: SVC gives the best overall accuracy for our data. So we are going to evaluate it on the test set to ensure that the model does not overfit and is generalizing well to the new data.
- Evaluate the selected model: Once you have selected the best model, evaluate its performance on the test set to ensure that it is not overfitting and generalizing well to new data.
- Refine and iterate: Lastly, we can refine our dataset by trying out different parameters via hyperparameter tuning, also by adding new features or by selecting a different model entirely and train the data over it. Iterate over this process until you find the model that performs well on the test set.

MODEL SELECTION AND EVALUATION

Metrics/Classes	Precision TP/(TP+FP)	Sensitivity TP/(TP+FN)	Specificity TN/(TN+FP)	F1 Score 2*(precision*recall)/(precision+recall)	Per-class accuracy (TP+TN)/(TP+TN+FP+FN)
Concentrating	1.0	1.0	1.0	1.0	1.0
Neutral	0.95	0.9565	0.9704	0.9513	0.96
Relaxed	0.97	0.9621	0.9689	0.9657	0.96

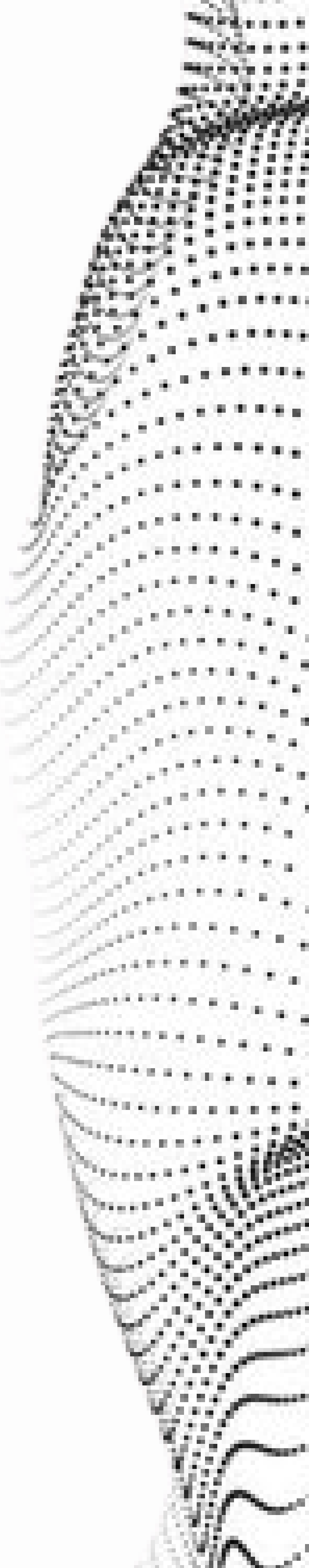
MODEL SELECTION AND EVALUATION





CONCLUSION AND FUTURE WORK

- The support vector machine (SVM) model achieved the highest accuracy of 90% in detecting brainwave states in students, indicating that it can accurately predict a student's brainwave state in real-time.
- The model can be used to promote self-awareness and help students better regulate their emotions and cognitive processes by understanding their own brainwave patterns, which can lead to better academic and personal outcomes.
- Educators and researchers can use the SVM model to gain insights into how students learn and develop, allowing for tailored educational interventions that better support student learning needs.
- The model can be used to detect when a student is in a distracted or relaxed state, providing personalized interventions and adjusting the learning environment to improve learning outcomes.
- The high accuracy of the SVM model suggests that it has great potential for practical applications in education and beyond.
- Future work could explore more advanced machine learning models, such as deep learning algorithms, to improve the accuracy of the brainwave state detection.



REFERECES

1. "Emotion Recognition Using Brain Waves" by R. Rajalakshmi and V. Radha, International Journal of Engineering and Technology, vol. 5, no. 3, pp. 2523-2528, 2013.
2. "A Comprehensive Review on EEG-Based Emotion Recognition Techniques" by P. Rajalakshmi and P. Priya, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 7, no. 1, pp. 462-469, 2017.
3. "EEG-based Emotion Detection and Recognition: A Systematic Review" by F. Liao, Z. Li, and G. Chen, IEEE Transactions on Affective Computing, vol. 10, no. 3, pp. 374-393, 2019.
4. "Emotion Detection using EEG Signals: A Survey" by S. Dey and S. Roy, Proceedings of the 3rd International Conference on Computing, Communication, Control and Automation, pp. 47-50, 2017.
5. "Emotion Detection from EEG Signals Using Wavelet Transform and Support Vector Machines" by A. Ahmadi, H. Saadat, and M. Gholami, International Journal of Engineering and Technology, vol. 7, no. 2.35, pp. 106-110, 2018.
6. "EEG-based Emotion Recognition using Convolutional Neural Network and Deep Belief Network" by T. Kim, B. Lee, and Y. Kim, Neurocomputing, vol. 311, pp. 216-223, 2018.

ROLES AND RESPONSIBILITIES

- Furqaan: Research and Visualization, responsible for conducting research on brainwave data and visualizing the results
- Tanvi: Data Cleaning and Feature Extraction, responsible for cleaning the raw data and extracting relevant features to be used in the model, Model Selection, Training, and Testing.
- Prachiti: Training, and Testing, responsible for selecting an appropriate machine learning model, training it on the preprocessed data, and testing its performance.
- Pankti: Data Analysis and Research, responsible for analyzing brainwave data, reading relevant research papers, and providing insights to inform model development.
- Pratiksha: Model Deployment and Real-World Applications, responsible for deploying the trained model and exploring its potential applications in real-world scenarios.