Systematic Evaluation of Baseline Machine Learning Methods and Novel Graph Neural Network Exploration for EEG-Based Brain State Classification on a Limited Dataset

Daksh Mor
Preliminary Task for GSOC NEURODYAD

April 6, 2025

Contents

Abstract

Electroencephalography (EEG) provides rich information about brain activity, but classifying brain states, especially with limited data, presents significant challenges. This report details a systematic evaluation performed as a preliminary task for the GSOC NEURODYAD project, focusing on classifying binary metabolic brain states using static EEG band power features from a 64-channel recording. The dataset comprised only 40 samples with 320 features (N « P), posing a high risk of overfitting. We first completed the requested tasks: evaluating standard machine learning classifiers (Logistic Regression, Support Vector Machines with various kernels, k-Nearest Neighbors) using robust 5-fold cross-validation, and exploring feature selection techniques (Univariate Feature Selection, Recursive Feature Elimination, Principal Component Analysis). Furthermore, motivated by the data's spatial structure, we conducted a novel exploration using Graph Neural Networks (GCNs), incorporating electrode topology and feature engineering, also evaluated via 5-fold cross-validation. Across all methods, including the GNN exploration, performance evaluated via cross-validation remained near or below chance levels (e.g., average AUC < 0.6), primarily attributed to the extreme data scarcity. These findings underscore the limitations of standard and advanced techniques on this specific dataset and motivate the use of methods like Contrastive Learning of Behavior Representations (CEBRA) on richer, time-series data, as planned for the main GSOC project. The full codebase and notebooks for this analysis are available at https://github.com/daksh-mor/cebra/tree/main.

1 Introduction

Electroencephalography (EEG) is a non-invasive neuroimaging technique that measures electrical activity generated by the brain via electrodes placed on the scalp [?]. It provides high temporal resolution for studying dynamic brain processes [?]. This report focuses on classifying brain states related to metabolic conditions using EEG band power features from a 64-channel system (Figure 1).

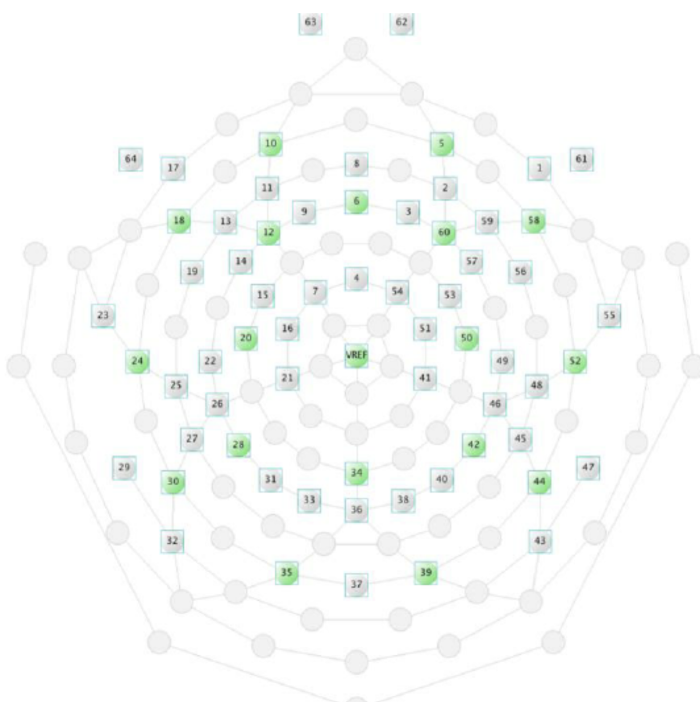

Figure 1: Standard 64-Channel EEG electrode configuration used for data acquisition.

The primary challenge is the dataset: 40 samples vs. 320 features (N « P), prone to overfitting [?].

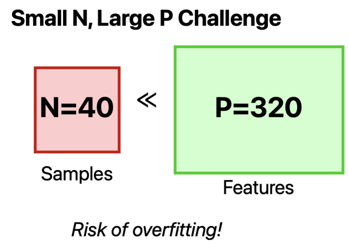

Figure 2: Conceptual representation of the N « P challenge (N=40, P=320).

This work, a preliminary task for GSOC NEURODYAD (https://github.com/daksh-mor/cebra/tree/main), involved:

1. **Task 1: Baseline Classification:** Evaluating standard algorithms (Logistic Regression, SVM, KNN) as requested.
2. **Task 2: Feature Selection:** Identifying informative features using UFS, RFE, and PCA as requested.
3. **Novel Exploration: Graph Neural Networks (GNNs):** Applying GCNs to leverage spatial topology as an additional investigation beyond the required tasks.

The goal was methodological evaluation and understanding data limitations to inform the main GSOC project.

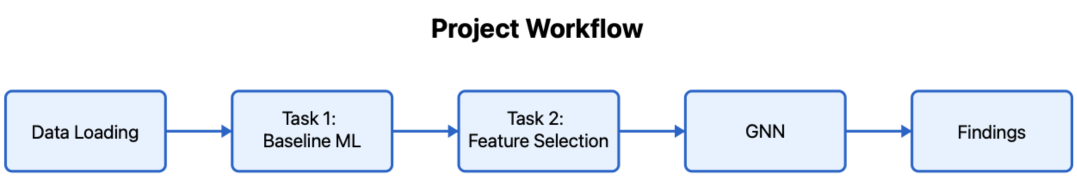

Figure 3: Overall project workflow, including Tasks 1-2 and the Novel GNN Exploration.

2 Methods

2.1 Dataset

As described: 40 samples (20 per class), 320 features (64 channels $\times$ 5 bands: Delta, Theta, Alpha, Beta, Gamma).

2.2 Data Preprocessing and Evaluation Strategy

- **Scaling:** StandardScaler applied within pipelines or CV folds.
- **Evaluation:** Primarily 5-Fold Stratified Cross-Validation. Metrics: Accuracy, F1-Score, AUC.

2.3 Baseline Machine Learning Models (Task 1)

Standard classifiers evaluated as per Task 1 requirements: Logistic Regression, SVM (Linear, Poly, RBF, Sigmoid kernels), KNN (various k).

2.4 Feature Selection Techniques (Task 2)

Top 5 features identified as per Task 2 requirements using: UFS (‘SelectKBest‘ with ‘f$_{c}lassif$‘), $RFE(withLinearSVM), PCA$

2.5 Novel Exploration: Graph Neural Network Approach

This approach was explored *in addition* to the requested tasks to leverage spatial information inherent in the EEG setup.

- **Graph Construction:**
  - *Nodes:* Each of the 64 EEG electrodes.
  - *Edges:* Connections established between electrodes based on physical adjacency defined in a predefined dictionary (derived from Figure 1). Self-loops were included.
  - *Node Features:* Initially, the 5 band power values for each electrode. In later experiments, feature engineering was applied.
- **Feature Engineering (GNN):** To enrich node representations, the following were added:
  - Regional Averages: Average power per band within predefined brain regions (e.g., frontal, central).
  - Band Ratios: Theta/Beta and Alpha/Beta power ratios, commonly used EEG metrics.
  - This resulted in 12 features per node (5 original + 5 regional + 2 ratio).
- **Model Architecture:** A Graph Convolutional Network (GCN) [?] was implemented using PyTorch Geometric [?]. The architecture typically included GCN layers, ReLU activation, Layer Normalization, Global Mean Pooling, and a final linear classifier with Dropout.
- **Training and Evaluation:** Trained using Adam optimizer and Binary Cross-Entropy with Logits loss, evaluated using 5-Fold Stratified Cross-Validation with feature scaling applied within each fold.

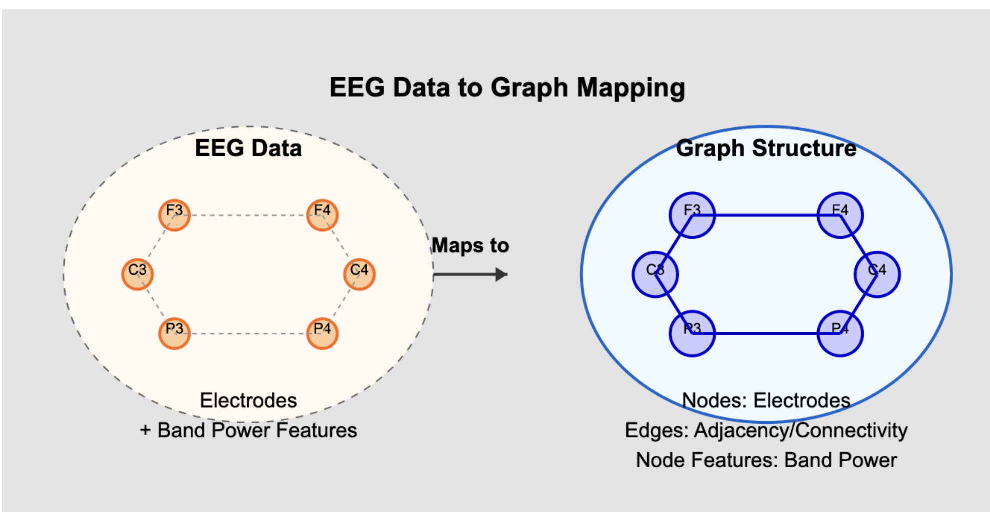

Figure 4: Conceptual illustration of constructing the graph representation from EEG data for the GNN exploration.

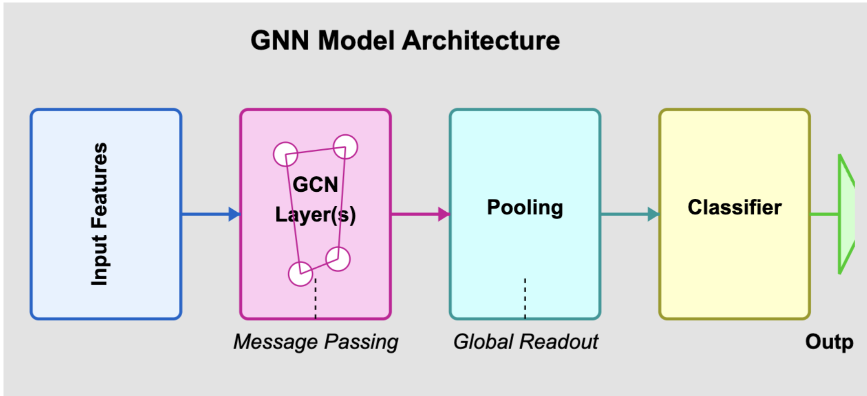

Figure 5: Simplified GCN architecture used in the novel exploration.

3 Results

3.1 Task 1: Baseline Model Performance

Robust 5-Fold Cross-Validation results (Table 1) showed poor generalization for all baseline models. Average performance hovered near or below random chance (AUC ≤ 0.288), with KNN models failing completely. SVM (Polynomial Kernel) had the highest average accuracy (0.400) but poor F1 and AUC scores. PCA visualization (Figure 6) confirmed the lack of clear linear separability.

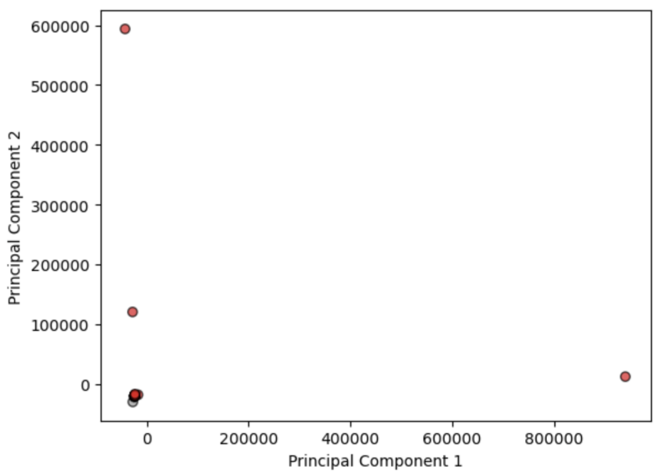

Figure 6: PCA visualization (first two components) of the dataset, colored by class.

Table 1: Baseline Model Performance (5-Fold Stratified Cross-Validation - Average ± Std Dev)

Model	Avg. Accuracy	Avg. F1 Score	Avg. AUC
Logistic Regression	0.300 ± 0.100	0.359 ± 0.090	0.288 ± 0.094
SVM (RBF Kernel)	0.375 ± 0.079	0.226 ± 0.192	0.213 ± 0.123
SVM (Linear Kernel)	0.300 ± 0.100	0.328 ± 0.082	0.275 ± 0.146
SVM (Poly Kernel)	**0.400 ± 0.050**	0.109 ± 0.218	0.138 ± 0.139
SVM (Sigmoid Kernel)	0.325 ± 0.100	0.223 ± 0.206	0.138 ± 0.092
KNN (k=5)	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
KNN (k=7)	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000

3.2 Task 2: Feature Selection Results

The top 5 features identified by each method were distinct (Table 2), reflecting their different criteria. Using these subsets in baseline models did not yield significant improvements in cross-validated performance.

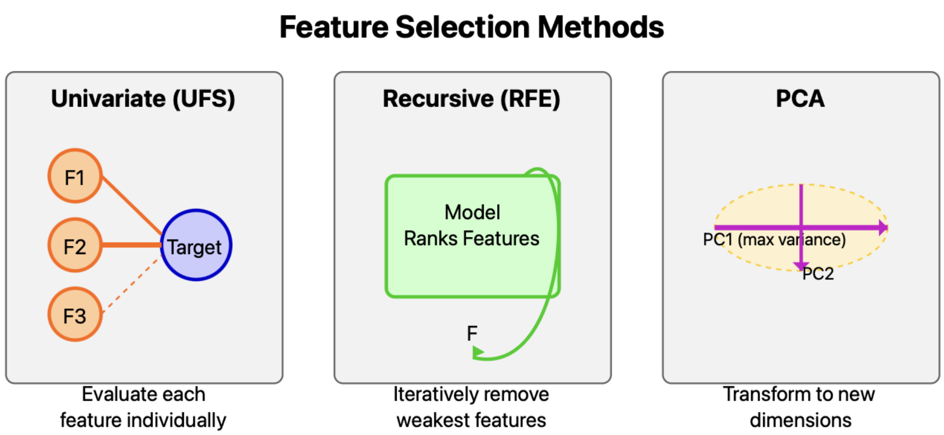

Figure 7: Conceptual difference between UFS, RFE, and PCA evaluation strategies.

Table 2: Top 5 Features Identified by Different Selection Methods (Example Run)

Method	Top 5 Features Selected
UFS (f_classif)	‘alpha10‘, ‘gamma23‘, ‘gamma34‘, ‘gamma38‘, ‘gamma45‘
RFE (Linear SVM)	‘alpha41‘, ‘beta43‘, ‘theta21‘, ‘theta29‘, ‘theta62‘
PCA (Loadings)	‘beta17‘, ‘theta1‘, ‘delta6‘, ‘delta47‘, ‘delta45‘

3.3 Novel Exploration: Graph Neural Network Performance

The GCN model explored as a novel approach, incorporating electrode topology and engineered features, was evaluated using 5-Fold Stratified CV (Table 3). While the average AUC (0.563) was marginally higher than the baseline models, performance remained poor overall with high standard deviation, indicating instability and failure to overcome the data limitation.

Table 3: GCN Performance from Novel Exploration (5-Fold Stratified Cross-Validation - Average ± Std Dev)

Model Configuration	Avg. Accuracy	Avg. F1 Score	Avg. AUC
GCN + Feat. Eng. (12 feat)	0.500 ± 0.137	0.493 ± 0.259	**0.563 ± 0.163**

4 Discussion

The results from both the required tasks and the novel GNN exploration consistently highlight the impact of the dataset's characteristics (N=40, P=320).

- **Task Completion and Baseline Limitations:** Tasks 1 and 2 were completed, establishing that standard ML models perform poorly under robust evaluation (5-Fold CV) and that simple feature selection does not remedy the core issue.
- **Data Scarcity as Primary Limitation:** The extremely small sample size is the most significant factor hindering model generalization across all approaches.
- **Novel Exploration Confirms Limits:** The additional GNN exploration, while theoretically motivated by the data's spatial structure, also yielded near-chance performance. This reinforces that even advanced modeling techniques incorporating domain knowledge (electrode adjacency) struggle fundamentally when N is critically low relative to P. The slight improvement in average AUC for GNNs is marginal and accompanied by high variance.
- **Implications for Main Project:** This preliminary work, including the GNN exploration, serves as a valuable negative result. It strongly indicates that relying solely on static band power features from a small cohort is insufficient. It reinforces the rationale for the main GSOC NEURODYAD project's focus on richer, **time-series EEG data** and advanced methods like **CEBRA**, designed for such data regimes.

5 Conclusion

This study successfully completed the preliminary GSOC tasks, evaluating baseline ML models and feature selection techniques on a limited (N=40, P=320) EEG dataset. Furthermore, a novel exploration using Graph Neural Networks was conducted to leverage spatial information. The pervasive challenge of data scarcity dominated the results, leading to poor generalization performance near or below chance levels across all methods, including GNNs, when assessed using robust 5-fold cross-validation.

The key conclusion is that the provided static features from this small dataset lack sufficient signal or sample size for reliable classification using the evaluated standard or advanced exploratory methods. This preliminary work successfully demonstrated evaluation methodologies but ultimately highlighted the need for alternative approaches, validating the direction of the main GSOC NEURODYAD project towards analyzing richer, time-locked dyadic EEG data with advanced representational learning methods like CEBRA. The codebase detailing this exploration is available at https://github.com/daksh-mor/cebra/tree/main.