EEG-based Prediction of Learning Outcomes using Machine Learning

Zuzanna Bąk

Abstract

This project investigates whether EEG band-power features can predict a learner's comprehension of instructional material. Using a public Kaggle dataset containing multichannel EEG recordings paired with binary comprehension labels, two model families—feed-forward neural networks (FNN) and XGBoost—were evaluated under a strict subject-wise split to avoid person-specific leakage. Three imbalance-handling strategies were tested: baseline, class weighting, and SMOTE. Results show that the FNN baseline achieved the most meaningful separability (ROC-AUC = 0.704), while XGBoost reached the highest raw accuracy but failed to discriminate between classes (AUC < 0.50). Confusion matrices confirm that XGBoost predominantly predicts the majority class. Overall, band-power features contain only a weak signal, and robust evaluation requires metrics beyond accuracy. Three accompanying Jupyter notebooks provide full EDA, modeling code, and imbalance experiments.

1 Introduction

Understanding whether a student has absorbed new information is a central challenge in cognitive science and educational technology. Traditional assessments provide delayed, indirect feedback, whereas EEG offers a non-invasive and time-resolved measure of neural activity associated with attention and information processing. This raises the question: can comprehension be inferred directly from EEG signals?

The dataset used in this project contains multichannel EEG recordings collected as participants watched short instructional videos, along with binary labels indicating whether they understood the material. The classification task is challenging due to: (1) strong class imbalance, (2) substantial inter-subject variability, and (3) overlapping feature distributions observed in exploratory data analysis.

Prior work suggests that simple spectral EEG features often carry only weak discriminative information for complex cognitive states such as workload or engagement [2, 1, 4]. This motivates a careful evaluation of whether band-power features are sufficient for comprehension prediction.

The goal of this work is to build and evaluate machine learning models that classify comprehension from EEG band-power features. Two model families—FNN and XGBoost—were trained under a subject-wise split to prevent leakage of person-specific patterns. Multiple imbalance-handling techniques were tested to assess whether improving representation of the minority class increases detection of non-comprehension. The following sections describe the approach, present results, and discuss implications.

2 Approach

2.1 Dataset

The dataset was obtained from a publicly available Kaggle repository [3]. It contains multichannel EEG recordings paired with binary comprehension labels:

- **class 1**: participant understood the material,
- **class 0**: participant did not understand.

The distribution is strongly imbalanced, with class 1 dominating. To prevent subject-level leakage, samples were grouped by participant and split into **train/test by subject**, ensuring no individual appears in both sets.

2.2 Feature Extraction and Preprocessing

Each EEG segment was transformed into approximately 70 **band-power (POW) features** across standard frequency bands (alpha, beta, gamma, etc.) and channels. Preprocessing included:

- removing `video_id` and `subject_id` to avoid non-physiological shortcuts,
- z-score standardization of all numeric features,
- enforcing subject-wise partitioning.

Early experiments with 2D time–frequency images and CNNs did not improve performance; thus final experiments focus on POW features, which are compact and widely used in EEG classification.

2.3 Models

Feed-Forward Neural Network (FNN). The FNN uses several dense layers with ReLU activation and dropout for regularization. The Adam optimizer and early stopping were applied. This architecture captures non-linear relationships between EEG features.

XGBoost. XGBoost was configured with standard hyperparameters (learning rate, max depth, subsampling, and boosting rounds with early stopping). This model provides a strong tabular-data baseline complementing the FNN.

2.4 Imbalance Handling

Three imbalance strategies were evaluated:

1. **Baseline**: no class weighting or oversampling,
2. **Class weighting**: weights inversely proportional to class frequencies,
3. **SMOTE**: synthetic oversampling of class 0 in the training set.

These were tested to examine whether improved minority representation enhances detection of non-comprehension.

2.5 Evaluation Protocol

All results are computed under the same subject-wise split. Metrics:

- **Accuracy** (misleading in imbalanced settings),
- **F1-score** (dominated by majority class),
- **ROC-AUC** (primary metric; measures class separability).

Confusion matrices are included to visualize prediction distributions. All experiments used fixed random seeds.

3 Results

3.1 FNN Performance

Table 1: FNN performance under imbalance-handling strategies.

Variant	Accuracy	F1-score	ROC-AUC
Baseline	0.734	0.836	0.704
Class weights	0.732	0.838	0.677
SMOTE	0.718	0.832	0.627

Confusion Matrix (FNN Baseline):

Table 2: FNN baseline confusion matrix.

	Pred 0	Pred 1
True 0	617	1772
True 1	1444	6888

The FNN baseline achieved the strongest separability. Although recall for class 0 remained low, the model captured meaningful differences in EEG patterns, reflected in the AUC of 0.704. Both class weighting and SMOTE reduced separability.

3.2 XGBoost Performance

Table 3: XGBoost performance under imbalance-handling strategies.

Variant	Accuracy	F1-score	ROC-AUC
Baseline	0.763	0.865	0.460
Class weights	0.577	0.710	0.445
SMOTE	0.609	0.733	0.489

Confusion Matrix (XGBoost Baseline):

Table 4: XGBoost baseline confusion matrix.

	Pred 0	Pred 1
True 0	152	2237
True 1	860	7472

Although XGBoost achieved the highest accuracy, the confusion matrix and AUC reveal that it predicts almost exclusively class 1. The classifier fails to learn EEG patterns distinguishing the two states.

3.3 Comparison

The FNN baseline is the only model exhibiting moderate discriminative ability. XGBoost's high accuracy is misleading due to extreme class imbalance. Rebalancing strategies did not improve performance, reinforcing that the primary limitation lies in the weak separability of band-power EEG features rather than the imbalance itself.

4 Discussion

Band-power EEG features exhibit substantial overlap between comprehension states, consistent with the confusion matrices and low AUC for most models. The FNN captured non-linear structure and achieved moderate separability (AUC = 0.704). XGBoost relied heavily on predicting the majority class, confirming that accuracy alone inflates perceived performance.

Imbalance-handling techniques did not improve detection of class 0. SMOTE increased representation of minority samples but introduced noise and reduced separability. Class weighting improved sensitivity marginally but did not enhance overall discrimination.

These findings highlight the difficulty of decoding comprehension from short EEG windows using simple spectral features. Subject-level variability and weak physiological markers contribute to the challenge, consistent with prior evidence that band-power EEG features provide limited discriminative power for higher-level cognitive states [2].

5 Conclusion

This project evaluates whether EEG band-power features can predict comprehension of instructional material. Under subject-wise evaluation, the FNN baseline provides the only meaningful separability, while XGBoost fails to discriminate despite high accuracy. Imbalance-handling methods do not materially improve performance. Overall, band-power features contain only a weak predictive signal for this task. Future work should explore richer time–frequency representations, connectivity features, and temporal deep learning models.

Additional Deliverables: Three Jupyter notebooks are included:

- **EDA.ipynb** — exploratory data analysis, feature distributions, imbalance assessment,
- **Modeling.ipynb** — FNN and XGBoost training pipeline,
- **ClassImbalance.ipynb** — SMOTE, class weighting, and extended evaluation.

References

[1] B. Blankertz et al. *Optimizing Spatial Filters for Robust EEG Single-Trial Analysis*. IEEE Signal Processing Magazine, 2007.

[2] F. Lotte et al. *A Review of Classification Algorithms for EEG-based Brain–Computer Interfaces*. Journal of Neural Engineering, 2018.

[3] Kaggle. *EEG Brainwave Dataset: Predict Student Understanding*. Dataset available at: `https://www.kaggle.com/datasets/madyanomar/eeg-data-distance-learning-environment/code` (accessed: Nov, 7th 2025).

[4] T. Zander and C. Kothe. *Towards Passive Brain–Computer Interfaces: Understanding Implicit User States*. Journal of Neural Engineering, 2011.