A Cluster-Aware EEG Classification Framework with Brain Topographic Imaging and Gradient-Boosted Trees

Joowon Oh, Soyoung Jang, Hyeonji Rhee, Suah Hong, Chaewon Kim, Yeojin Shin, Jaeho Lee*

Duksung Women's Univ., *Duksung Women's Univ.

2000violet@duksung.ac.kr, *izeho@duksung.ac.kr

뇌 지형 EEG 이미지와 클러스터 인식 특징을 활용한 그래디언트 부스팅 기반 분류 프레임워크

오주원, 장소영, 이현지, 홍수아, 김채원, 신여진, 이재호*
덕성여자대학교, *덕성여자대학교

Abstract

This study proposes a cluster-aware EEG classification framework for mood disorder identification based on brain topographic imaging and gradient-boosted decision trees. EEG spectral features are transformed into scalp topography images, encoded via a pretrained CNN, and enriched with latent structure information through band-wise clustering. Experimental results demonstrate that the proposed approach achieves robust and reliable classification performance, highlighting its potential as an objective EEG-based diagnostic support method.

Ⅰ. Introduction

Psychiatric disorders, particularly mood disorders such as major depressive disorder and bipolar disorder, are highly prevalent and are characterized by persistent or episodic emotional disturbances that impair daily functioning, making accurate discrimination from healthy individuals a clinically important task. However, psychiatric diagnosis largely depends on subjective clinical assessments, leading to variability and limited reproducibility, which highlights the need for objective, data-driven diagnostic support. Electroencephalography (EEG) provides a non-invasive and cost-effective means of capturing neural activity with high temporal resolution, and recent advances increasingly adopt image-based EEG representations to preserve spatial relationships among channels. In this study, spectral features from the alpha, beta, high-beta, and gamma bands are transformed into scalp topographic images and encoded using a convolutional neural network as a fixed feature extractor. Density-based clustering is then applied in a reduced latent space to characterize intrinsic feature structure and reduce the influence of noisy samples, followed by gradient-boosted decision tree classification for robust and interpretable mood disorder identification. Through this integrated framework, we aim to enhance the objectivity and reliability of EEG-based diagnostic support for mood disorders.

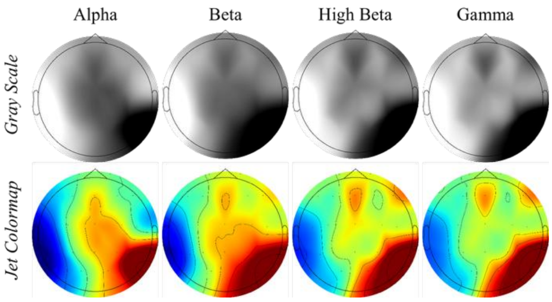

Figure 1. EEG Topographic Maps Across Frequency Bands

Ⅱ. Method

a. EEG to Image

Multichannel EEG spectral features were transformed into two-dimensional scalp topographic images using a standardized 10- 20 electrode layout. For each frequency band, channel-wise spectral powers were min- max normalized and projected onto the scalp surface via topographic interpolation, producing single-channel grayscale brain maps that shown in Fig. 1 preserve spatial relationships among EEG channels.

b. Feature Extraction via CNN Model

A pretrained EfficientNet-B0 [1] model was employed as a fixed CNN-based feature extractor to obtain robust and generalizable visual representations from EEG brain topography images. The images were fed into the frozen network and encoded into 1280-dimensional feature vectors using the global pooled output, without any fine-tuning.

c. Band-wise Unsupervised Clustering via DBSCAN

To capture intrinsic structure in the learned feature space, band-wise unsupervised clustering was applied to the CNN embeddings. For each frequency band, features were standardized, reduced using PCA, and clustered using DBSCAN to identify dense regions and noise samples. Cluster-aware geometric descriptors derived from the clustering results were combined with the PCA-projected features for downstream classification, with band-wise feature distributions visualized in Fig. 2.

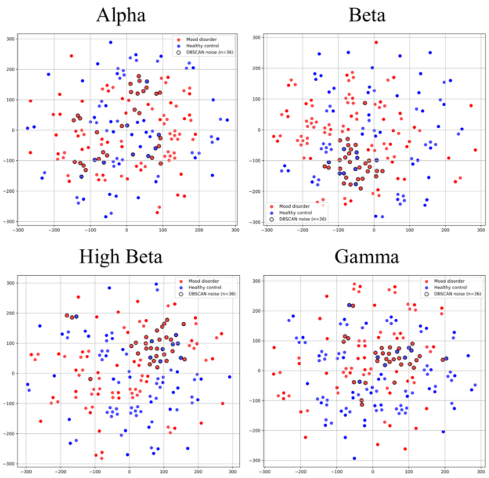

Figure 2. t-SNE Visualization of Band-wise DBSCAN Clustering Results

d. Classification Using XGBoost with Band-wise Cluster-Aware Features

The final classification stage employed an XGBoost model to leverage the structured, cluster-aware EEG features. For each frequency band, sample-level features were aggregated at the subject level by averaging, and band-specific representations were merged to form a unified multi-band feature vector. The model was trained using a binary logistic objective with constrained tree depth, a low learning rate, and subsampling to promote generalization and reduce overfitting.

The XGBoost classifier was trained using carefully selected hyperparameters summarized in Table 1. As shown in Fig. 3, the training log-loss rapidly decreases and stabilizes, while the validation curve remains relatively flat, indicating stable convergence. The final classification results in Table 2 demonstrate high overall accuracy and balanced precision- recall performance across classes, confirming the effectiveness of the proposed cluster-aware classification framework.

Table 1. XGBoost Training Parameters

Category	Parameter	Value
Tree growth	tree_method	hist
Tree complexity	max_depth	3
Learning rate	eta	0.01
Row subsampling	subsample	0.8
Column subsampling	colsample_bytree	0.6
Histogram bins	max_bin	128
Evaluation metrics	eval_metric	logloss, error
Boosting rounds	num_boost_rounds	3500

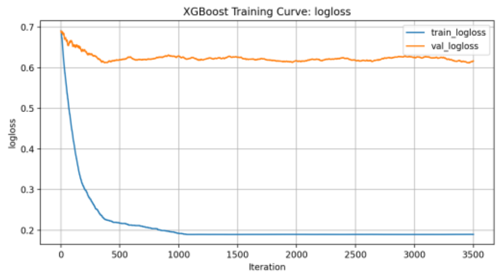

Figure 3. XGBoost Training Curves for Log-Loss

Table 2. Test Evaluation Classification Result

	Precision	Recall	F1-Score
Healthy Control	0.8182	1.0000	0.9000
Mood Disorder	1.0000	0.7778	0.8750
Accuracy			**0.8889**
Macro avg	0.9091	0.8889	0.8875
Weighted avg	0.9091	0.8889	0.8875

Ⅲ. Conclusion

This study demonstrated the feasibility of a cluster-aware EEG classification framework using brain topographic imaging and XGBoost for mood disorder identification. The results indicate that spatial EEG representations enriched with latent structure information can support robust and objective classification, with potential for extension to larger datasets and related psychiatric applications.

ACKNOWLEDGMENT

This work was conducted without external funding.

REFERENCES

[1] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.

[2] Chen, T. (2016). XGBoost: A Scalable Tree Boosting System. *Cornell University*.