논문 2026-2-17 http://dx.doi.org/10.29056/jsf.2026.06.17

EEG-Based Classification of Pediatric ADHD via Multi-Classifier, Multi-Band Ablation: Verifying the Standalone Predictive Capacity of the Delta-Band Biomarker

Fika Fauzia*, Young-In Kim*†

Abstract

While EEG-based machine-learning models achieve high accuracy, purely data-driven feature selection may inadvertently obscure mutually correlated, biologically significant markers like the delta-band (0.5 – 3.0 Hz). To evaluate its independent predictive capacity, we propose a five-stage spatial – spectral ablation with four classifiers (Random Forest, SVM, XGBoost, and a Multilayer Perceptron). Using a public dataset of 61 ADHD and 60 control children (19 channels, 128 Hz), we extract 1,159 features and evaluate every feature-set under stratified 10-fold cross-validation. The full-feature model reaches 98.4 – 99.6% accuracy (0.998 – 1.000 AUC). Notably, utilizing only the 57 delta-band power/entropy features the models attained a mean AUC of 0.846, far exceeding a label-permuted control (0.499). These results provide strong evidence for the delta band as a standalone biomarker, highlighting the necessity of balancing mathematical optimization with biological interpretability in medical AI.

keywords : Electroencephalography, Attention-deficit/hyperactivity disorder, Multi-classifier comparison, Ablation study, Feature-selection bias

1. Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a prevalent neurodevelopmental condition [5] primarily diagnosed via subjective clinical interviews [6], which risks misdiagnosis [1]. Consequently, electroencephalography (EEG) paired with machine learning has emerged as an objective diagnostic alternative, achieving classification accuracies above 90% [1]-[3]. However, current literature exhibits two structural weaknesses: over-reliance on single classifiers (which obscures whether success stems from the feature space or model idiosyncrasies) and a potential vulnerability in purely data-driven feature selection. Specifically, optimization algorithms can inadvertently de-emphasize mutually correlated

* Department of Applied IT and Engineering, Pusan National University
† Corresponding Author : Young-In Kim(email: kimyi@pusan.ac.kr)

Submitted: 2026.06.01. Accepted: 2026.06.06. Confirmed: 2026.06.20.

features [10], [11], risking the marginalization of valid diagnostic markers. This phenomenon is particularly relevant to the delta band (0.5 - 3.0 Hz). While elevated resting-state slow-wave activity is a core ADHD marker supporting the cortical maturational-lag hypothesis [7], [8], recent SHAP-optimized models [2] largely omit delta-band power and entropy features.

To resolve this tension, we propose a robust evaluation framework with threefold contributions: (i) we introduce a five-stage spatial-spectral ablation evaluated across four classifiers (RF, SVM, XGBoost, MLP) to ensure model-independent conclusions; (ii) we extend the feature space to 1,159 dimensions by incorporating the gamma band; and (iii) we isolate the de-prioritized delta-band features, testing them against a label-permuted negative control to provide strong evidence for their standalone diagnostic capacity.

2. Related Work

2.1 EEG-Based ADHD Classification

Work on automated EEG-based ADHD classification falls into three strands. The first couples quantitative-EEG (qEEG) power-spectral indices with shallow classifiers and has historically treated elevated theta power or an elevated theta/beta ratio (TBR) as a single biomarker [7]; meta-analytic evidence, however, indicates that TBR lacks sufficient diagnostic specificity [9]. The second applies end-to-end deep learning to raw EEG or time - frequency representations; Loh et al. [1], for instance, combine channel-wise continuous-wavelet correlation maps from 12-channel EEG with a 2D-CNN to distinguish ADHD, conduct disorder, and their comorbidity, and supply Grad-CAM explanations. The third combines multi-domain features—spectral power, nonlinear dynamics, and functional connectivity—with ensemble classifiers, exemplified by Mao et al. [2] and Kim et al. [3]. The third strand attains the highest reported accuracies but typically depends on a single classifier and on a data-driven selection stage.

2.2 Positioning of This Work

Table 1 situates representative studies on the same public dataset [4], or the same

Table 1. Comparison with prior EEG-based ADHD classification studies
표 1. 기존 EEG 기반 ADHD 분류 연구와의 비교

Study	Dataset	Features	Classifier	Selection	Accuracy
Loh et al. [1] (2024)	Self-collected (12 channels)	CWT correlation	2D-CNN + Grad-CAM	None	98.19% (note)
Mao et al. [2] (2025)	Nasrabadi (public)	Multi-domain (931-dim)	CatBoost	SHAP (206)	99.58%
Kim et al. [3] (2025)	Clinical (168 subj.)	PSD, CFPAC	XGBoost	SHAP	90.81%
This work	Nasrabadi (public)	Multi-domain (1,159-dim)	RF / SVM / XGB / MLP	Hierarchical ablation	98.4-99.6%

ADHD-versus-control task, against the present work. We adopt the multi-domain feature design of Mao et al. [2] but omit the SHAP-based second-stage selection; in its place, we compare four classifiers under one protocol and extend the analyzed spectrum to the gamma band.

3. Materials and Methods

3.1. Dataset

We use the public EEG dataset by Nasrabadi et al. [4] on IEEE DataPort, comprising 61 children with ADHD (DSM-IV diagnosed, methylphenidate-treated ≥6 months) and 60 demographically matched healthy controls (HC) without psychiatric history, aged 7–12. EEG was recorded at 128 Hz from 19 channels (10–20 system) during a visual attention task.

3.2. Preprocessing and Epoching

As outlined in Fig. 1, the raw EEG data were band-pass filtered at 0.5–45 Hz using a 4th-order Butterworth filter. The 45 Hz cutoff was selected to preserve gamma-band activity in the 30–45 Hz range. FastICA [13] was applied as a conservative artifact attenuation step, consistent with the common use of ICA-based methods for EEG artifact reduction [14]. In this study, the single highest-variance independent component was removed to reduce the dominant high-variance artifact source, which may reflect large-amplitude ocular activity. This fixed one-component strategy was adopted to maintain a reproducible pipeline and to avoid the potential subjectivity associated with manual component selection [14]. Furthermore, it prevents excessive component rejection that could inadvertently remove neurophysiologically meaningful low-frequency activity, particularly the delta-band signals central to this study. However, this procedure should not be interpreted as complete removal of all EOG and EMG artifacts. Because dedicated EOG/EMG reference channels and manual ICA component labels were unavailable in the public dataset, residual artifacts may remain, as further discussed in Section 5. Finally, the signals were segmented into 4-second non-overlapping epochs of 512 samples, yielding 4,173 total epochs: 2,330 ADHD epochs and 1,843 healthy control epochs.

3.3. Feature Extraction

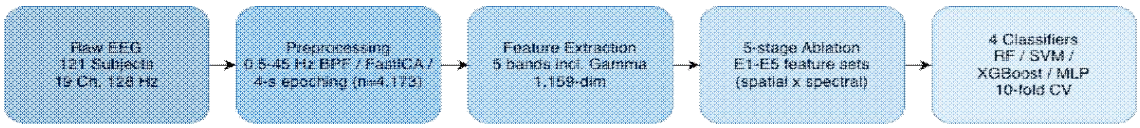

Fig. 1. End-to-end analysis pipeline for EEG-based pediatric ADHD classification
그림 1. EEG 기반 소아 ADHD 분류를 위한 전체 분석 파이프라인

From each 4 s epoch we extract five feature types—absolute power (AP), relative power (RP), theta/beta ratio (TBR), fuzzy entropy (FuzEn), and mutual information (MI)—over five bands: delta (δ, 0.5 - 3 Hz), theta (Θ, 4 - 7 Hz), alpha (α, 8 - 13 Hz), beta (β, 14 - 30 Hz), and gamma (ɣ, 30 - 45 Hz), for 1,159 features in total. Absolute power integrates the Welch power spectral density [12] over each band,

$$AP_b = \int_{f \in b} P(f)\, df, \quad (1)$$

relative power normalizes each band by total power, and TBR is the ratio of theta- to beta-band absolute power. Fuzzy entropy quantifies nonlinear signal complexity [15] with embedding dimension m = 2 and tolerance r = 0.2σ,

$$FuzEn(m, r) = -\ln[\Phi m+1(r) / \Phi m(r)], \quad (2)$$

and mutual information is a nonparametric measure of nonlinear statistical dependence between two channels,

$$I(X;Y) = \Sigma\Sigma\, p(x,y)\, \log[p(x,y)/(p(x)p(y))]. \quad (3)$$

MI is computed for all 171 channel pairs in each of the five bands. Table 2 summarizes the dimensionality of the five feature types.

3.4. Five-Stage Ablation Design

Table 3 defines five feature subsets decomposing the feature space along spatial and spectral axes. E1 (the minimal anchor) contains two features: delta-band AP at O2 and P4. E2 isolates the spatial axis by extending this delta AP to all 19 channels. E3 isolates localized multi-band/multi-type diversity, comprising all 207 features from channels O2 and P4. E4 captures the delta band's overall predictive capacity using all 228 delta features (AP, RP, FuzEn, and MI) across all channels. E5 is the full 1,159-dimensional space.

Table 2. EEG feature types and dimensionality
표 2. EEG 특징 유형 및 차원 수

Type	Definition	Channels × bands	Count
AP	Welch PSD band integral	19 × 5	95
RP	Total-power normalized	19 × 5	95
FuzEn	Fuzzy entropy	19 × 5	95
TBR	Theta/beta power ratio	19 × 1	19
MI	Channel-pair MI	171 × 5	855
Total	–	–	**1,159**

Crucially, these subsets form parallel expansions rather than a single nested chain. E1 is the common origin for E2 (spatial expansion) and E3 (spectral/type expansion). E4 provides a global delta-band slice, and E1 - E4 are all subsets of E5. This parallel architecture enables the independent evaluation of spatial versus spectral contributions.

3.5. Classifiers and Evaluation

We evaluate four classifiers—random forest (n_estimators=300), RBF-SVM, XGBoost [16], and MLP [17], using stratified 10-fold cross-validation (random_state=42). To prevent data leakage, a StandardScaler is fitted exclusively on each training fold. Because all models share identical seeded splits, performance variations reflect algorithmic

differences rather than data partitioning. We report accuracy, precision, recall, F1 score, and AUC (ADHD as the positive class).

Table 3. Definition of the five ablation stages
표 3. 다섯 단계 절제 실험의 정의

ID	Axis probed	Scope	Feat.
E1	Minimal anchor	O2, P4 × δ × AP	2
E2	Spatial axis	19 channels × δ × AP	19
E3	Localized diversity	All features of channels O2, P4	207
E4	Delta slice	19 ch × δ × {AP, RP, FuzEn, MI}	228
E5	Full space	19 ch × 5 bands × 5 types	1,159

4. Results

4.1. Classifier and Feature-Set Comparison

Table 4 and Fig. 2 report 10-fold cross-validation accuracies for all feature-set/classifier pairs. Against a 55.8% majority-class baseline, all four classifiers achieved >98.3% accuracy and >0.998 AUC on the full feature space (E5). This confirms that high classification performance is robust and strictly classifier-independent.

4.2. Spatial Versus Spectral Diversity

E3 (207 features, localized spatial diversity) and E4 (228 features, full delta band) yield comparable accuracy, 88–95%, across all classifiers, but their ordering depends on the classifier (Fig. 3). For RF and XGBoost, E3 slightly exceeds E4 (89.77% vs. 88.09%; 94.61% vs. 94.44%); for SVM and MLP the ordering reverses (92.45% vs. 91.83%; 93.58% vs. 93.31%).

Table 4. Classification accuracy by classifier and feature set (10-fold mean ± SD, %)
표 4. 분류기 및 특징 집합별 분류 정확도(10-fold 평균 ± 표준편차, %)

Feat. set	RF	SVM	XGBoost	MLP
E1 (2)	58.93± 2.06	57.27± 1.09	55.91± 1.96	61.92± 2.08
E2 (19)	79.18± 2.11	63.46± 1.88	80.88± 1.49	75.97± 2.06
E3 (207)	89.77± 1.95	91.83± 1.30	94.61± 0.85	93.31± 1.20
E4 (228)	88.09± 1.88	92.45± 1.39	94.44± 1.26	93.58± 1.25
E5 (1,159)	98.37 ± 0.46	99.31 ± 0.38	99.57 ± 0.35	99.09 ± 0.40

Each cell is the 10-fold cross-validation accuracy (mean ± SD). For E5, the AUC is 0.999 (RF), 1.000 (SVM), 1.000 (XGBoost), and 0.998 (MLP).

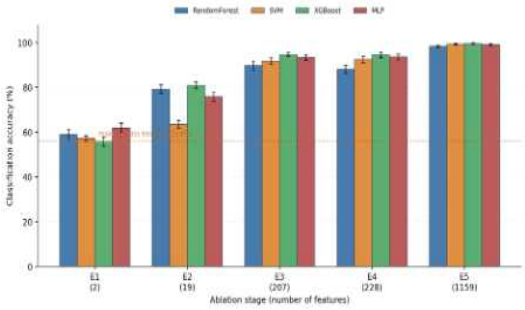

Fig. 2. Classification accuracy across five ablation stages and four classifiers
그림 2. 다섯 단계 절제 실험과 네 가지 분류기에 따른 분류 정확도

All four gaps are within 2 percentage points and overlap with the 10-fold standard

deviation.

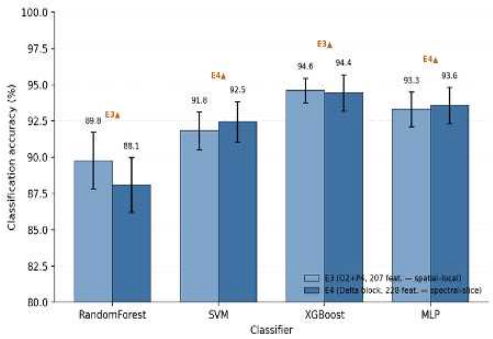

Fig. 3. Accuracy comparison between E3 and E4 by classifier
그림 3. 분류기별 E3와 E4의 정확도 비교

Two points follow. First, the multi-band, multi-type information of just two channels (E3) and the all-channel information of a single band (E4) carry comparable discriminative power for ADHD classification. Second, which axis appears stronger depends on the inductive bias of the classifier, so a general claim such as "spatial diversity outweighs spectral diversity" cannot be drawn from a single classifier. This is precisely why a multi-classifier comparison is needed.

4.3. Standalone Predictive Capacity of Delta Power/Entropy Features

Our central test concerns whether the delta-band power and entropy features de-prioritized by data-driven selectors in fact carry diagnostic information. We isolate the 57 features formed by delta-band AP, RP, and FuzEn (hereafter the E4′ probe), classify ADHD versus HC with them, and compare against a negative control in which the labels are randomly permuted. If the probe AUC clearly exceeds the control AUC, the feature set carries standalone diagnostic information.

Table 5. Cross-validation results for the E4′ probe
표 5. E4′ probe에 대한 교차검증 결과

Classifier	Probe ACC	Probe F1	Probe AUC	Ctrl. AUC	Gain
RF	0.785	0.825	0.868	0.502	+0.366
SVM	0.728	0.783	0.792	0.497	+0.295
XGBoost	0.828	0.853	0.911	0.492	+0.420
MLP	0.746	0.780	0.811	0.505	+0.307
Mean	**0.772**	**0.810**	**0.846**	**0.499**	**+0.347**

Table 5 and Fig. 4 show the E4′ probe yields an AUC of 0.792 - 0.911 (mean 0.846) across classifiers, far exceeding chance-level permuted controls (mean 0.499). This consistent AUC gain (mean +0.347) confirms the strong, classifier-independent diagnostic value of delta-band power and entropy features that may be de-prioritized by data-driven selectors.

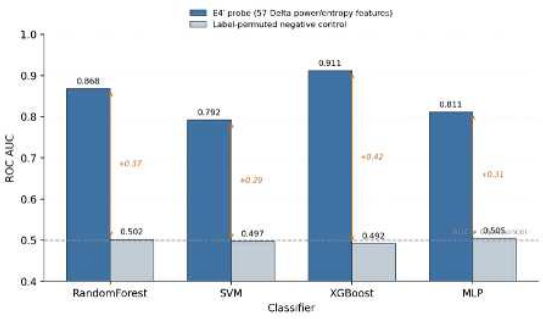

Fig. 4. AUC comparison between the E4′ probe and label-permuted control
그림 4. E4′ probe와 라벨 무작위 치환 대조군의 AUC 비교

Because these 57 features comprise the full delta block, they may not perfectly match the exact subset excluded by Mao et al. [2]. Thus, this represents an upper-bound test of the block's predictive capacity, leaving precise comparisons for future work (Section 5).

4.4. ROC Curves and Confusion Matrices

Fig. 5 and Fig. 6 display the 10-fold aggregated ROC curves for the random forest and XGBoost classifiers, respectively.

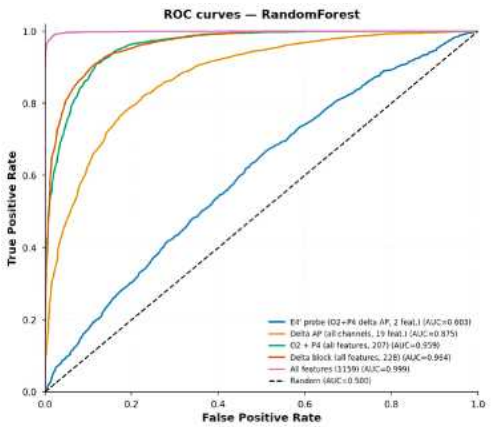

Fig. 5. ROC curves of the random forest classifier
그림 5. 랜덤 포레스트 분류기의 ROC 곡선

Fig. 7 and Fig. 8 present the corresponding row-normalized confusion matrices. For both classifiers, AUC increases with feature expansion, rising from the minimal E1 feature set to the full E5 feature space. These additional XGBoost visualizations further confirm that the strongest-performing classifier follows the same overall feature-expansion trend observed in the random forest results.

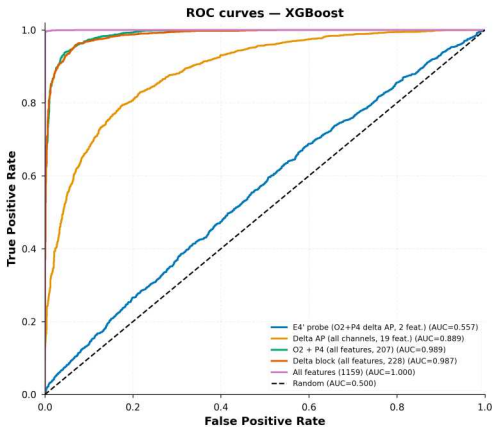

Fig. 6. ROC curves of the XGBoost classifier
그림 6. XGBoost 분류기의 ROC 곡선

5. Discussion

This study re-evaluates the diagnostic value of the delta band in pediatric ADHD, contextualizing its frequent omission by data-driven feature selectors [2]. Our five-stage ablation shows that the 57 delta-band power and entropy features yield strong standalone predictive capacity (mean AUC 0.846) across four distinct classifiers, far exceeding a label-permuted control. This aligns with the well-established cortical maturational-lag hypothesis, wherein increased resting-state slow-wave activity reflects a delayed developmental trajectory in ADHD [7], [8].

We propose that the omission of delta features in prior optimal subsets [2] may be partially attributed to the inherent mechanics of certain importance-based selectors, which often inadvertently mask the contribution of highly correlated features [10], [11]. By bypassing this

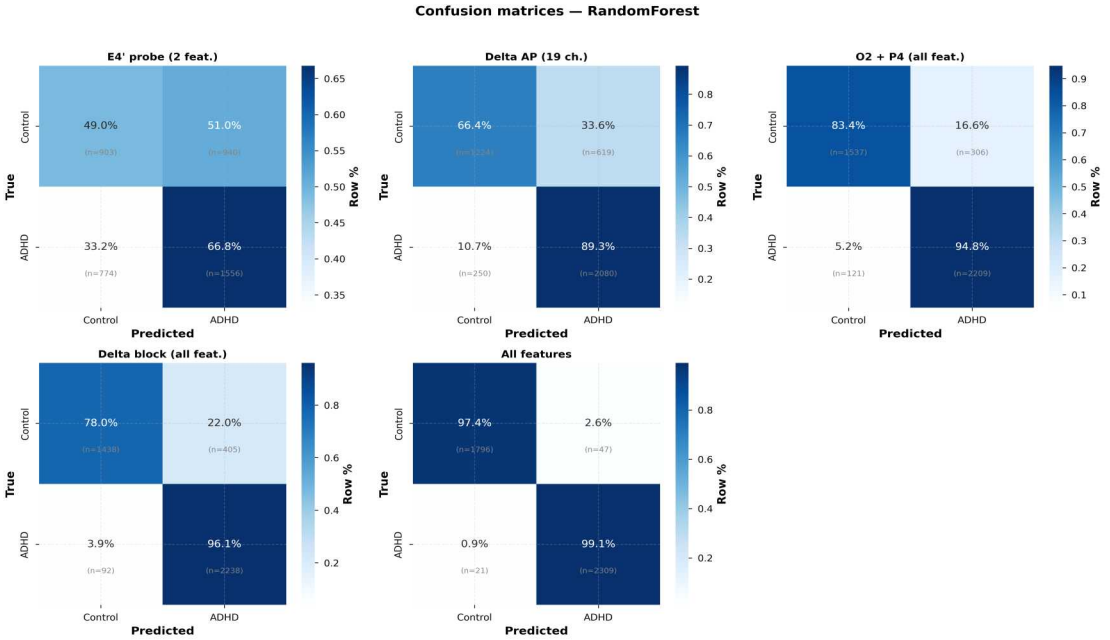

Fig. 7. Confusion matrices of the random forest classifier
그림 7. 랜덤 포레스트 분류기의 혼동행렬

selection step, we demonstrate that models can achieve 98.4–99.6% accuracy independently of classifier choice. This highlights a critical methodological implication: mathematically optimized feature subsets must be balanced against neurophysiological interpretability.

Our findings are constrained by epoch-level cross-validation, which risks subject-level data leakage and may inflate absolute accuracies relative to leave-one-subject-out protocols [3]. In addition, the simplified FastICA-based artifact attenuation strategy may not fully remove all EOG and EMG artifacts, because only the single highest-variance component was removed and no dedicated EOG/EMG reference channels or manual ICA component labels were available in the public dataset. Future work will employ subject-level cross-validation across diverse clinical cohorts and incorporate more rigorous artifact-correction procedures, such as EOG/EMG-assisted ICA, automated component classification, and manual expert inspection, to further validate the delta band's role as a robust, interpretable biomarker.

6. Conclusion

We examined the tension by which data-driven feature selectors de-prioritize the delta band in EEG-based pediatric ADHD classification, and we analyzed it with a five-stage spatial–spectral ablation run under four classifiers. On a public dataset [4] with a 1,159-dimensional, five-band feature space we

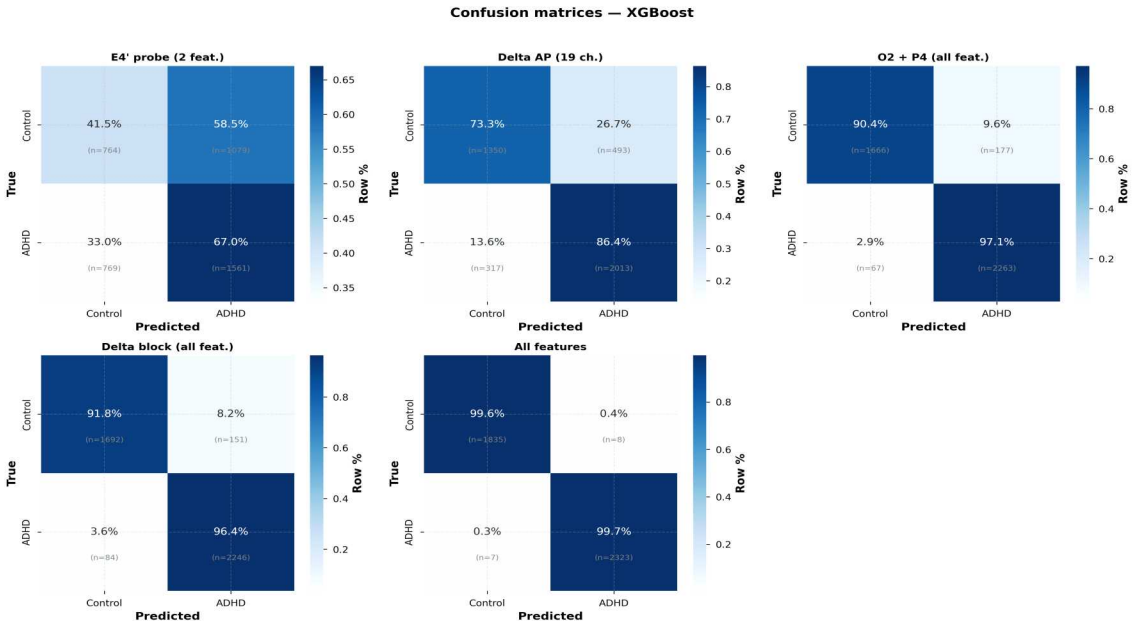

Fig. 8. Confusion matrices of the XGBoost classifier
그림 8. XGBoost 분류기의 혼동행렬

found that (i) the full-feature model reaches 98.4 - 99.6% accuracy for RF, SVM, XGBoost, and MLP alike, so high performance is independent of classifier choice; (ii) localized spatial diversity (E3) and the full delta band (E4) yield comparable 88 - 95% accuracy, with the ordering of the two axes depending on the classifier; and (iii) the 57 delta-band power/entropy features, which may be de-prioritized in data-driven feature-selection pipelines, achieve a mean AUC of 0.846 across the four classifiers, far above the label-permuted control at 0.499.

These findings provide strong evidence that the delta band retains standalone predictive value for ADHD classification and support the need to evaluate data-driven optimization results alongside neurophysiological interpretability. The epoch-level cross-validation limits external generalization of the absolute accuracies, and subject-level validation together with a precise comparison against the excluded features of [2] remains for future work.

This work is based on the author's master's thesis at Pusan National University.

References

[1] H. W. Loh, C. P. Ooi, S. L. Oh, P. D. Barua, Y. R. Tan, U. R. Acharya, and D. S. S. Fung, "ADHD/CD-NET: automated EEG-based characterization of ADHD and CD using explainable deep neural network technique,"Cognitive Neurodynamics, vol. 18, no. 4, pp. 1609-1625, Aug. 2024, DOI : https://doi.org/10.1007/s11571-023-10028-2

[2] Y. Mao, X. Qi, L. He, S. Wang, Z. Wang,

and F. Wang, "Advanced machine learning techniques reveal multidimensional EEG abnormalities in children with ADHD: a framework for automatic diagnosis," Frontiers in Psychiatry, vol. 16, art. 1475936, 2025,
DOI : https://doi.org/10.3389/fpsyt.2025.1475936

[3] J. W. Kim, B. N. Kim, J. I. Kim, C. M. Yang, and J. Kwon, "Electroencephalogram (EEG) based prediction of attention deficit hyperactivity disorder (ADHD) using machine learning,"Neuropsychiatric Disease and Treatment, vol. 21, pp. 271-279, Feb. 2025,
DOI : https://doi.org/10.2147/NDT.S509094

[4] A. M. Nasrabadi, A. Allahverdy, M. Samavati, and M. R. Mohammadi, "EEG data for ADHD/control children," IEEE DataPort, 2020,
DOI : https://doi.org/10.21227/rzfh-zn36

[5] K. Sayal, V. Prasad, D. Daley, T. Ford, and D. Coghill, "ADHD in children and young people: prevalence, care pathways, and service provision,"The Lancet Psychiatry, vol. 5, no. 2, pp. 175-186, Feb. 2018,
DOI: https://doi.org/10.1016/S2215-0366(17)30167-0

[6] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 5th ed., Washington, DC: American Psychiatric Publishing, 2013, DOI:https://doi.org/10.1176/appi.books.9780890425596

[7] R. J. Barry, A. R. Clarke, and S. J. Johnstone, "A review of electrophysiology in attention-deficit/hyperactivity disorder: I. Qualitative and quantitative electro encephalography," Clinical Neurophysiology, vol. 114, no. 2, pp. 171-183, Feb. 2003,
DOI : https://doi.org/10.1016/s1388-2457(02)00362-0

[8] A. R. Clarke, R. J. Barry, and S. Johnstone, "Resting state EEG power research in attention-deficit/hyperactivity disorder: a review update,"Clinical Neurophysiology, vol. 131, no. 7, pp. 1463-1479, Jul. 2020,
DOI : /10.1016/j.clinph.2020.03.029

[9] M. Arns, C. K. Conners, and H. C. Kraemer, "A decade of EEG theta/beta ratio research in ADHD: a meta-analysis," Journal of Attention Disorders, vol. 17, no. 5, pp. 374-383, Jul. 2013, DOI : https://doi.org/10.1177/1087054712460087

[10] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems, vol. 30, Long Beach, CA, 2017, pp. 4765-4774, DOI : https://doi.org/10.48550/arXiv.1705.07874

[11] C. Strobl, A. L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests," BMC Bioinformatics, vol. 9, art. 307, 2008, DOI : https://doi.org/10.1186/1471-2105-9-307

[12] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms,"IEEE Transactions on Audio and Electroacoustics, vol. 15, no. 2, pp. 70-73, Jun. 1967, DOI : https://doi.org/10.1109/TAU.1967.1161901

[13] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," Neural Networks, vol. 13, no. 4-5, pp. 411-430, Jun. 2000,
DOI: https://doi.org/10.1016/S0893-6080(00)00026-5

[14] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from EEG signals: a review," Sensors, vol. 19, no. 5, art. 987, Mar. 2019,
DOI : https://doi.org/10.3390/s19050987

[15] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy,"IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 15, no. 2, pp. 266-272, Jun. 2007, DOI : https://doi.org/10.1109/TNSRE.2007.897025

[16] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system,"in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, 2016, pp. 785-794, DOI : https://doi.org/10.1145/2939672.293978

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., "Scikit-learn: machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011, DOI : https://doi.org/10.48550/arXiv.1201.0490

Authors

Fika Fauzia

2010.8 BS in Mathematics, University of Indonesia

2024.9 - present MS Candidate in Applied IT & Engineering, Pusan National University

<Research interests> Artificial Intelligence, Machine Learning, Data Mining

Young-In Kim

1996.2 Ph.D. in Computer Engineering, MyongJi University

1996.4-2006.2 Faculty member in Miryang National University

2007.8-2008.7 University of Missouri Visiting Scholar

2006.3-present Faculty member in Pusan National University

<Research interests> Database Systems, Data Mining, Machine Learning