Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.38708

From EEG Classification to Personalized Neurofeedback: An Explainable AI Framework for Alzheimer's Disease

Amy Zi-Wei Li

Quarry Lane School, Dublin, USA
amy.ziwei.li@gmail.com

Abstract. This study developed an explainable machine learning (ML) framework for Alzheimer's disease classification using resting-state EEG-BCI recordings from the OpenNeuro ds004504 cohort, which consists of 36 patients with AD and 29 Healthy Control (HC). Four EEG biomarkers and a normative-deviation score are integrated with feature weighting and logistic classification, while SHAP, gradient saliency, and layer-wise relevance propagation provide complementary explanations of model predictions. The performance of the framework was evaluated using stratified 5-fold cross-validation and a locked 15-subject holdout set (8 AD, 7 HC). The model achieved a mean cross-validated area under the receiver operating characteristic curve (ROC-AUC) of 0.823 ± 0.095 and a holdout ROC-AUC of 0.929. The Explainable AI (XAI) feature attribution identified posterior alpha power, alpha peak frequency, and frontal theta/beta ratio as significant discriminators between AD and HC ($p < 0.05$), with posterior alpha power showing the strongest association with classification ($p < 0.001$). Posterior alpha power alone achieved a ROC-AUC of 0.900 ± 0.074. These results demonstrate the feasibility of using explainable AI to extract important spectral features that can distinguish AD from healthy controls and may inform the future development of adaptive neurofeedback protocols tailored to individual brain patterns. However, a larger and more diverse data set is needed for future studies of real-time, personalized neurofeedback applications.

Keywords: Machine Learning, EEG-BCI, Neurofeedback, Explainable AI (XAI), Alzheimer's disease

1. Introduction

Neurofeedback is a noninvasive approach that may help support cognitive function in older adults. People can learn how to adjust their activities to stimulate certain brain functions based on the use of real-time electroencephalographic (EEG) signals [1]. Previous studies have shown its improvements in attention, working memory, and executive function. Traditionally neurofeedback method loop method is used to train people to adjust their EEG patterns, for example, by increasing alpha activity or reducing theta activity [2, 3].

However, this approach may not be efficient for everyone. EEG abnormality can be a result of normal aging, and it can also be more significant in neurodegenerative diseases. EEG signals include slower alpha peak frequency, lower posterior alpha power, increased theta activity, and

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.38708

reduced functional connectivity are often being studied for its role in cognitive change [4]. Because these signal changes can vary significantly from person to person, generic neurofeedback protocols can easily miss a patient's distinct neurophysiological profile.

Machine learning (ML) and deep learning (DL), such as convolutional neural networks, autoencoders, and attention-based architectures, can be used to extract subtle temporal and spectral dynamics beyond conventional spectral analysis [5]. These computational techniques have increasingly been used to classify Alzheimer's disease (AD) and mild cognitive impairment (MCI). However, many high-performing ML and DL algorithms function as opaque "black boxes," making it difficult to interpret how decisions are reached whether those. This lack of transparency poses a significant barrier in clinical setting.

Explainable artificial intelligence (XAI) can bring transparency to the models by identifying which input features most strongly contribute to prediction. Methods such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) estimate each feature's contribution to a classification output [6, 7]. The XAI has enabled researchers to identify features that may be linked to cognitive decline, such as alpha peak frequency, posterior alpha power, theta-to-beta ratio, and interhemispheric coherence. However, there is not enough study to show how the information could be used for adaptive neurofeedback.

To address this gap, we present an explainable machine learning framework designed to characterize key EEG biomarkers from AD group. Ultimately the work enables mapping these interpretable biomarkers directly into adaptable neurofeedback protocols, establishing a concept and practical bridge between explainable diagnostic modeling and patient-tailored brain training.

2. Method

2.1. Dataset

Resting-state EEG recordings were downloaded from the OpenNeuro ds004504 [8], which contains recordings from participants with AD, frontotemporal dementia (FTD), and cognitively healthy controls (HC). Because this study focused on binary AD classification, this work excluded the 23 participants with FTD, leaving 65 subjects (36 with AD and 29 with HC). The EEG data was recorded using a standard 19-channel international 10-20 montage at a 500 Hz sampling rate during eyes-closed resting-state acquisition. Original dataset clinical diagnoses were used as ground-truth classification labels. The EEG recordings were already preprocessed using band-pass filtering, re-referencing, Artifact Subspace Reconstruction (ASR), and Independent Component Analysis (ICA). Non-neural components including eye movements, muscle, cardiac, and channel noise were removed before reconstructing the signals.

2.2. EEG preprocessing

Th For consistent spectral analysis across subjects, cleaned EEG signals were band-pass filtered (1–40 Hz), notch-filtered at 50 Hz, and segmented into non-overlapping 4-second epochs. Epochs with peak-to-peak amplitudes exceeding 150 μV were discarded to remove residual artifacts. Power spectral density features were then derived from the remaining epochs via Welch's method.

DOI: 10.54254/2755-2721/2026.38708

2.3. Feature extraction

Welch's method was used to calculate the power spectral density (PSD) for each remaining epoch, showing the amount of EEG signal power at different frequencies. Four EEG biomarkers were calculated from the PSD data: alpha peak frequency, posterior alpha power, frontal theta/beta ratio, and interhemispheric alpha coherence. Specific frequency ranges and EEG channels were used for each biomarker. Frontal theta and beta power were also included as additional features in the normative-deviation model.

The four EEG biomarkers were calculated as follows:

- Alpha peak frequency - Evaluated across posterior electrode channels (P7, P3, Pz, P4, P8, O1, O2) by computing the mean power spectral density (PSD) and identifying the frequency corresponding to maximum power within the alpha band (8-12 Hz).
- Posterior alpha power - Evaluated as the average spectral power within the 8-12 Hz alpha frequency range across the posterior channel group (P7, P3, Pz, P4, P8, O1, O2).
- Frontal theta/beta ratio Calculated across frontal electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8) by dividing mean theta power (4-8 Hz) by mean beta power (13-30 Hz).
- Interhemispheric alpha coherence - Estimated across homologous left-right electrode pairs (e.g., F3-F4, P3-P4) using 2-second sliding windows, averaged across the alpha band (8-12 Hz) and across all retained epochs.

In addition, two supporting quantities, frontal theta and beta power values were retained as auxiliary inputs for the normative-deviation model described in section 2.4. However, they are not treated as biomarkers in this study.

2.4. Normative score and ML classification

Before training the model, the EEG biomarkers were standardized using z-scores. The mean and standard deviation were calculated from the training data. The classification model had three main steps: calculating normative deviations, weighing the EEG features using attention, and classifying participants into two groups.

2.4.1. Normative deviation encoding

A bottleneck autoencoder was trained on EEG biomarkers from the HC group only, following the Variational Information Bottleneck approach. The model learned a smaller representation of the EEG patterns found in healthy participants. The encoder consisted of two tanh-activated hidden layers (8→4→8 units) with L_2 regularization and was optimized using backpropagation. For each participant, the latent features were compared with the HC training group. The z-scores were averaged across the latent features to produce a normative deviation score, which represented the difference from the healthy EEG pattern.

2.4.2. Feature weighting & classification

The four EEG biomarkers and the normative deviation score were passed through an attention - based feature weighting layer followed by a logistic regression classifier to estimate the probability of AD.

The model was tested using a stratified 70/30 train-test splitting and stratified 5-fold cross validation. A separate set of 15-participants holdout set (8 AD, 7HC) was kept out of model

DOI: 10.54254/2755-2721/2026.38708

development and used for the final evaluation.

2.5. Explainable artificial intelligence

To evaluate the interpretability of model decisions, three complementary Explainable AI methods - SHapley Additive exPlanations (SHAP), gradient-based saliency, and Layer-wise Relevance Propagation (LRP) were used to quantify individual feature contributions to the model's predictions. The three methods were then compared to see whether they identified the same biomarkers as important.

2.5.1. SHAP

SHAP values were calculated using *KernelExplainer*. For each participant, the values showed which EEG biomarkers are the driving factor for AD. All EEG biomarkers were standardized (z -scored) prior to attribution. The SHAP values quantified the magnitude and direction (+/−) by which each biomarker shifted the predicted probability relative to a background reference baseline. Following attribution computation, feature values were transformed back into their original physical units (e.g., Hz, $\mu V^2/Hz$) to maintain physiological relevance for clinical review.

2.5.2. Gradient-based saliency

Gradient-based saliency was used to measure how the model's prediction changed when there is small shift in each EEG biomarker [9]. Feature gradients were computed numerically using central finite differences and normalized by each biomarkers' standard deviation to ensure cross-biomarker comparability. Unlike SHAP, which compares a feature with a baseline, saliency looks at how the prediction changes around an individual participant's feature values.

2.5.3. Layer-wise relevance propagation (LRP)

We implemented LRP through the ε-LRP rule [10]. Relevance scores were traced backward from the classifier output through the logistic regression layer, feature-weighting layer, and bottleneck encoder to measure how much each biomarker contributed to each subject's prediction [11].

2.5.4. Deriving individualized neurofeedback targets

We ran attribution scoring for each subject and biomarker across all XAI methods. This step gave us both group-level averages (global explanations) and subject-specific profiles (local explanations). Section 2.6 details the statistical differences between AD and HC groups. Biomarkers showing significant group differences in both SHAP and LRP were selected as target candidates for personalized neurofeedback. Gradient saliency differed from the other methods, identifying interhemispheric alpha coherence as its only significant biomarker; as a result, we analyzed it separately. Overall, these XAI profiles test whether target selection is feasible, rather than proving therapeutic success.

DOI: 10.54254/2755-2721/2026.38708

2.6. Statistical analysis

2.6.1. Cross-validation and performance evaluation

The performance across 5-fold cross-validation is presented as mean ± standard deviation. ROC-AUC metrics and 95% confidence intervals were evaluated for both cross-validation and holdout splits, using a *t*-distribution (df = 4) for fold means and nonparametric bootstrapping (2,000 resamples) for the holdout split. Confusion matrices creates the summary of classifier's performance at each stage.

2.6.2. Explainability & attribution analysis

Signed attribution scores from SHAP, gradient saliency, and LRP were compared between the AD and HC groups using a two-sided Mann-Whitney U test. This test was used because the groups were small and unequal in size (36 AD and 29 HC), and normality was not assumed. Group means are shown with standard error bars.

A separate ablation analysis was used to test all 15 possible nonempty combinations of the four EEG biomarkers. Accuracy and ROC-AUC were calculated using the same 5-fold cross-validation procedure. The results were used to examine which biomarkers had the greatest effect on classifier performance and to compare them with the features identified by the XAI methods.

3. Results

3.1. Classification performance

This work evaluated the classifier using three complementary protocols: a single 70/30 train–test split, 5-fold cross-validation, and an independently locked holdout set (see Table 1).

In the single 70/30 split, the model achieved an accuracy of 0.850, precision of 1.000, recall of 0.727, F1 score of 0.842, and ROC-AUC of 0.869 (95% bootstrap CI: 0.646–1.000). Across the five cross-validation folds, the model achieved a mean accuracy of 0.769 ± 0.094, precision of 0.817 ± 0.140, recall of 0.775 ± 0.083, F1 score of 0.789 ± 0.079, and ROC-AUC of 0.823 ± 0.095; 95% CI: 0.703–0.940).

Final generalization was assessed once on the locked 15-subject holdout set (n_{AD} = 8 and n_{HC} = 7), which remained untouched during model architecture and hyper parameter tuning. The model achieved an accuracy of 0.800, precision of 0.857, recall of 0.750, F1 score of 0.800, and ROC-AUC of 0.929. Specifically, the model correctly classified six of eight AD participants and six of seven HCs. The two false-negative AD cases received predicted AD probabilities of 0.42 and 0.43, whereas the single false-positive healthy control received a predicted AD probability of 0.685. Performance remained consistent across validation tiers, though estimates should be interpreted with appropriate caution given the small overall sample size (N = 65).

Table 1. Classifier performance under three evaluation protocols (n = 65: 36 AD and 29 HC; locked holdout n = 15). Cross-validation results are reported as mean ± standard deviation across five folds

Metric	Single 70/30 split	5-fold CV	Locked holdout
Accuracy	0.850	0.769 ± 0.094	0.800
Precision	1.000	0.817 ± 0.140	0.857
Recall	0.727	0.775 ± 0.083	0.750

DOI: 10.54254/2755-2721/2026.38708

Table 1. (continued)

F1	0.842	0.789 ± 0.079	0.800
ROC-AUC	0.869 (95% CI 0.646-1.000)	0.823 ± 0.095 (95% CI 0.703-0.940)	0.929 (95% CI 0.732-1.000)

3.2. Explainability: biomarker significance

Table 2 reports the group-mean signed attribution and the AD-vs-HC significance (Mann–Whitney U, two-sided, uncorrected) for each of the three explanation methods. Using SHAP and LRP, posterior alpha power, alpha peak frequency, and frontal theta/beta ratio were identified as significant discriminators ($p < 0.05$ for all three under both methods), with posterior alpha power the strongest in each ($p < 0.001$). Interhemispheric alpha coherence was not significant under either method (SHAP $p = 0.531$; LRP $p = 0.900$).

In contrast, gradient-based saliency identified interhemispheric coherence as the only significant biomarker ($p = 0.031$), whereas SHAP and LRP showed attribution sign differences between groups; the saliency gradient for each biomarker had the same sign in both groups.

3.3. Predictive performance of biomarker combinations (ablation analysis)

An ablation analysis (5-fold CV) evaluated all 15 non-empty combinations of the four biomarkers by replacing excluded biomarkers with their training-fold mean values (see Table 3). The single-biomarker model using posterior alpha power alone achieved the highest ROC-AUC of any subset tested (0.900 ± 0.074), exceeding the full four-biomarker model (0.823 ± 0.095), as shown in Figure 1 and Figure 2. The best performing two-biomarker combination (frontal theta/beta ratio + posterior alpha power) achieved an accuracy of 0.785 and an ROC-AUC of 0.880, exceeding the four-biomarker baseline on both metrics.

Table 2. Group-level signed biomarker attributions and Mann–Whitney U comparisons between AD and HC participants for SHAP, saliency, and LRP

Biomarker	SHAP (AD/HC, p)	Saliency (AD/HC, p)	LRP (AD/HC, p)
Posterior alpha power	+0.037/−0.026, p < 0.001***	−0.061/−0.064, p = 0.276 ns	+0.094/−0.159, p < 0.001***
Alpha peak frequency	+0.031/−0.008, p = 0.002**	−0.038/−0.037, p = 0.380 ns	+0.042/−0.075, p = 0.005**
Frontal theta/beta ratio	+0.040/+0.005, p = 0.009**	+0.054/+0.066, p = 0.232 ns	+0.082/−0.111, p = 0.013*
Interhemispheric coherence	+0.004/+0.005, p = 0.531 ns	−0.021/−0.033, p = 0.031*	−0.014/−0.028, p = 0.900 ns

Note: Signed feature attributions are reported as AD cohort mean / HC cohort mean. *p*-values reflect two-sided Mann-Whitney U comparisons between groups ($n_{AD} = 36$, $n_{HC} = 29$). Significance thresholds: ***p < 0.001; **p < 0.01; *p < 0.05; $^{ns}p \geq 0.05$

Table 3. Selected biomarker subsets evaluated using 5-fold cross-validation. Accuracy and ROC-AUC are reported as mean values across folds. The complete ranking of all 15 subsets is provided in results/biomarker_sensitivity.csv

Rank	Biomarker subset	Accuracy, mean ± SD	ROC-AUC, mean ± SD
1	Frontal theta/beta ratio + posterior alpha power	0.785 ± 0.100	0.880 ± 0.079
2	All four biomarkers (baseline)	0.769 ± 0.094	0.823 ± 0.095
3	Posterior alpha power + Alpha peak frequency	0.769 ± 0.122	0.857 ± 0.095
4	Drop Interhemispheric alpha coherence	0.769 ± 0.144	0.837 ± 0.065
5	Posterior alpha power only	0.754 ± 0.084	0.900 ± 0.074

DOI: 10.54254/2755-2721/2026.38708

Table 3. (continued)

6	Drop Alpha peak frequency	0.738 ± 0.103	0.871 ± 0.080

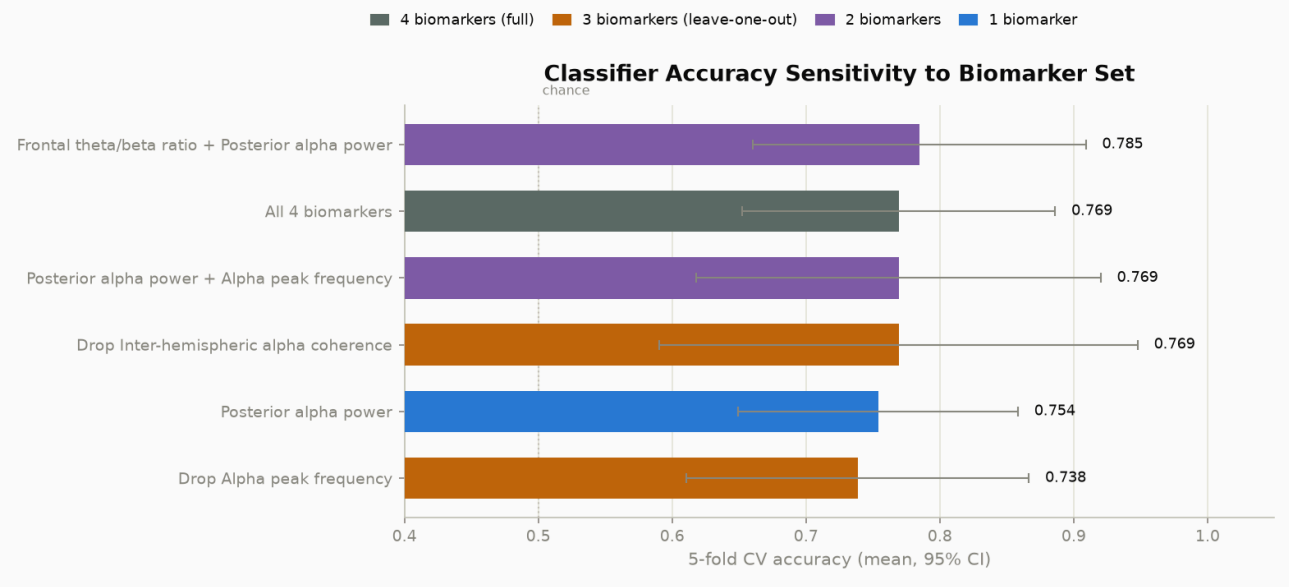

Figure 1. Classification accuracy for different biomarker combinations

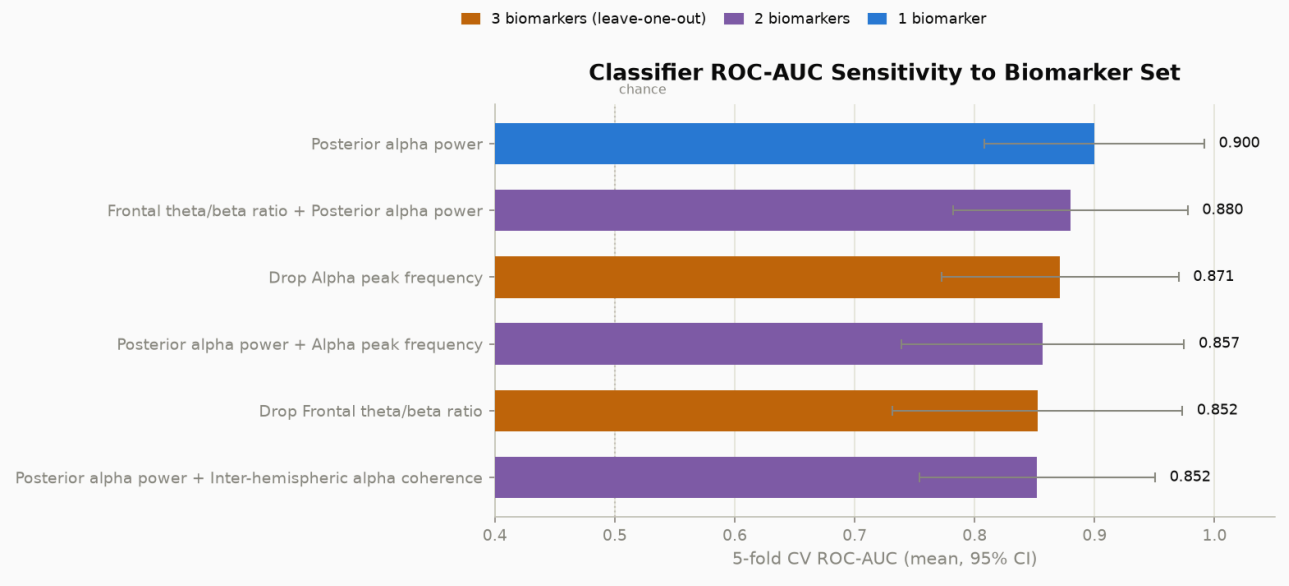

Figure 2. ROC-AUC for different biomarker combinations

3.4. Feasibility of explainability-guided personalized biomarker selection

The model result offers an insight into personalized biomarker selection based on feature importance. The selected biomarkers show a concordance of 96.9% between the feature weighting layer and local SHAP attributions in the 65 cases. The two non-concordant cases both occurred in healthy control subjects and showed minor weighting shifts in the frontal theta/beta ratio [12].

These EEG profiles can serve as references for clinicians when choosing neurofeedback targets. The advantage of it is that the treatment method can be customized based on everyone's EEG pattern

DOI: 10.54254/2755-2721/2026.38708

rather than using a generic one. Researchers can also monitor the changes in these biomarkers longitudinally during rehabilitation. However, this study did not test whether this approach could improve cognitive outcomes in real cases. Cognitive assessments and measurements before and after rehabilitation are needed to determine the efficiency of the method in practice.

4. Discussion

This study demonstrates a framework of using explainable artificial intelligence (XAI) method for EEG-BCI biomarkers' classification and interpretation in Alzheimer's disease research. The method can transform the traditional "black box" prediction method into a transparent and reliable tools for AD prognosis and longitudinal monitor.

Across additive SHAP and structural LRP, posterior alpha power spectral signal emerged as the primary discriminative biomarker for AD classification, followed by alpha peak frequency and the frontal theta/beta ratio. From the concordance between SHAP and LRP attributions, we can confirm that the analysis is performed based on reliable neurophysiological signals rather than other artifacts.

The gradient-based saliency generated a different attribution pattern, where the interhemispheric alpha coherence is found to be the only significant feature ($p = 0.031$) while the other three did not show significance. These results demonstrated great alignment with the previous clinical study results that alpha slowing and spectral shifts are the main characteristics of early neurodegenerative decline [13, 14].

From the feature ablation study, we found that a larger feature set does not always generate better performance. The classifier trained with posterior alpha power alone achieved the highest ROC-AUC value of 0.900; it outperforms the baseline model that used all four biomarkers (ROC-AUC of 0.823). The small data set used for machine learning model training could be the limitation in generating a more reliable and reproducible result. Features such as interhemispheric coherence did not improve performance when combined with stronger biomarkers. However, this feature attribution plays an important role in model optimization and feature selection. By removing uninformative inputs from the data set, researchers can build a cleaner interpretable model with better classification accuracy.

Beyond the application for classification purposes, the individual feature attribution can provide a structured framework for designing personalized neurofeedback and brain-computer interface (BCI) protocols. Conventional neurofeedback protocols only apply population-average training targets (such as generic alpha enhancement or theta suppression) across all participants, the method ignores the substantial inter-individual variability introduced by age and disease related spectral dynamics. By implementing the explainability AI models into the real-time feature attribution discovery, clinicians can tailor their target selection to accommodate patient specific spectral fluctuations.

Several methodological limitations should be considered when interpreting these findings. First, the dataset included only 65 participants. A larger dataset would be needed for machine learning model training. Second, only resting-state EEG was analyzed. Other task-based EEG would also be worth studying, particularly when the framework is applied to neurofeedback. Third, the model used fixed feature weights to the standardized EEG biomarkers, this limits its ability to capture more complex brain signal changes for different. In future studies, we recommend using the personalized neurofeedback loop training in clinical practice while comparing its effectiveness in cognitive score and learning rate compared to generalized protocol.

DOI: 10.54254/2755-2721/2026.38708

5. Conclusion

This study establishes an explainable machine learning framework that not only classifies Alzheimer's disease through resting-state EEG spectral features but also identify patient specific targets for personalized neurofeedback.

The methodology combines attention-based feature weighting with three complementary feature attribution algorithms - SHapley Additive exPlanations (SHAP), Layer-wise Relevance Propagation (LRP), and gradient-based saliency. It provides a transparent and individualized interpretation of the brain signals from patients. SHAP and LRP attributions consistently identified posterior alpha power, alpha peak frequency and the frontal theta/beta ratio as the primary discriminative biomarkers separating ADs and Healthy Controls. In contrast, gradient-based saliency shows a different attribution pattern, which clearly suggests evaluating different models are necessary to prevent any analysis bias toward a specific method.

The feature ablation method demonstrated that a single biomarker model trained solely on posterior alpha power can achieve the highest ROC-AUC score and its result is superior to other multiple biomarkers combined models. At the individual level, we compared the feature attribution profiles between internal model feature weights and SHAP explanations, the concordance rate is at 96.9%, which suggests the technical feasibility for identifying unique brain signal patterns.

Overall, this framework offers a practical application of XAI that beyond AD classification, it can successfully identify EEG features that contribute to model prediction and thus be selected as the biomarkers for further neurofeedback loop training. With further validation, this approach could help guide the development of personalized BCI-EEG neurofeedback treatment for people with cognitive impairment.

References

[1] Jiang, Y., Abiri, R., & Zhao, X. (2017). EEG-based neurofeedback for cognitive enhancement in older adults. *Frontiers in Aging Neuroscience, 9*, 52.

[2] Belkacem, A. N., Jamil, N., Palmer, J., Ouhbi, S., & Chen, C. (2020). Brain–computer interfaces for neurorehabilitation: Traditional protocols and beyond. *Frontiers in Neuroscience, 14*, 692.

[3] Gómez-Pilar, J., Corralejo, R., Nicolás-Alonso, L. F., Álvarez, D., & Hornero, R. (2016). Assessment of cognitive training effects induced by an EEG-based brain-computer interface. *Medical & Biological Engineering & Computing, 54*, 1655–1666.

[4] Leroy, S., Bublitz, V., von Dincklage, F., Antonenko, D., & Fleischmann, R. (2024). Age-related alterations in resting-state EEG dynamics. *medRxiv*.

[5] Grillenberger, A. J., Shenton, N., Lauritzen, M., Benedek, K., & Puthusserypady, S. (2026). Deep learning frameworks for EEG-based diagnosis of Alzheimer's disease. *Computers in Biology and Medicine, 204*, 111538.

[6] Alatrany, A. S., Khan, W., Hussain, A., Kolivand, H., & Al-Jumeily, D. (2024). Explainable artificial intelligence for interpretability in neural engineering. *Scientific Reports, 14*, 2637.

[7] Rajpura, P., Cecotti, H., & Meena, Y. K. (2024). Feature attribution methods in brain signal classification. *Journal of Neural Engineering, 21*, 041003.

[8] Miltiadous, A., Tzimourta, K. D., Afrantou, T., Ioannidis, P., Grigoriadis, N., Tsalikakis, D. G., Angelidis, P., Tsipouras, M. G., Glavas, E., Giannakeas, N., & Tzallas, A. T. (2023). A dataset of EEG recordings from patients with Alzheimer's disease, frontotemporal dementia, and healthy controls. *OpenNeuro*.

[9] Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv Preprint*, arXiv:1312.6034.

[10] Bach, S., Binder, A., Montavon, G., Klauschen, F., Samek, W., & Müller, K. R. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLOS ONE, 10*, e0130140.

[11] Nam, H., Kim, J. M., Choi, W., Bak, S., & Kam, T. (2023). Relative attributions: Towards explaining natural language and deep learning models. *Frontiers in Human Neuroscience, 17*, 1205881.

DOI: 10.54254/2755-2721/2026.38708

[12] Lombardi, A., Diacono, D., Amoroso, N., Biecek, P., Monaco, A., Bellantuono, L., Pantaleo, E., Logroscino, G., Blasi, R. D., Tangaro, S., & Bellotti, R. (2022). Intra-group feature importance variability in pathological cohorts. *Brain Informatics, 9*, 17.

[13] Babiloni, C., Del Percio, C., Boccardi, M., et al. (2015). Occipital sources of resting-state alpha rhythms are related to local gray matter density in subjects with amnesic mild cognitive impairment and Alzheimer's disease. *Neurobiology of Aging, 36*(2), 556–570.

[14] Vecchio, F., Babiloni, C., Lizio, R., Fallani, F. D. V., Blinowska, K. J., Verrienti, G., Frisoni, G. B., & Rossini, P. M. (2013). Resting state cortical connectivity in Alzheimer's disease. *Supplements to Clinical Neurophysiology, 62*, 223–236.