

# Predictive EEG Signatures of Alzheimer's Disease: A Cross-Validated, Cross-Method Analysis Using ANCOVA Filtering and Sparse Logistic Regression

Garima Joshi

[garima.0227j@gmail.com](mailto:garima.0227j@gmail.com)

Unaffiliated <https://orcid.org/0009-0006-3916-9884>

---

## Research Article

**Keywords:** EEG, Alzheimer's Disease, sample entropy, power spectral density, phase-locking value, machine learning classification

**Posted Date:** October 20th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-11390599/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** The authors declare no competing interests.

---

# Predictive EEG Signatures of Alzheimer's Disease: A Cross-Validated, Cross-Method Analysis Using ANCOVA Filtering and Sparse Logistic Regression

Garima Joshi

Independent Researcher

Correspondence: [garima.0227j@gmail.com](mailto:garima.0227j@gmail.com)

ORCID: 0009-0006-3916-9884

## Abstract

Alzheimer's Disease (AD) changes how the brain's electrical activity behaves, and these changes can, in principle, be picked up directly from a resting-state EEG recording without any invasive testing. This study asks a much narrower question: does a small subset of four EEG-derived measurement families, namely signal complexity, spectral power, functional connectivity and network topology, reliably separate individuals with AD from Healthy Controls (HC), and does that separation hold up under strict, leakage-free testing? Using a public dataset of 36 AD and 29 HC subjects, candidate features were selected through permutation-based ANCOVA with false discovery rate correction, controlling for age and gender, and each selected feature was then re-checked using a complementary method (Mann-Whitney U with FDR correction, plus network-based statistics for connectivity), applied to the same subjects, to confirm the finding wasn't an artifact of the first method's particular assumptions. Agreement between the two methods was strong for signal complexity (Jaccard = 0.85) and spectral power (Jaccard = 0.87), essentially perfect for two of the four network measures (Jaccard = 1.00), and weaker for connectivity (Jaccard = 0.60, assessed at the level of frequency bands). A sparse logistic regression classifier was then trained using 5-fold cross-validation in which feature selection was independently repeated inside each fold, a design specifically intended to stop information from test subjects leaking into which features get chosen. This classifier reached  $78.5\% \pm 5.8\%$  accuracy at discriminating AD from HC (sensitivity  $0.803 \pm 0.116$ , specificity  $0.753 \pm 0.100$ , AUC  $0.808 \pm 0.134$ ). A single spectral power feature was among the two features retained in three of the five folds, more than any other individual feature. Taken together, the results point to a real but modest EEG signature of AD in this dataset, identified using a validation design built specifically so that reported accuracy reflects genuine generalization rather than an artifact of how the features were selected.

**Keywords:** EEG; Alzheimer's Disease; sample entropy; power spectral density; phase-locking value; machine learning classification

## 1. Introduction

Long before someone with Alzheimer's Disease (AD) shows obvious memory loss or confusion, their brain circuits are already functioning differently: the visible symptoms arrive years after the underlying disruption begins (Yuan & Zhao, 2025). The tools currently used to catch that early disruption, brain imaging and spinal fluid testing in particular, tend to be expensive, physically invasive, or unavailable outside specialized clinics; cognitive exams avoid those specific drawbacks but still require in-person specialist administration and can miss disruption that precedes measurable cognitive change.

Resting-state EEG sidesteps all of these problems: a person sits still with their eyes closed for a few minutes, and the recording captures ongoing brain electrical activity with no needles, no radiation, and no specialized facilities required.

What exactly changes in that recording is reasonably well documented, even if it is not fully agreed upon. Several independent lines of evidence point to AD brains producing less complex, more predictable signal patterns (measurable via entropy-based metrics), shifted power distributions across frequency bands (typically weaker alpha activity alongside stronger slow-wave activity), and altered coordination between brain regions captured by phase-locking value and similar synchrony measures (Yuan & Zhao, 2025; Zhang et al., 2021). The major problem here is that these findings do not always replicate well from one study to another, most likely due to various reasons: these studies differ in how AD is diagnosed, they work with small and varied subject pools, and each of them applies their own particular flavor of these computational measures (Zhang et al., 2021; Henao Isaza et al., 2026).

This study starts from two working assumptions and tests both of them directly. First, that AD status leaves a measurable fingerprint after accounting for age- and gender-based variabilities across all four EEG-derived measurement families: signal complexity, spectral power, connectivity and the network structure built from that connectivity. Second, that whichever features turn out to carry the fingerprint can actually be used to correctly differentiate a new person as either AD or HC, provided that the person's data was never involved in choosing or training the model in the first place. A further, related expectation was that if two genuinely different statistical tests, built on different mathematical assumptions, arrive at the same set of significant features when applied to the same data, that agreement itself counts as stronger evidence than either test could offer alone.

## 2. Methods

### 2.1 Data

EEG recordings were obtained from a publicly available OpenNeuro dataset “A Dataset of Scalp EEG Recordings of Alzheimer’s Disease, Frontotemporal Dementia and Healthy Subjects from Routine EEG” (DOI: 10.3390/data8060095). Of 88 participants in the full dataset, 65 met inclusion criteria for the present analysis: 36 with AD and 29 Healthy Controls (HC); participants with frontotemporal dementia were excluded. Recordings used 19 scalp electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2) arranged per the international 10-20 system, sampled at 500 Hz. Age and gender were recorded for each participant and used as covariates in the significance testing (Section 2.5). The analysis used the preprocessed (“derivatives”) version of the dataset, whose preprocessing is described in the dataset publication (Miltiadous et al., 2023); no further artifact rejection was applied here.

### 2.2 Pre-Processing

Legacy electrode labels (T3, T4, T5, T6) were renamed to their modern equivalents (T7, T8, P7, P8), and a standard 10-20 montage was applied. Signals were re-referenced to the common average. Each recording was cropped to the 30-300 second window to exclude the initial settling artifacts and later session fatigue effects and then segmented into non-overlapping 4-second epochs.

### 2.3 Channel and Band Index Reference

For compactness, individual features in later tables and in-text results are labeled by numeric channel and band index rather than by name, as listed in the tables below.

**Table: Channel Index**

| Index   | 0   | 1   | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---------|-----|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Channel | Fp1 | Fp2 | F3 | F4 | C3 | C4 | P3 | P4 | O1 | O2 | F7 | F8 | T7 | T8 | P7 | P8 | Fz | Cz | Pz |

**Table: Band Index**

| Index | 0     | 1     | 2     | 3    | 4         |
|-------|-------|-------|-------|------|-----------|
| Band  | Delta | Theta | Alpha | Beta | Low-Gamma |

PLV connections are indexed as edges according to the upper-triangular ordering of the 19×19 channel matrix, proceeding row-wise (edge 0 = channel 0–channel 1, edge 1 = channel 0–channel 2, and so on). The specific connections named in Results and Tables 4–5 correspond to: edge 2 = Fp1–F4, edge 35 = F3–F4, edge 82 = C4–O1.

## 2.4 Feature Extraction

For every subject, every epoch, and each of the five canonical bands (delta, theta, alpha, beta, low gamma), four feature families were computed. SampEn and PLV were computed from band-pass filtered signals (delta 0.5–4, theta 4–7.5, alpha 8–12, beta 12–30, low-gamma 30–45 Hz).

- Sample Entropy (SampEn), a measure of signal complexity per channel per band (order 2, tolerance =  $0.2 \times$  the signal's standard deviation). Within each subject, per-epoch SampEn values were winsorized (5% in each tail, separately for each channel and band) before averaging across epochs.
- Power Spectral Density (PSD): multitaper spectra were averaged across epochs and converted to decibels, and the dB values were summed across the frequency bins in each band. Because the bin count per band is the same for every subject, this equals the band's mean dB up to a constant, so effect sizes are unaffected. The 181 spectral bins (true range 0 to 45 Hz, 0.25 Hz resolution) were assigned frequencies spaced evenly from 0.5 to 45 Hz. As a result, the PSD bands correspond to true ranges of approximately 0.75–3.5 (delta), 3.75–7.5 (theta), 7.75–11.5 (alpha), 11.75–29.75 (beta) and 30–44.75 Hz (gamma), up to about 0.5 Hz below the nominal edges.
- Phase-Locking Value (PLV), a measure of functional connectivity between every pair of the 19 electrodes (171 unique connections) computed via the Hilbert transform, averaged across epochs, then Fisher z-transformed.
- Graph-theoretic network measures (global efficiency, clustering coefficient, characteristic path length, small-worldness), derived from each subject's averaged, z-transformed PLV connectivity matrix, treating each electrode as a node and PLV strength as edge weight. Connectivity matrices were not thresholded prior to the network analysis; all pairwise connections were retained as weighted edges. Small-worldness was computed against an ensemble of 10 random reference graphs per subject and band. Because the matrices were not thresholded, every graph was complete; the reference graphs were therefore generated by drawing edge weights with replacement from all entries of the subject's matrix (including its zero diagonal) and rescaling them to preserve the mean weight. They are matched to the observed graph on node and edge count, not degree distribution.

Each feature was summarized to one value per subject by averaging across epochs (SampEn after winsorizing; PSD before the dB conversion; PLV before the Fisher z-transform). Network measures were computed once per subject from the averaged matrix. The resulting per-subject values by group and band are shown in Figures 1 to 3. Several families, most clearly PLV, had long upper tails with many high-value outliers, and the group distributions overlapped substantially.

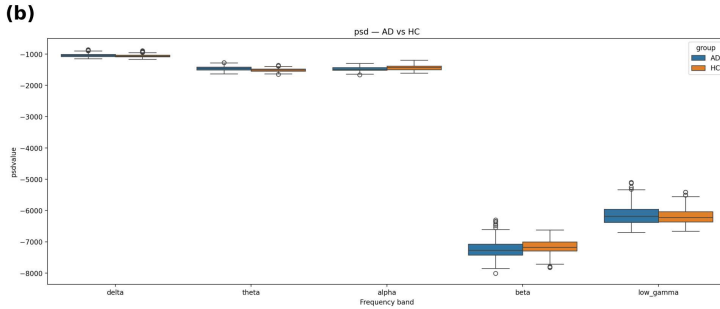
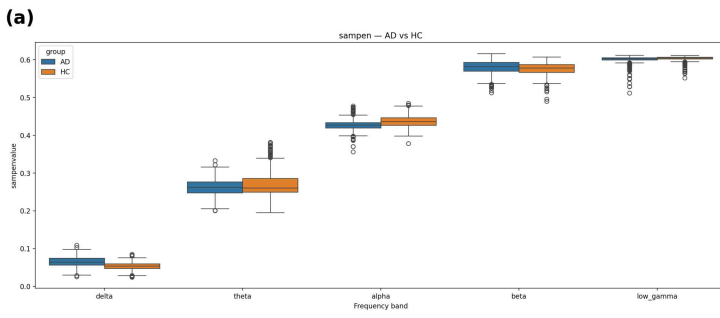


Figure 1. Per-subject sample entropy (a) and power spectral density (b) in AD and HC by frequency band. Values are pooled across all channels. PSD values are decibel values summed across the frequency bins of each band (Section 2.4), so they are large and negative. Values are not adjusted for age or gender.

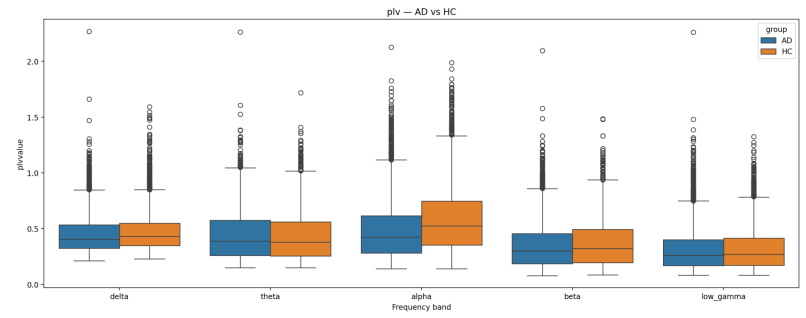


Figure 2. Per-subject phase-locking value (Fisher z-transformed) in AD and HC by frequency band, pooled across all 171 connections. Long upper tails with many high-value outliers are present in every band. Values are not adjusted for age or gender.

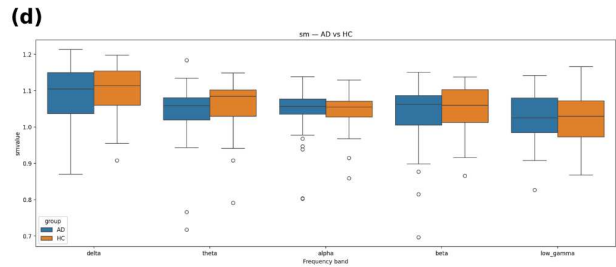
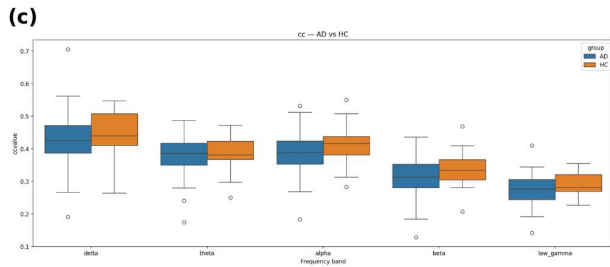
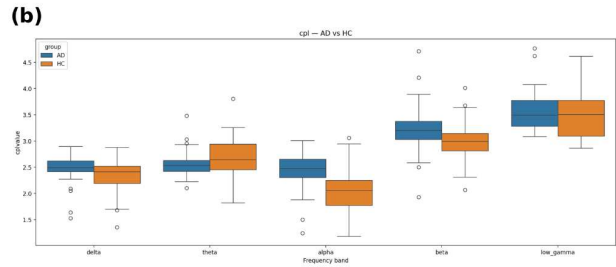
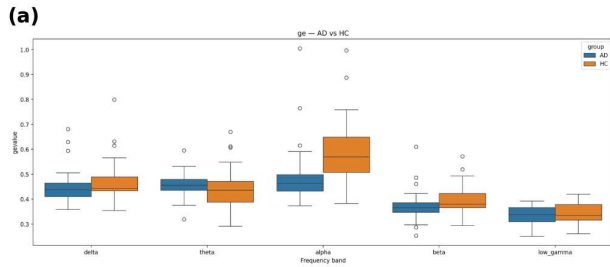


Figure 3. Per-subject network measures in AD and HC by frequency band: (a) global efficiency, (b) characteristic path length, (c) clustering coefficient, (d) small-worldness. Each subject contributes one value per band. Values are not adjusted for age or gender.

## 2.5 Statistical Analysis

Each feature was residualized on age and gender via ordinary least squares regression, with the regression refit separately for each analysis reported below: on the full 65-subject sample for the significance-testing results in Sections 3.1–3.2, and independently within each training fold, using only that fold's subjects, for the cross-validated predictive-modeling results in Sections 2.6 and 3.3. Group differences were then assessed on the residuals using permutation-based ANCOVA (500 permutations, group labels shuffled to build the null distribution; the test statistic was the difference in mean residuals between AD and HC, and two-sided p-values were the proportion of permuted statistics at least as extreme as the observed one). The distributions of the feature values were inspected before testing (Figure 4); several were skewed, such as global efficiency in the delta and alpha bands and characteristic path length in the delta band, so no normality assumption was made, and the Mann-Whitney U check described below also does not rely on one. Effect sizes were quantified using Cohen's d, computed on

the unadjusted feature values as  $(\text{mean AD} - \text{mean HC}) / \text{pooled SD}$ , so positive values indicate higher values in AD; p-values come from the age- and gender-adjusted analysis. Every individual feature was tested separately: each channel/band combination for SampEn and PSD, each network measure per band, and each of the individual 171 PLV connections per band (855 tests for PLV alone). FDR correction (Benjamini & Hochberg, 1995) was applied separately within each of these sets of tests: SampEn (95 tests), PSD (95 tests), each of the four network measures (5 tests each, one per band), and PLV (855 tests). A feature was considered significant only at FDR-corrected  $p < 0.05$ .

As a robustness check, this analysis was independently repeated using the Mann-Whitney U test (a non-parametric, rank-based test not reliant on the same distributional assumptions as ANCOVA) on the same age/gender residual values, with FDR correction applied within the same sets of tests. For PLV, network-based statistics (NBS; Zalesky et al., 2010) was used in place of per-connection testing, applied separately to each of the five bands on connection-wise age/gender residuals of the PLV matrices: edge-level differences between groups were first assessed with a two-tailed t-test at each of the 171 connections, edges exceeding a component-forming threshold of  $t = 2.5$  were grouped into connected components, and the size of each observed component was compared against a null distribution built from 1,000 permutations of group labels, with a component considered significant at  $p < 0.05$  (family-wise corrected across connections within a band via the permutation procedure itself, per Zalesky et al., 2010; no further correction was applied across the five bands). Agreement between the two methods was quantified per feature family using the Jaccard index, based on a complementary analytical method applied to the same subjects rather than an independent sample. For SampEn and PSD, agreement was assessed on the sets of significant channel-band features; for each network measure, on the set of significant bands; and for PLV at the band level, where a band counted as significant under ANCOVA if at least one of its 171 connections was significant and under NBS if at least one component was significant.

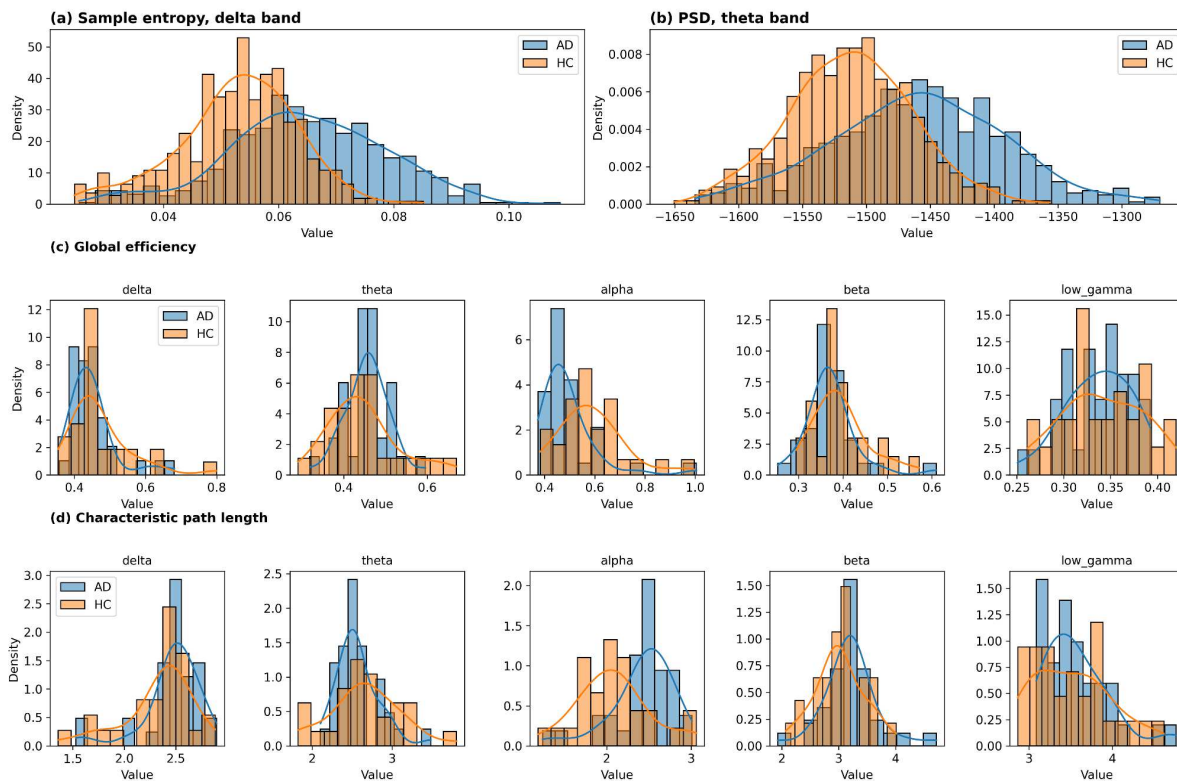


Figure 4. Histograms of per-subject feature values in AD (blue) and HC (orange): (a) sample entropy, delta band, and (b) power spectral density, theta band, both pooled across channels; (c) global efficiency and (d) characteristic path length, shown separately for each frequency band. Values are not adjusted for age or gender.

## 2.6 Predictive Modeling

To assess whether significant features could discriminate AD from HC in unseen subjects, 5-fold stratified cross-validation was used, re-running the entire feature selection procedure independently within each fold using only that fold's training subjects. This design specifically prevents information from held-out test subjects from influencing which features are treated as candidate predictors, since doing otherwise can introduce data leakage that can substantially inflate reported classifier accuracy in secondary EEG-biomarker analyses.

Within each fold, candidate features were capped using an events-per-variable (EPV) rule standard in logistic regression methodology:  $\max \text{features} = \lfloor \min(n_{\text{AD,train}}, n_{\text{HC,train}}) / 10 \rfloor$ , retaining the features with the largest absolute effect size when the significant-feature count exceeds this threshold. The threshold was fixed prior to model fitting, independent of any resulting accuracy value. Retained features were standardized (fit on training data only), and used to train an L1-regularized (LASSO; Tibshirani, 1996) logistic regression classifier; regularization strength was selected per fold via inner 3-fold cross-validation on the training data alone. Performance was then evaluated on each fold's held-out subjects and is reported as mean  $\pm$  standard deviation across five outer folds (standard deviations computed with divisor  $n$ ). Sensitivity is the proportion of AD participants correctly identified and specificity the proportion of HC participants correctly identified. AUC is unaffected by which class is treated as positive. Features selected in multiple folds were recorded as indicators of the result stability. Age and gender adjustment was applied only within the feature-selection tests in Section 2.5; the classifier itself was trained on unadjusted, standardized feature values.

All analyses were implemented in Python (NumPy, SciPy, statsmodels, scikit-learn, MNE-Python, NetworkX, antropy, pandas, h5py, and the Brain Connectivity Toolbox for Python).

## 3. Results

### 3.1 Feature Significance

Under permutation ANCOVA with FDR correction, 32 SampEn features, 28 PSD features, one global efficiency measure, one characteristic path length measure, and 78 individual PLV connections (spanning 5 of 5 frequency bands) reached significance (FDR-corrected  $p < 0.05$ ). No clustering coefficient or small-worldness features reached significance. Effect sizes (Cohen's  $d$ ) among significant features are summarized in Table 1, ranging from moderate to large across every family with at least one significant result. The direction of the effects depended on the band. Among the 32 significant SampEn features, 16 were higher in AD (all in the delta band, including the five largest effects, at Fz, F3, O1, P8 and P7) and 16 were lower in AD (12 in the alpha band, 4 in the low-gamma band). Among the 28 significant PSD features, 17 were higher in AD (15 theta, 2 delta) and 11 were lower (6 alpha, 5 beta). PLV connections made up the largest number of significant features (78 of 140 in total), whereas SampEn had the largest share of its tests reaching significance (32 of 95, versus 28 of 95 for PSD and 78 of 855 for PLV); as shown in Section 3.4, neither the number nor the share of significant features by itself predicts which features the predictive model would later rely on most heavily. Global efficiency and characteristic path length, despite contributing only one significant feature each, showed perfect agreement between the two statistical methods on that single feature (Section 3.2), reflecting consistency between the two tests rather than the strength of the underlying finding.

**Table 1. Effect Sizes (Cohen's  $d$ ) Among Significant Features, by Feature Family**

| Feature Family | $n$ | Mean $ d $ | Range | Strongest Feature ( $ d $ ) |
|----------------|-----|------------|-------|-----------------------------|
|----------------|-----|------------|-------|-----------------------------|

|                            |    |       |                 |                                   |
|----------------------------|----|-------|-----------------|-----------------------------------|
| SampEn                     | 32 | 0.928 | -1.105 to 1.301 | Channel 16, delta band: d = 1.301 |
| PSD                        | 28 | 0.923 | -1.121 to 1.354 | Channel 10, theta band: d = 1.354 |
| PLV                        | 78 | 0.918 | -1.325 to 0.994 | Alpha band, edge 2: d = -1.325    |
| Global Efficiency          | 1  | 0.775 | -0.775          | Alpha band: d = -0.775            |
| Characteristic Path Length | 1  | 0.882 | 0.882           | Alpha band: d = 0.882             |
| Clustering Coefficient     | 0  | n/a   | n/a             | n/a                               |
| Small-Worldness            | 0  | n/a   | n/a             | n/a                               |

### 3.2 Cross-Method Agreement

Agreement between the permutation ANCOVA and the Mann-Whitney U/NBS approach was strong for SampEn (29 out of 34 flagged features agreed; Jaccard = 0.853), PSD (27 out of 31 flagged features agreed; Jaccard = 0.871), global efficiency (Jaccard = 1.000), and characteristic path length (Jaccard = 1.000). Both methods independently found no significant clustering coefficient or small-worldness features. Agreement was weaker for PLV connectivity at the band level (3 out of 5 bands flagged by both; Jaccard = 0.600); notably, every band flagged by the Mann-Whitney/NBS approach was also flagged by permutation ANCOVA, indicating the disagreement reflects differing sensitivity rather than conflicting conclusions (Table 2).

**Table 2. Cross-Method Agreement Between Permutation ANCOVA and Mann-Whitney U/NBS ( $\alpha = 0.05$ )**

| Feature family             | ANCOVA Significant | Mann-Whitney U Significant | Agree | ANCOVA Only | Mann-Whitney U Only | Jaccard | Agreed Feature(s) | Agreed Band(s)         |
|----------------------------|--------------------|----------------------------|-------|-------------|---------------------|---------|-------------------|------------------------|
|                            |                    |                            |       |             |                     |         |                   |                        |
| SampEn                     | 32                 | 31                         | 29    | 3           | 2                   | 0.853   | -                 | -                      |
| PSD                        | 28                 | 30                         | 27    | 1           | 3                   | 0.871   | -                 | -                      |
| Global Efficiency          | 1                  | 1                          | 1     | 0           | 0                   | 1.000   | [2]               | Alpha                  |
| Clustering Coefficient     | 0                  | 0                          | 0     | 0           | 0                   | n/a     | -                 | -                      |
| Characteristic Path length | 1                  | 1                          | 1     | 0           | 0                   | 1.000   | [2]               | Alpha                  |
| Small-Worldness            | 0                  | 0                          | 0     | 0           | 0                   | n/a     | -                 | -                      |
| PLV (band level)           | 5                  | 3                          | 3     | 2           | 0                   | 0.600   | [2, 3, 4]         | Alpha, Beta, Low Gamma |

Note: For PLV, the Mann-Whitney U column refers to network-based statistics (NBS), and agreement is assessed at the band level (see Section 2.5). Band indices: 0 delta, 1 theta, 2 alpha, 3 beta, 4 low gamma.

### 3.3 Classification performance

Under 5-fold cross-validation with per-fold feature re-selection (EPV = 10, capping each fold to 2 candidate features), the LASSO classifier achieved a mean accuracy of  $78.5\% \pm 5.8\%$  across folds (Table 3), with a mean sensitivity of  $0.803 \pm 0.116$ , specificity of  $0.753 \pm 0.100$ , and AUC of  $0.808 \pm 0.134$ . This accuracy held despite substantial variation in how many candidate features were available at the significance-testing stage before capping, ranging from 35 significant features in the smallest fold to 175 in the largest, a nearly fivefold difference. Restricting each fold to two features regardless of that variation, following the events-per-variable (EPV) rule described in Section 2.6, kept the features-per-subject ratio fixed at 0.04 across every fold and the selected regularization strength constant at  $C = 1.0$ , indicating that the cap, not the regularization search, was the primary mechanism controlling model complexity in this dataset.

Per-fold accuracy ranged from 69.2% (Fold 5) to 84.6% (Folds 2 and 4), with no consistent relationship between the number of pre-cap significant features and resulting accuracy; Fold 5, which drew from the largest pool of pre-cap significant features (175), produced the lowest accuracy and also the lowest sensitivity (0.571) and AUC (0.571) among all folds, while Fold 4, drawing from the second-largest pool (96), produced one of the two highest accuracies and the highest AUC (0.976). This suggests the EPV-based cap was largely successful at neutralizing the risk posed by a larger, noisier candidate pool, rather than pool size directly predicting downstream performance. The standard deviation of 5.8 percentage points across folds reflects how much accuracy varied with the subjects held out; with 13 test subjects per fold, each fold's estimate is itself imprecise.

**Table 3. Per-Fold Cross Validation Summary of Feature Selection and Classification Performance**

| Fold Number | Training AD count | Training HC Count | EPV Value | Total training subjects | Pre-Cap significant feature found | Cap/Post-Cap feature limit | Feature per subject ratio | Selected C | Significant features used (post-cap) | Specificity       | Sensitivity       | AUC               | Accuracy          |
|-------------|-------------------|-------------------|-----------|-------------------------|-----------------------------------|----------------------------|---------------------------|------------|--------------------------------------|-------------------|-------------------|-------------------|-------------------|
| 1           | 28                | 24                | 10        | 52                      | 35                                | 2                          | 0.04                      | 1.0000     | 2                                    | 0.600             | 0.875             | 0.875             | 0.769             |
| 2           | 29                | 23                | 10        | 52                      | 46                                | 2                          | 0.04                      | 1.0000     | 2                                    | 0.833             | 0.857             | 0.786             | 0.846             |
| 3           | 29                | 23                | 10        | 52                      | 47                                | 2                          | 0.04                      | 1.0000     | 2                                    | 0.667             | 0.857             | 0.833             | 0.769             |
| 4           | 29                | 23                | 10        | 52                      | 96                                | 2                          | 0.04                      | 1.0000     | 2                                    | 0.833             | 0.857             | 0.976             | 0.846             |
| 5           | 29                | 23                | 10        | 52                      | 175                               | 2                          | 0.04                      | 1.0000     | 2                                    | 0.833             | 0.571             | 0.571             | 0.692             |
| Mean        | -                 | -                 | -         | -                       | -                                 | -                          | -                         | -          | -                                    | $0.753 \pm 0.100$ | $0.803 \pm 0.116$ | $0.808 \pm 0.134$ | $0.785 \pm 0.058$ |

### 3.4 Feature Stability across Folds

Across five folds, the classifier selected two features per fold, for a total of ten feature-selection events (Table 4). These selections were drawn exclusively from three of the four feature families tested, SampEn, PSD, and PLV, despite global efficiency and characteristic path length both reaching significance under both statistical methods with perfect cross-method agreement (Section 3.2). Because each fold's candidates were first capped to the two with the largest absolute effect sizes (Section 2.6), and the global efficiency and characteristic path length features had smaller full-sample effect sizes ( $|d| = 0.775$  and  $0.882$ ) than the strongest features in the other families (up to  $|d| = 1.35$ ), this absence most likely reflects that effect-size ranking rather than a direct test of discriminative power; fold-level effect sizes were not examined, and the features were not tested in competition without the cap. Figures 5 to 7 show the full effect-size landscape for PLV, PSD and SampEn, respectively, with LASSO-selected features outlined; because the features that recur reflect effect-size ranking within each training set as well as the classifier's own weighting, they should not be read as the classifier's independent choice among all candidates.

Two features were selected in more than one fold and are reported separately in Table 5: a PSD feature (Channel 10, band 1), selected in three of five folds, and a PLV connection (Band 2, edge 2), selected in two of five folds. Of the ten total feature-selection events across the cross-validation procedure, the PSD feature at channel 10, band 1, was selected in three of five folds — the highest selection rate of any individual feature observed in this analysis. This recurrence is reported descriptively, as the most frequently selected feature among those tested, rather than as a formally validated stability metric.

SampEn tells a different story than PSD, PLV, GE, or CPL.

Two SampEn features were individually selected, at channel 2, band 0 in Fold 1 and channel 16, band 0 in Fold 2, but neither repeated in any subsequent fold, even though SampEn had the largest share of its tests reaching significance (32 of 95; Section 3.1). This pattern need not reflect weak discriminative power. The strongest effects in the three families were nearly tied (PSD channel 10 theta,  $d = +1.35$ ; PLV alpha edge 2,  $|d| = 1.33$ ; SampEn channel 16 delta,  $d = +1.30$ ), so which features ranked in the top two in a given training set could vary, and no SampEn feature was selected in more than one fold (Table 4).

Coefficient direction, by contrast, was where the two consistently repeated features told a clean and stable story. Class labels were coded AD = 0 and HC = 1, so a negative coefficient pushes a prediction toward AD and a positive coefficient toward HC. The PSD feature at channel 10, band 1, carried a negative coefficient in every fold it appeared in (Folds 2, 3, and 4:  $-1.0327$ ,  $-1.5086$ , and  $-1.1046$ ), consistently predicting AD, in line with its positive effect size in Table 1 (higher theta-band power at F7 in AD,  $d = +1.354$ ). The PLV connection at band 2, edge 2, carried a positive coefficient in both folds it

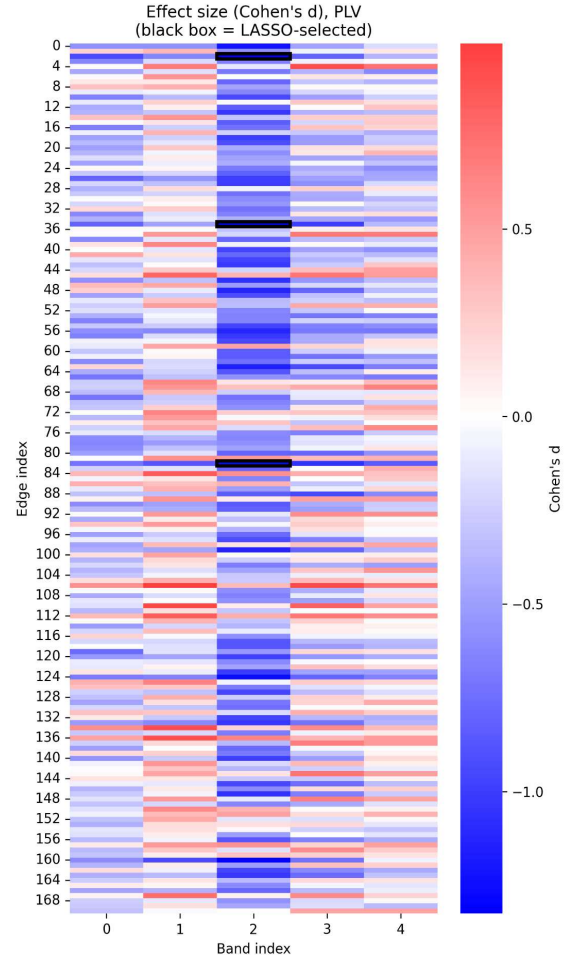


Figure 5. Effect size (Cohen's  $d$ ) for every tested phase-locking value feature (171 edges  $\times$  5 bands). Positive values indicate higher values in AD than in HC. Color reflects the raw effect size regardless of statistical significance. Black outlines mark features selected by LASSO during cross-validation. Not every colored cell reached FDR-corrected significance (Section 3.1), and not every significant feature was selected by the classifier (Section 3.4).

appeared in (Fold 1: 1.2152; Fold 4: 1.1192), consistently predicting HC, in line with its negative effect size in Table 1 (lower alpha-band synchrony in AD,  $d = -1.325$ ). Both repeated features maintained the same coefficient direction across every fold in which they were selected, though, as above, no formal test was run to establish whether this consistency exceeds what would occur by chance.

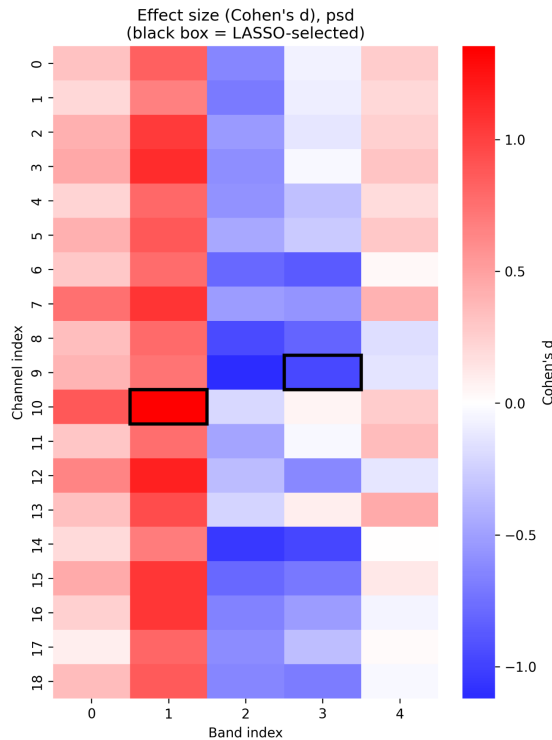


Figure 6. Effect size (Cohen's  $d$ ) for every tested power spectral density feature (19 channels  $\times$  5 bands). Positive values indicate higher values in AD than in HC. Color reflects the raw effect size regardless of statistical significance. Black outlines mark features selected by LASSO during cross-validation. Not every colored cell reached FDR-corrected significance (Section 3.1), and not every significant feature was selected by the classifier (Section 3.4).

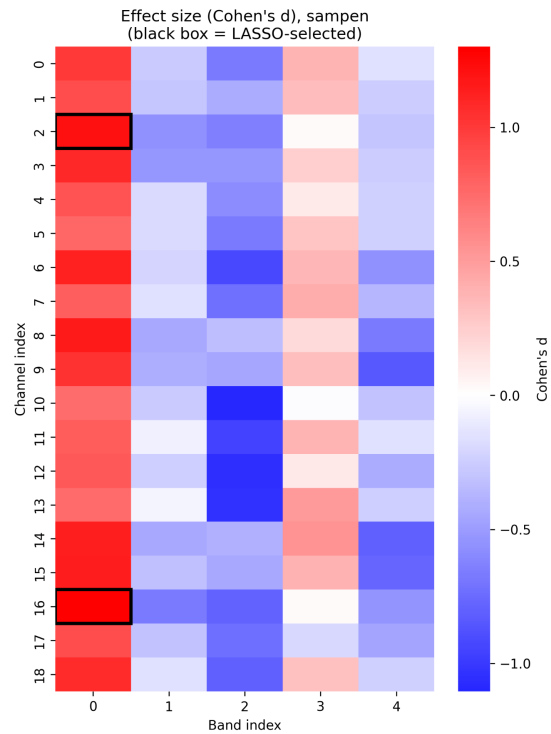


Figure 7. Effect size (Cohen's  $d$ ) for every tested sample entropy feature (19 channels  $\times$  5 bands). Positive values indicate higher values in AD than in HC. Color reflects the raw effect size regardless of statistical significance. Black outlines mark features selected by LASSO during cross-validation. Not every colored cell reached FDR-corrected significance (Section 3.1), and not every significant feature was selected by the classifier (Section 3.4).

**Table 4. Individual Features Selected by LASSO in Each Cross-Validation Fold, with Direction and Magnitude of the Effect**

| Fold | Feature 1                  | Coefficient 1 | Feature 2               | Coefficient 2 |
|------|----------------------------|---------------|-------------------------|---------------|
| 1    | SampEn, channel 2, band 0  | -1.0994       | PLV, band 2, edge 2     | 1.2152        |
| 2    | SampEn, channel 16, band 0 | -1.2106       | PSD, channel 10, band 1 | -1.0327       |
| 3    | PSD, channel 10, band 1    | -1.5086       | PSD, channel 9, band 3  | 1.3566        |
| 4    | PSD, channel 10, band 1    | -1.1046       | PLV, band 2, edge 2     | 1.1192        |
| 5    | PLV, band 2, edge 82       | 1.4611        | PLV, band 2, edge 35    | 1.4314        |

**Table 5. Features Selected by LASSO in Multiple Cross-Validation Folds**

| Metric type | Feature identity   | Selected in no.of folds |
|-------------|--------------------|-------------------------|
| PSD         | Channel 10, band 1 | 3/5                     |
| PLV         | Band 2, edge 2     | 2/5                     |

#### 4. Discussion

The study found that resting-state EEG features could distinguish AD from HC, and that the features recurring across folds reflected two different properties of the signal: spectral power and phase-locking connectivity. Spectral power describes how electrical activity is distributed across frequencies; PLV describes coordination between brain regions. These two features are not arbitrary; a prior study using a comparably smaller AD/HC sample (26 AD, 29 HC) also relied on power spectrum and PLV to identify AD-associated differences (Wang et al., 2025). The significant PSD effects also followed the familiar slowing pattern in AD: higher theta (15 of 28 features) and delta (2) power, and lower alpha (6) and beta (5) power.

Alzheimer's Disease is understood to involve both localized pathological changes and a broader breakdown in connectivity between brain regions, a framework described in the literature as a "disconnection syndrome" (Geschwind, 1965; Brier et al., 2014). PSD reflects the first, a localized signal; PLV reflects the second, a connectivity signal. Feature selection combined a significance filter, an effect-size cap, and a cross-validated sparse classifier. The two features that recurred across folds represented each of these two levels, a localized measure (PSD) and a connectivity measure (PLV), a pattern consistent with AD's documented dual character (Brier et al., 2014). Within that pairing, the PSD feature recurred more often (three of five folds, versus two for the PLV connection); its full-sample effect was slightly larger ( $d = +1.35$  versus  $|d| = 1.33$ ).

The feature family with the highest share of significant results, SampEn (32 of 95 tests), was not the family that supplied the most consistently selected predictor. PSD, with 28 of 95, supplied the one feature selected in three of five folds (theta band at F7). A large number of significant results therefore did not make SampEn the strongest discriminator in this study, where each fold kept the two features with the largest effects. The SampEn results show that complexity clearly differs between AD and HC, in line with altered EEG complexity being well documented in AD (Zhang et al., 2021), although the direction depends on the band (Section 3.1). Why it did not rank as a discriminator was not tested, so the reason cannot be identified. One possibility, offered as speculation, is the nature of the measures. SampEn summarizes the overall irregularity of a signal, so it can differ between groups in many places without any single difference standing out. PSD and PLV measure specific properties, power at a given frequency and phase coupling between a particular pair of sites, which makes them more likely to produce a strong single effect that this selection procedure favors. If so, SampEn is most probably a mismatch for this approach, not a measure that cannot rank.

Several limitations restrict these findings. This is a secondary analysis of a public dataset, not original data collected for this study; participants with frontotemporal dementia present in the source dataset were excluded, leaving only the AD and HC groups (65 subjects total) used here. This modest sample size limits how confidently these results generalize, and the model was evaluated only against healthy controls, not other neurological conditions, so this separation cannot be shown to be specific to AD. Cross-validation protects against feature-selection leakage within this dataset but does not substitute for

testing on an independent cohort. Disease staging was not modeled, so this study cannot determine whether PSD's relative strength over PLV reflects disease stage, progression rate, or another explanation. The design is cross-sectional, so the EEG differences are associations with diagnosis; this study cannot say whether they precede, accompany or follow the disease. Because the classifier was trained on unadjusted feature values, part of its performance could reflect residual age or gender differences between the groups. In addition, because each fold's candidate features were capped to the two with the largest absolute effect sizes before the classifier was fit, the features that recur across folds largely reflect which features had the largest effect sizes in each training set, rather than a selection made by the classifier among many candidates.

Taken together, these results point to a modest EEG signal associated with AD in this dataset. Spectral power and connectivity emerged as the discriminating features, in line with the localized and disconnection accounts of the disease, while sample entropy, despite its many significant results, did not rank as a discriminator, which, as speculated above, is in line with what we know about AD. Neither pattern was imposed on the analysis, since no feature was chosen in advance and both emerged on their own. This suggests that this type of data-driven approach is meaningful and worth pursuing for studies of progression and for early detection in neurodegenerative disease. More in-depth and further work must be done.

## 5. Conclusion

This study offers one demonstration that a rigorously validated, data-driven computational pipeline, one that did not prespecify any particular EEG feature as the expected discriminator, can recover features consistent with existing AD neurophysiology. The significance of that result is methodological as much as clinical: it shows this kind of approach is worth pursuing further, not that it currently provides a diagnostic answer.

**Ethics Statement:** This study uses a publicly available, de-identified EEG dataset (Miltiadous et al., 2023). No new data was collected for this study, and no human or animal subjects were directly involved. Ethical approval for the original data collection is documented in the source publication.

**Data Availability Statement:** The EEG data analyzed in this study are publicly available via OpenNeuro (Miltiadous et al., 2023; DOI: 10.3390/data8060095). Code used to perform all analyses is available at [https://github.com/garimaaa01/predictive\\_model](https://github.com/garimaaa01/predictive_model).

**Funding:** This research received no external funding or grant support.

**Conflict of Interest:** The author declares no competing interests.

**Author Contributions:** G.J. conceived the study, designed the analysis, wrote the analysis code, performed all analyses, and wrote the manuscript, with AI assistance as described in the statement below.

**Use of Generative AI:** The study concept, analysis design, analysis code, and interpretation of results are the author's own. Generative AI (Claude, Anthropic) was used as a reviewer of the analysis code, which flagged data-leakage risks and reporting errors that the author then corrected and as a checker of reported numbers against analysis output and of references. All design decisions, and the corrections that followed from code review, were made by the author. The author reviewed all content and takes full responsibility for the accuracy, integrity, and originality of the work. No AI tool is listed as an author.

## References

1. Miltiadous, A., Tzamourta, 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

scalp EEG recordings of Alzheimer's disease, frontotemporal dementia and healthy subjects from routine EEG. *Data*, 8(6), 95. <https://doi.org/10.3390/data8060095>

2. Yuan, Y., & Zhao, Y. (2025). The role of quantitative EEG biomarkers in Alzheimer's disease and mild cognitive impairment: applications and insights. *Frontiers in Aging Neuroscience*, 17, 1522552. <https://doi.org/10.3389/fnagi.2025.1522552>
3. Zhang, H., Geng, X., Wang, Y., Guo, Y., Gao, Y., Zhang, S., Du, W., Liu, L., Sun, M., Jiao, F., Yi, F., Li, X., & Wang, L. (2021). The significance of EEG alpha oscillation spectral power and beta oscillation phase synchronization for diagnosing probable Alzheimer disease. *Frontiers in Aging Neuroscience*, 13, 631587. <https://doi.org/10.3389/fnagi.2021.631587>
4. Wang, J., Zhao, J., Chen, X., Yin, B., Li, X., & Xie, P. (2025). Alzheimer's disease diagnosis using rhythmic power changes and phase differences: a low-density EEG study. *Frontiers in Aging Neuroscience*, 16, 1485132. <https://doi.org/10.3389/fnagi.2024.1485132>
5. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
6. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
7. Zalesky, A., Fornito, A., & Bullmore, E. T. (2010). Network-based statistics: identifying differences in brain networks. *NeuroImage*, 53(4), 1197–1207. <https://doi.org/10.1016/j.neuroimage.2010.06.041>
8. Henao Isaza, V., Aguillon, D., Tobón-Quintero, C. A., Lopera, F., & Ochoa-Gómez, J. F. (2026). Comprehensive methodology for sample enrichment in EEG biomarker studies for Alzheimer's risk classification. *PLOS ONE*, 21(3), e0343722. <https://doi.org/10.1371/journal.pone.0343722>
9. Geschwind, N. (1965). Disconnexion syndromes in animals and man. *Brain*, 88(2), 237–294. <https://doi.org/10.1093/brain/88.2.237>
10. Brier, M. R., Thomas, J. B., & Ances, B. M. (2014). Network dysfunction in Alzheimer's disease: refining the disconnection hypothesis. *Brain Connectivity*, 4(5), 299–311. <https://doi.org/10.1089/brain.2014.0236>

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableS1significantfeatures.csv](#)