

Information-Theoretic Geometry HCFIM features for data-limited EEG screening of Alzheimer's disease and Frontotemporal Dementia

Héctor Rojas-Pescio

hector.rojas.p@usach.cl

Universidad de Santiago de Chile

Research Article

Keywords: neurophysiological screening, EEG, Alzheimer's disease, frontotemporal dementia, entropy, statistical complexity, Fisher information, information-theoretic geometry, machine learning

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10834216/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Information-Theoretic Geometry HCFIM features for data-limited EEG screening of Alzheimer's disease and Frontotemporal Dementia

Héctor Rojas-Pescio^{1, 2}

¹ Departamento de Ingeniería Informática, Universidad de Santiago de Chile, Av. Víctor Jara N° 3659, Estación Central, Santiago 9190864, Chile

² Nouha Biomedical Technologies, Dr. Barros Borgoño N° 71, Providencia, Santiago 7500593, Chile

Abstract. Resting-state EEG is a scalable modality for neurophysiological screening, but conventional spectral descriptors capture only part of the signal organization altered in neurodegenerative disease. This study evaluates a reproducible information-theoretic geometry framework based on entropy (H), statistical complexity (C) and Fisher Information Measure (FIM) for data-limited EEG characterization of Alzheimer's disease (AD) and frontotemporal dementia (FTD). Using the public OpenNeuro resting-state EEG dataset comprising 29 healthy controls, 36 AD patients and 23 FTD patients, preprocessed eyes-closed 19-channel recordings were segmented into epochs, transformed into finite probability distributions, projected into HC and $HFIM$ planes and $HCFIM$ space, and aggregated at subject level. Synthetic signal regimes were first used to verify the behavior of the probability-domain geometry. Regional and lateralized cortical schemas were then compared against PSD baselines using leakage-aware subject-level validation. Then, various machine learning classification models were applied to distinguish between healthy and neurologically affected individuals. In the regional schema, the most consistent AD-related geometric displacement was found in temporal alpha rhythms, using Permutation entropy with $m = 4$, $\tau = 1$ (inter/intra ratio = 0.847; silhouette = 0.141), while in FTD it was found in temporal broadband (inter/intra ratio = 1.140; silhouette = 0.073); both silhouette values indicate modest, overlapping group geometry rather than well-separated clusters. For AD classification, PSD + HCFIM with Extra Randomized Trees achieved ROC AUC = 0.818 and balanced accuracy = 0.803; adding lateralized differences increased ROC AUC to 0.832 and sensitivity to 0.833, with lower specificity. For FTD, regional HCFIM-only features achieved the highest ROC AUC (0.810), while regional PSD + HCFIM achieved the best balanced accuracy (0.740). Lateralization highlighted left temporal broadband displacement in FTD but did not improve classification. These results support HCFIM as an interpretable, reproducible complement to spectral EEG features for preclinical methodological research, while emphasizing the need for independent validation before clinical or device-specific claims.

Keywords. neurophysiological screening; EEG; Alzheimer's disease; frontotemporal dementia; entropy; statistical complexity; Fisher information; information-theoretic geometry; machine learning.

1. Introduction

Neurodegenerative dementias are characterized by progressive cognitive, behavioral, and functional impairments (Scheltens et al., 2021; Bang et al., 2015). Alzheimer's disease (AD) is the most common dementia subtype, and early diagnosis remains a significant clinical and public health problem since pathophysiological alterations can occur years before overt cognitive symptoms appear (Sperling et al., 2011; Scheltens et al., 2021).

Resting-state electroencephalography (EEG) has emerged as an appealing non-invasive choice for investigating AD-related neurophysiological alterations due to its cheap cost, scalability, and direct sensitivity to cortical oscillatory dynamics (Cassani et al., 2018). These characteristics are particularly crucial for preclinical screening and longitudinal follow-up, when repeated measurements and operational feasibility are critical (Babiloni et al., 2021; Cassani et al., 2018).

Spectral changes are among the most reproducible EEG findings in dementia (Jeong, 2004; Dauwels et al., 2010; Cassani et al., 2018). Resting-state EEG in Alzheimer's disease typically reveals slowing, which is defined as a transfer of power from faster alpha and beta rhythms to slower delta and theta activity (Cassani et al., 2018). Expert recommendations for AD clinical trials include anomalies in posterior alpha rhythms and broad delta/theta activity as potential EEG indicators for disease classification and progression monitoring (Babiloni et al., 2021). However, spectral power only describes a single feature of the signal. EEG is a multichannel, noisy, nonlinear, and non-stationary physiological time series, and neurodegeneration can alter temporal organization, irregularity, probability structure, and local dynamical sensitivity (Jeong, 2004; Dauwels et al., 2010; Sun et al., 2020).

For this reason, information-theoretic and nonlinear descriptors have been widely explored as complementary representations of brain activity (Dauwels et al., 2010; Sun et al., 2020). Permutation entropy (PE), introduced by Bandt and Pompe, estimates the uncertainty of ordinal patterns extracted from a time series and is computationally efficient, robust to observational noise and invariant under monotonic transformations (Bandt and Pompe, 2002). When combined with disequilibrium measures such as Jensen-Shannon divergence, PE supports the construction of entropy-complexity planes that help distinguish between regular, chaotic and stochastic dynamics (Martin et al., 2006; Rosso et al., 2007). In cognitive impairment and AD, PE and statistical complexity have shown that alterations in complexity are not necessarily global or monotonic; instead, they can depend on frequency band, cortical location and disease stage (Echegoyen et al., 2020). This observation is important because it discourages the simplistic assumption that pathological EEG must always exhibit uniformly lower complexity.

Multiscale and multivariate entropy approaches further support the idea that neurodegenerative EEG alterations should be studied across frequency bands and brain regions (Morabito et al., 2012; Deng et al., 2017). Multivariate multiscale permutation entropy was proposed to exploit the multidimensional nature of EEG and to reveal inter-channel and scale-dependent structure (Morabito et al., 2012). Subsequent work using multivariate multiscale weighted permutation entropy reported AD-related complexity changes distributed across temporal, occipitoparietal and fronto-occipitoparietal regions, depending on frequency band and time scale (Deng et al., 2017). Together, these studies motivate regional and band-specific analysis rather than a purely global summary of EEG complexity.

The present work extends to a doctoral research line in which physiological signals are represented in spaces defined by information-theory measures (Rojas-Pescio et al., 2024; Rojas-Pescio et al., 2025). In this framework, a signal is treated as an observable realization of an underlying dynamical system. A probability distribution is induced from the signal, and that distribution is projected into geometric coordinates such as entropy (H), statistical complexity (C) and Fisher Information Measure (FIM). The Fisher-Shannon information plane has previously been proposed for the analysis of complex and non-stationary signals, showing that Fisher information and entropy can provide complementary descriptions of signal behavior (Vignat and Bercher, 2003). The HCFIM representation developed here integrates these ideas into a three-dimensional information-geometric space where entropy, statistical complexity and Fisher information are considered jointly.

This study applies the HCFIM framework to a public resting-state EEG dataset including healthy controls, AD patients and frontotemporal dementia (FTD) patients. The dataset contains eyes-closed scalp EEG recordings from 88 subjects: 36 AD patients, 23 FTD patients and 29 healthy controls, acquired with 19 electrodes according to the international 10-20 system and distributed in BIDS format with both raw and preprocessed data (Miltiadous et al., 2023; NEMAR/OpenNeuro ds004504). The present analysis uses the preprocessed OpenNeuro derivatives directly, without retrospective adaptation to operational acquisition devices. This design keeps the scientific question focused on the methodological value of HCFIM rather than on device transferability.

Recent analyses of the same public cohort have evaluated spectral, nonlinear, connectivity and graph-theoretical features for AD and FTD characterization. For example, multidimensional EEG analysis has reported frequency-dependent and spatially heterogeneous abnormalities in AD and FTD, with AD showing more posterior abnormalities and FTD showing comparatively different spatial patterns (Cai et al., 2026). Other EEG-based AD/FTD classification studies using this cohort have explored functional connectivity and machine-learning approaches (Mlinaric et al., 2026).

These works show that the cohort is useful for methodological exploration, but they also reinforce the need for careful validation, leakage control and restrained interpretation in a data-limited setting.

The central question of this paper is whether resting-state EEG signals from controls, AD and FTD occupy distinguishable regions in HCFIM space, and whether these differences become more informative when organized by frequency band, cortical region and hemispheric lateralization. The contribution is therefore methodological and preclinical. We do not claim that HCFIM is a clinically validated diagnostic biomarker. Instead, we propose a reproducible pipeline for constructing EEG-derived probability distributions, projecting them into HC and HFIM planes and a HCFIM space, quantifying group separability, and evaluating classification utility under subject-level validation.

The study addresses four research questions: 1) Do synthetic dynamical regimes occupy coherent and interpretable regions in HC, HFIM and HCFIM spaces under the selected probability transformations? 2) Do resting-state EEG signals from controls, AD and FTD show separable information-geometric patterns across frequency bands and cortical regions? 3) Does lateralized regional aggregation improve the interpretability and stability of HCFIM descriptors? And 4) Do HCFIM features provide information complementary to classical PSD-based features under leakage-aware, subject-level classification?

By following this sequence, the paper positions classification as a downstream test of utility, not as the primary scientific object. The primary object is the informational-geometric representation of EEG dynamics and its ability to support interpretable, reproducible and data-conscious exploration of neurodegenerative signal alterations.

2. Materials and Methods

2.1 Subjects and measurement

This study used the public resting-state EEG dataset described by Miltiadous et al. (2023), distributed through OpenNeuro/NEMAR as dataset ds004504/on004504. The dataset includes eyes-closed resting-state scalp EEG recordings from 88 participants: 36 patients diagnosed with Alzheimer’s disease (AD), 23 patients diagnosed with frontotemporal dementia (FTD), and 29 cognitively normal healthy controls (CN/HC). Cognitive status was assessed using the Mini-Mental State Examination (MMSE), where lower scores indicate greater cognitive impairment. According to the dataset descriptor, the mean MMSE score was 17.75 (SD = 4.5) for the AD group, 22.17 (SD = 8.22) for the FTD group, and 30 for the control group. Mean age was 66.4 years (SD = 7.9) for AD, 63.6 years (SD = 8.2) for FTD, and 67.9 years (SD = 5.4) for controls (Miltiadous et al., 2023).

All recordings were acquired from routine clinical EEG examinations. The AD and FTD diagnoses were reported by the dataset authors as being established according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R/DSM-IV), ICD-10, and the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) criteria. Disease duration was reported in months, with a median of 25 months and an interquartile range of 24-28.5 months. The dataset descriptor also states that no dementia-related comorbidities were reported for the AD group.

EEG recordings were acquired at the 2nd Department of Neurology of AHEPA General Hospital of Thessaloniki by an experienced neurology team. A clinical Nihon Kohden EEG 2100 system was used with 19 scalp electrodes placed according to the international 10-20 system: Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2. Two mastoid electrodes, A1 and A2, were used for impedance checking and as reference electrodes. Each recording was performed according to a clinical protocol while participants were seated, in a resting state, with eyes closed. The recording montage was referential using Cz for common mode rejection. Signals were sampled at 500 Hz with a reported resolution of 10 μ V/mm.

Recording duration differed across subjects and groups. The average recording duration was approximately 13.5 min for the AD group (range: 5.1-21.3 min), 12.0 min for the FTD group (range: 7.9-16.9 min), and 13.8 min for the

control group (range: 12.5-16.5 min). In total, the dataset contains approximately 485.5 min of AD recordings, 276.5 min of FTD recordings, and 402 min of control recordings.

The dataset is organized according to the Brain Imaging Data Structure (BIDS) standard and includes both raw and preprocessed EEG data. The present study uses the preprocessed derivatives released with the dataset, without retrospective adaptation to operational acquisition devices. This decision was made to keep the methodological analysis focused on the HCFIM representation and to avoid confounding the scientific evaluation with acquisition-transfer assumptions.

The original data acquisition was approved by the Scientific and Ethics Committee of AHEPA University Hospital, Aristotle University of Thessaloniki, under protocol number 142/12-04-2023. The investigations were conducted in accordance with the Declaration of Helsinki, and informed consent was obtained from all subjects involved in the study (Miltiadous et al., 2023).

2.1.1 Dataset Summary

Group	n	Mean age (years)	MMSE	Mean recording duration
Alzheimer’s disease	36	66.4 (SD = 7.9)	17.75 (SD = 4.5)	13.5 min
Frontotemporal dementia	23	63.6 (SD = 8.2)	22.17 (SD = 8.2)	12.0 min
Healthy controls	29	67.9 (SD = 5.4)	30.00 (SD = 0.0)	13.8 min

Note. The MMSE standard deviation for healthy controls was 0.0 because all control participants in the released metadata had an MMSE score of 30.

2.1.2 Acquisition Summary

Attribute	Description
Modality	Resting-state scalp EEG
Condition	Eyes closed, seated resting state
Recording setting	Routine clinical EEG
Acquisition site	2nd Department of Neurology, AHEPA General Hospital of Thessaloniki
EEG system	Nihon Kohden EEG 2100
Electrode montage	19 scalp electrodes, international 10-20 system
Scalp electrodes	Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2
Reference electrodes	A1 and A2 mastoids
Sampling rate	500 Hz
Reported resolution	10 uV/mm
Data organization	BIDS, raw and preprocessed derivatives
Ethics approval	Scientific and Ethics Committee of AHEPA University Hospital, Aristotle University of Thessaloniki, protocol 142/12-04-2023
Consent	Informed consent obtained from all subjects

2.2 Study design and reproducibility protocol

The study was designed as a reproducible methodological analysis of information-theoretic geometry in resting-state EEG. The central objective was not to validate an operational acquisition device, but to evaluate whether

probability-space representations of EEG segments can reveal group-dependent geometric organization in HC, HFIM and HCFIM spaces.

The analysis followed four consecutive stages. First, synthetic signals with known dynamical regimes were used to visually and numerically characterize the geometry induced by each signal-to-probability transformation. Second, preprocessed EEG derivatives from the public dataset were segmented into epochs and transformed into finite probability distributions. Third, each distribution was projected into information-theoretic coordinates: normalized entropy, statistical complexity and Fisher Information Measure. Fourth, candidate representations were evaluated through geometric separability and subject-level classification under leakage-aware validation.

All analyses were structured at the subject level. Epochs were used to estimate within-subject distributions and feature summaries, but no epoch from a given subject was allowed to appear simultaneously in training and test partitions. This constraint was applied to avoid optimistic performance estimates caused by window-level data leakage.

2.3 EEG derivatives and epoching

The reproducible paper analysis uses the preprocessed EEG derivatives released with the OpenNeuro/NEMAR dataset. These derivatives were selected as the scientific input to keep the analysis independent of device-transfer or operational normalization assumptions.

Each EEG recording was segmented into fixed-length windows. The experimental pipeline used 10 s epochs with 50% overlap. For each subject, epochs were indexed with their subject identifier, diagnostic group, channel set, temporal position and quality-control subset. Two quality subsets were considered: *valid_epoch*, corresponding to windows passing minimum technical validity criteria, and *clean_epoch*, corresponding to a stricter subset used for sensitivity analyses.

The final subject-level matrices were obtained by aggregating epoch-level features within each subject. Therefore, the statistical and classification unit was always the participant, not the epoch.

2.4 Synthetic signal regimes

Before applying the method to EEG, synthetic signals were generated to inspect the geometry produced by the candidate probability transformations. This stage was used as a methodological sanity check: if a transformation produced non-interpretable or unstable organization for signals with known regimes, it was treated cautiously in the EEG analysis.

The synthetic inventory included stochastic multiplicative signals, regular and variable oscillations, periodic signals, chaotic logistic-map dynamics, white noise, pink noise, brownian noise, blue noise and violet noise. For each signal family, multiple realizations were generated and projected into the corresponding probability space. The resulting point clouds were visualized in the induced probability simplex in a Principal Component Analysis biplot (Figure 1a), the HC plane (Figure 1b), the HFIM plane (Figure 1c) and the three-dimensional HCFIM space (Figure 1d).

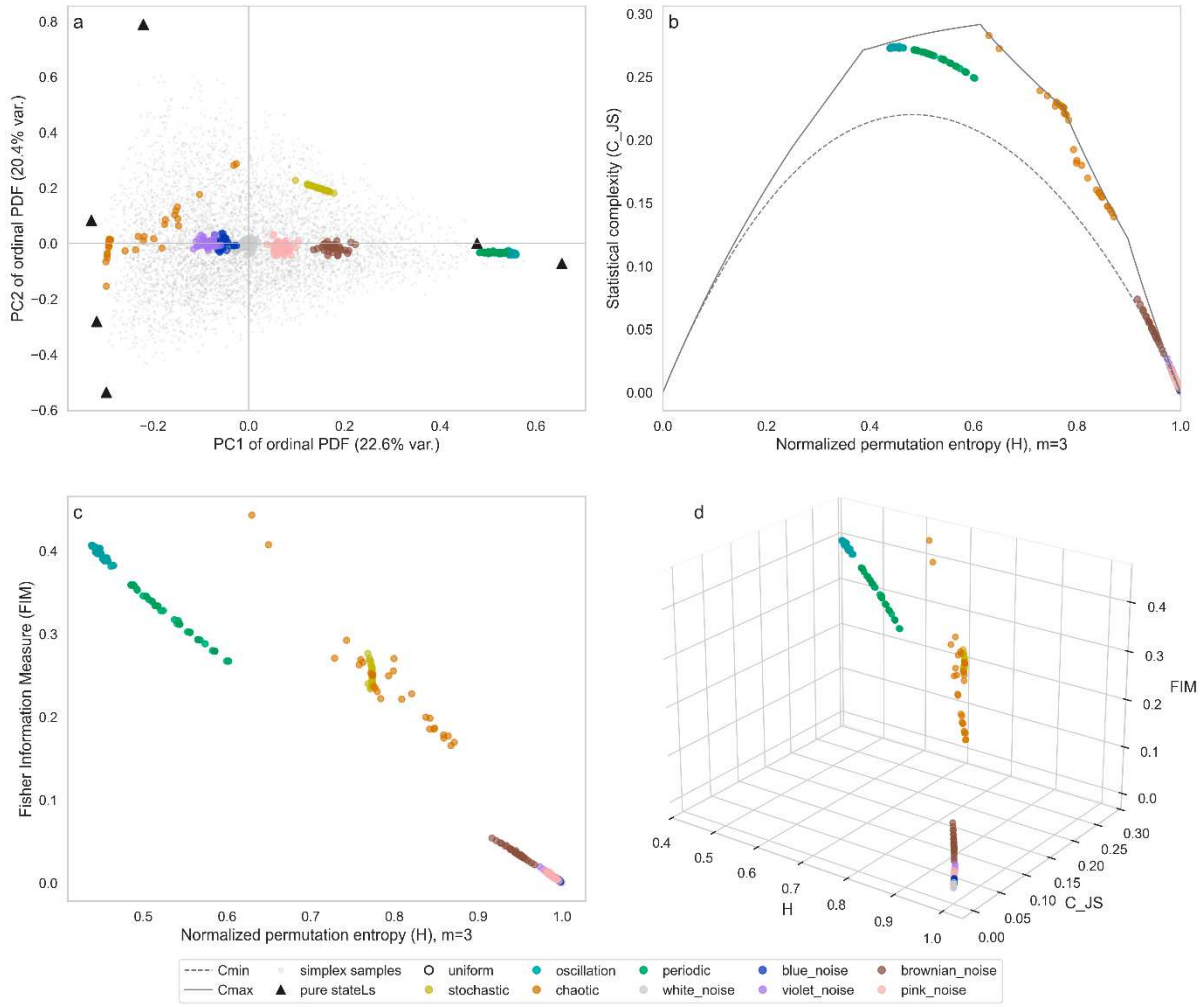


Figure 1. Synthetic information-theoretic geometry used to validate probability-domain behavior before EEG projection: simplex, HC plane, HFIM plane and HCFIM space.

2.5 Signal-to-Probability transformations

The proposed framework treats each EEG epoch as an observation of an underlying physiological dynamical process. A transformation operator maps the time series window into a finite probability distribution: $x(t) \rightarrow P = (p_1, p_2, \dots, p_M)$, $p_i \geq 0$, $\sum p_i = 1$.

The dimension M depends on the transformation. For ordinal patterns, $M = m!$; for histogram-based methods, M is the number of bins; for gridded representations, M is the number of occupied cells; and for wavelet representations, M corresponds to the number of energy components.

2.5.1 Direct Probability Domains

The main analysis considered probability transformations that naturally induce a finite distribution suitable for HC, HFIM and HCFIM analysis. These included Bandt-Pompe ordinal distributions for permutation entropy, ordinal extensions of Rényi and Tsallis entropy, dispersion-symbol distributions, distribution entropy based on distances between embedded vectors, gridded distributions in normalized phase-space representations, increment-symbol distributions, phase-angle distributions, slope-symbol distributions, normalized spectral power distributions, wavelet-energy distributions and normality-symbol distributions.

Low-dimensional parameterizations were used in the synthetic geometry stage to make simplex and HC/HFIM/HCFIM spaces interpretable. Ordinal representations were initially explored with embedding dimensions $m = 3, 4$ and 5 , and delay $\tau = 1$. Rényi and Tsallis ordinal variants were explored with parameters below and above the Shannon limit, including $\alpha = 0.8$, $\alpha = 1.2$, $q = 0.8$ and $q = 1.2$.

2.5.2 Derived Probability Domains for Regularity and Algorithmic Measures

Some measures commonly used in EEG do not directly define a probability simplex that is equivalent to the entropy-complexity plane used for ordinal or finite symbolic distributions. This applies particularly to Approximate Entropy, Sample Entropy, Fuzzy Entropy, Lempel-Ziv Complexity, compression complexity and multiscale regularity measures.

For this reason, these measures were treated as derived information-theoretic/algorithmic-complexity descriptors. When they were projected into HC, HFIM or HCFIM coordinates, an explicit derived probability domain was defined first. For Approximate Entropy and Sample Entropy, the derived domain was based on histograms of distances between embedded vectors. For Fuzzy Entropy, it was based on fuzzy similarity values between centered embedding vectors. For correlation/Kolmogorov-type descriptors, it was based on normalized correlation-integral increments across radii. For Lempel-Ziv descriptors, it was based on phrase-length distributions after binary, quantile or ordinal symbolization. Compression complexity was treated as an algorithmic descriptor rather than as a standard statistical complexity coordinate unless an explicit probability distribution was constructed. This distinction was made to avoid equating all nonlinear descriptors with classical statistical complexity.

2.6 Information-Theoretic Coordinates

For each probability distribution P , the analysis computed three coordinates: normalized entropy H , statistical complexity C and Fisher Information Measure FIM .

Normalized entropy quantified the global uncertainty of the induced distribution. For Shannon entropy:

$$H = S[P] = -\sum_{i=1}^n p_i \ln(p_i), H[P] = S[P]/S[P_e] \quad (1)$$

where P_e is the uniform distribution over the same finite state space. Analogous normalized forms were used for Rényi and Tsallis entropy when these families were evaluated.

Statistical complexity was computed as the product of normalized entropy and a disequilibrium term:

$$C[P] = H[P]Q[P] \quad (2)$$

The disequilibrium term quantified the distance between the observed distribution and the uniform distribution:

$$Q[P] = Q_0 D(P, P_e) \quad (3)$$

where D is a probability distance or divergence and Q_0 is a normalization constant. The distances systematically evaluated in the geometric search were Jensen-Shannon, Wootters and Euclidean distances. Additional divergences, such as Kullback-Leibler, Hellinger-Euclidean and total variation, were implemented but reserved for appendix-level inventory unless used in the final selected analysis.

The Fisher Information Measure was computed over an ordered discrete probability distribution:

$$FIM = F_0 \sum_{i=1}^{N-1} (\sqrt{p_i + 1} - \sqrt{p_i})^2 \quad (4)$$

Because this quantity is sensitive to local changes in the probability vector, the ordering of probability states was kept fixed across epochs, subjects and groups for each transformation. This constraint is essential for making *FIM* values comparable.

2.7 HC, HFIM And HCFIM Spaces

The HC plane was defined by the pair $(H[P], C[P])$, the *HFIM* plane was defined by $(H[P], F[P])$, and the *HCFIM* space was defined by $(H[P], C[P], F[P])$. The *HC* plane emphasizes the relationship between global uncertainty and statistical structure. The *HFIM* plane emphasizes the relationship between global uncertainty and local probability gradients. The *HCFIM* space integrates both perspectives and was used as the principal geometric representation.

For entropy-complexity planes where the probability domain and disequilibrium definition support it, minimum and maximum complexity curves were used to interpret the feasible region. These curves were not imposed on *HFIM* plots, because Fisher information does not define the same boundary geometry as statistical complexity.

2.8 Cortical Regional and Lateralized Aggregation

EEG features were computed at channel, regional and lateralized-regional levels. The regional schema grouped electrodes into frontal, temporal, central, parietal and occipital regions. The lateralized schema separated homologous left and right regions when possible: frontal left/right, temporal left/right, parietal left/right and occipital left/right, with a central/midline group for C3, Cz, C4, Fz and Pz.

Because the dataset uses traditional 10-20 labels, temporal electrodes were handled consistently across the analysis. If modern canonical names are used in figures or code, the mapping must be declared explicitly $T3 \rightarrow T7$, $T4 \rightarrow T8$, $T5 \rightarrow P7$, $T6 \rightarrow P8$.

For each subject, region, frequency band and epoch, the framework computed $X_{s,r,b,e} \rightarrow P_{s,r,b,e} \rightarrow (H_{s,r,b,e}, C_{s,r,b,e}, F_{s,r,b,e})$.

Epoch-level coordinates were then summarized at subject level using the mean, standard deviation, median, interquartile range, 25th percentile and 75th percentile.

2.9 Frequency Bands and Spectral Baseline Features

The core EEG bands were delta: 0.5-4 Hz; theta: 4-8 Hz; alpha: 8-13 Hz; beta: 13-30 Hz; and broadband: 0.5-40 Hz. For each channel or region, spectral baseline features were computed to provide a conventional comparator to the *HCFIM* representation. These included relative band power in delta, theta, alpha and beta bands, theta/alpha ratio, alpha/beta ratio and spectral entropy. Low-gamma power between 30 and 40 Hz was implemented during operational experimentation, but it is not treated as a central family feature in the present resting-state AD/FTD analysis unless the final reproducible analysis explicitly supports its inclusion.

2.10 Geometric Separability Analysis

Candidate HCFIM configurations were evaluated before classification using geometric separability criteria. For each diagnostic comparison, frequency band, cortical region, entropy family, probability transformation and distance definition, the following quantities were computed at the subject level: centroid distance between groups; within-group spread; inter/intra spread ratio; Fisher ratio; silhouette score; directional displacement in H, C and FIM (Δ_H , Δ_C , Δ_{FIM}); stability across *valid_epoch* and *clean_epoch* subsets.

This stage was used to rank candidate information-theoretic representations that increased between-group separation while preserving within-group stability. Because the dataset was small relative to the number of transformations, bands, regions, distances and classifiers explored, this ranking was treated as an exploratory feature-

discovery procedure rather than as confirmatory evidence of an optimal configuration. The classifier stage was therefore interpreted as a downstream evaluation of selected geometric descriptors under leakage-aware validation, not as an unconstrained search yielding a definitive best model.

2.11 Classification and Validation

Two binary classification problems were evaluated healthy controls versus Alzheimer’s disease, and healthy controls versus frontotemporal dementia.

Feature families were organized into four main groups: PSD-only features, HCFIM-only features, combined PSD + HCFIM features and lateralized HCFIM ablations. The tabular classification pipeline used median imputation, standard scaling and feature selection inside the cross-validation pipeline. Feature selection was never performed before splitting the data.

The core classifiers included Logistic Regression with elastic-net regularization, SVM with RBF kernel, Random Forest, Extra Randomized Trees and Gradient Boosting. Additional classifiers, including LDA with shrinkage, PLS-DA, Gaussian Process classification, calibrated linear models, Gaussian Naive Bayes, XGBoost, LightGBM and small/medium multilayer perceptrons, were explored as appendix-level comparative models. Given the limited sample size, neural network results were interpreted cautiously and used only as comparative evidence, not as the primary methodological claim.

Model evaluation was performed using subject-level cross-validation. For hyperparameter selection, nested validation was preferred when the final configuration required tuning. Out-of-fold probabilities were stored for threshold analysis, ROC curves, confusion matrices and confidence interval estimation.

2.12 Statistical Analysis and Reporting

Classification performance was reported using ROC AUC, balanced accuracy, sensitivity, specificity, F1 score and confusion matrices. Because the intended use case is screening-oriented, sensitivity was considered particularly important; however, specificity was reported jointly to avoid overinterpreting high-sensitivity configurations with excessive false positives.

Bootstrap confidence intervals were computed over subject-level out-of-fold predictions when available. Because many feature families, transformations, regions, frequency bands and classifiers were explored, model rankings were interpreted with caution and not as formal evidence of superiority after multiplicity correction. Permutation testing and independent external validation are required in future work to determine whether selected configurations remain significant under subject-label exchangeability. Feature stability across folds and across quality-control subsets was reported for selected models.

The results section should distinguish between methodological geometry, feature-family comparison and classifier performance. This distinction is important because the scientific claim of the study concerns the information-theoretic representation of EEG signals, while classification is used as downstream evidence of utility.

3. Results

3.1 Reproducible Cohort and Feature Matrices

The complete public OpenNeuro cohort included 88 resting-state EEG recordings: 29 healthy controls, 36 subjects with Alzheimer’s disease and 23 subjects with frontotemporal dementia. All analyses were performed at subject level. Epochs were used only to estimate window-level descriptors before aggregation and were never treated as independent participants in the classifier validation stage.

The regional analysis generated a subject-level matrix with 9,660 features: 210 spectral baseline features and 9,450 HCFIM features. The lateralized_v1 extension generated two matrices: a no-delta matrix with 17,388 features (378 spectral and 17,010 HCFIM features) and a full matrix with 25,116 features after adding 7,728 lateral-difference descriptors. No missing or non-finite numeric feature values were detected in the final subject-level matrices.

Schema	Feature family	Features	Role in analysis
Regional	PSD	210	Conventional spectral baseline
Regional	HCFIM	9,450	Main information-theoretic representation
Regional	PSD + HCFIM	9,660	Combined regional representation
Lateralized_v1	PSD + HCFIM	17,388	Hemisphere-aware extension without lateral deltas
Lateralized_v1	PSD + HCFIM + lateral deltas	25,116	Hemisphere-aware extension with explicit left-right differences

Table 1. Subject-level feature matrices used in the reproducible OpenNeuro analysis.

3.2 Synthetic Geometry

Synthetic signal experiments confirmed that the selected signal-to-probability transformations induce structured information-theoretic geometries. Regular oscillatory, periodic, chaotic, stochastic and colored-noise regimes occupied distinct regions of the probability simplex, HC plane, HFIM plane and HCFIM space. These synthetic results were not used to tune classifiers; their purpose was to verify that each probability transformation produced interpretable coordinates before projecting EEG subjects into the same family of spaces.

The geometry also showed why boundary curves must be interpreted with care. Minimum and maximum complexity curves are meaningful for finite probability domains and specific entropy-distance pairs, but they are not interchangeable across Shannon, Rényi, Tsallis, derived regularity domains or algorithmic-complexity descriptors. HFIM projections were therefore interpreted as empirical regime maps rather than as bounded HC-like manifolds.

3.3 HCFIM Geometric Separability

In the regional schema, the highest-ranked Alzheimer's disease geometry was observed in the temporal alpha band using Permutation entropy with $m=4$, $\tau=1$ and Euclidean disequilibrium. This configuration produced an inter/intra spread ratio of 0.847 and a silhouette score of 0.141. Since silhouette values close to zero indicate substantial overlap rather than clearly separated clusters, the finding should be interpreted as a modest but directionally coherent HCFIM displacement, with higher normalized entropy and statistical complexity in the Alzheimer's group and lower FIM relative to controls.

For frontotemporal dementia, the highest-ranked regional centroid displacement was observed in the temporal broadband signal using ordinal Shannon entropy with $m=4$, $\tau=1$ and Euclidean disequilibrium. The inter/intra spread ratio reached 1.140, with a silhouette score of 0.073. This low silhouette value indicates that the groups remained substantially overlapping; therefore, the result is better described as a weak-to-moderate geometric displacement dominated by higher entropy in the frontotemporal dementia group and lower FIM, rather than as evidence of a well-separated cluster structure.

The lateralized_v1 analysis produced a disease-specific pattern. For Alzheimer's disease, the most separable lateralized configuration by inter/intra ratio was *occipital_right* theta, but it did not improve the geometric separation obtained with the regional temporal alpha configuration. For frontotemporal dementia, lateralization sharpened the temporal finding: *temporal_left* broadband reached an inter/intra ratio of 1.187, suggesting that left temporal information-theoretic displacement is a relevant hypothesis for subsequent validation.

Comparison	Schema	Top region/band	Transform and distance	Ratio / silhouette	Main displacement
Control vs AD	Regional	Temporal alpha	Ordinal Shannon $m=4$, $\tau=1$, Euclidean	0.847 / 0.141	Higher H and C; lower FIM
Control vs AD	Lateralized_v1	Occipital right theta	Ordinal Shannon $m=4$, $\tau=1$, Euclidean	0.801 / 0.063	Lower H and C; higher FIM
Control vs FTD	Regional	Temporal broadband	Ordinal Shannon $m=4$, $\tau=1$, Euclidean	1.140 / 0.073	Higher H; lower FIM
Control vs FTD	Lateralized_v1	Temporal left broadband	Ordinal Shannon $m=4$, $\tau=1$, Euclidean	1.187 / 0.079	Higher H; lower FIM

Table 2. Highest-ranked HCFIM separability configurations for regional and lateralized schemas.

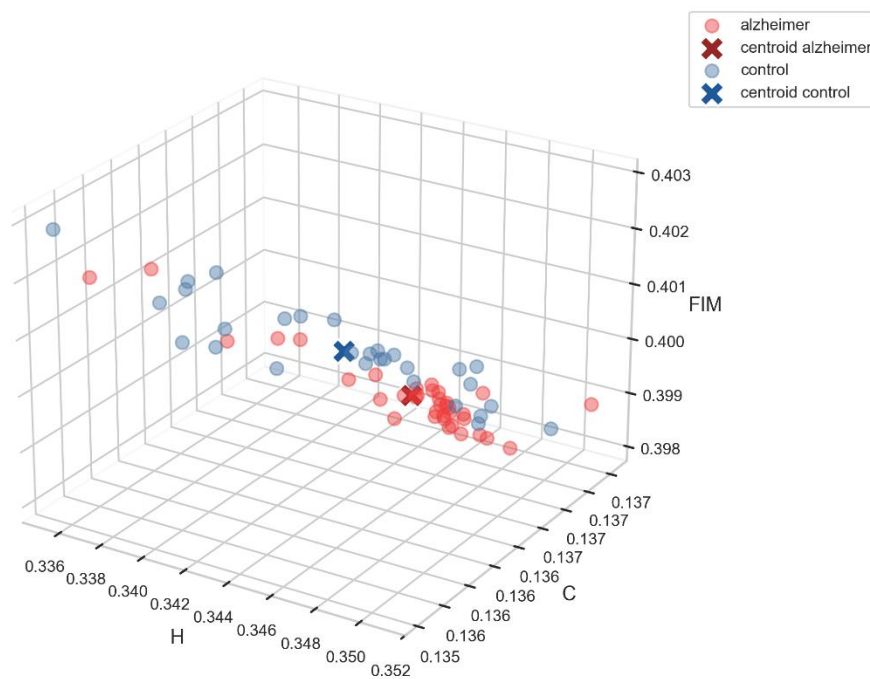


Figure 2. Regional HCFIM projection for the highest-ranked control versus Alzheimer's disease configuration.

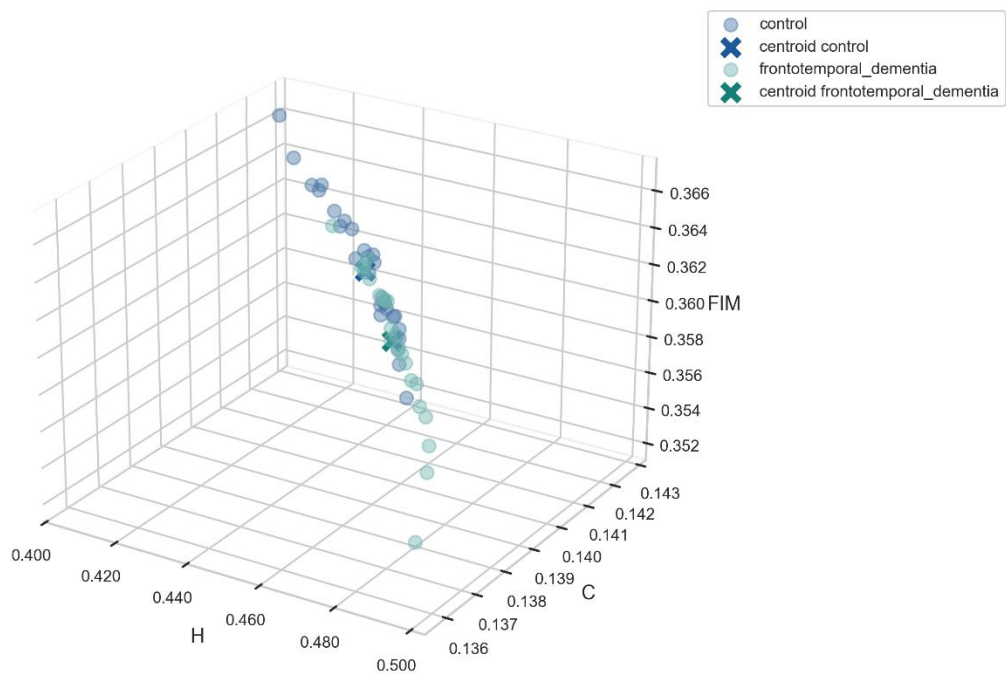


Figure 3. Regional HCFIM projection for the highest-ranked control versus frontotemporal dementia configuration.

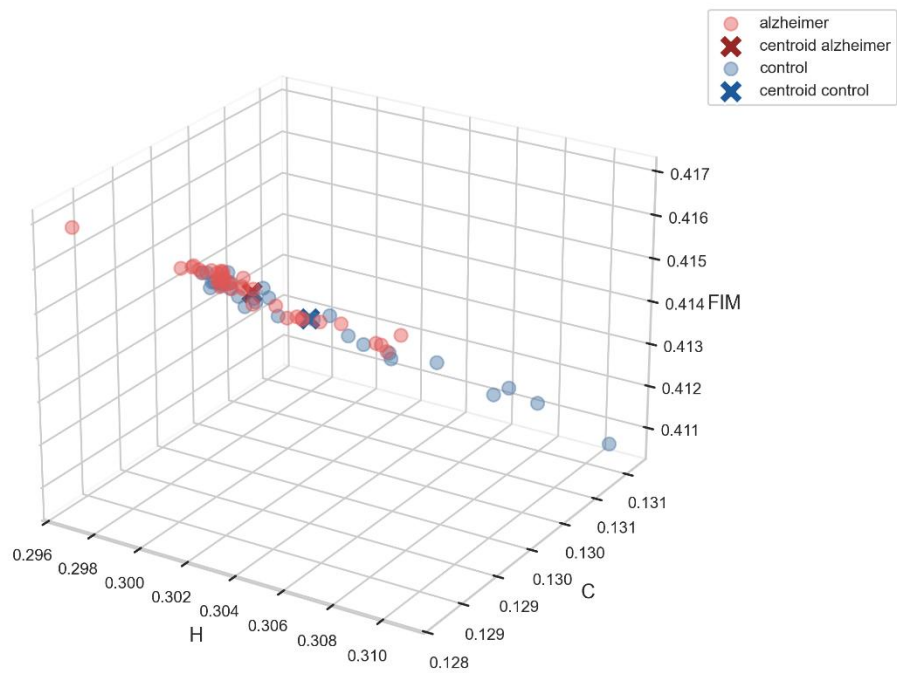


Figure 4. Lateralized_v1 HCFIM projection for control versus Alzheimer's disease.

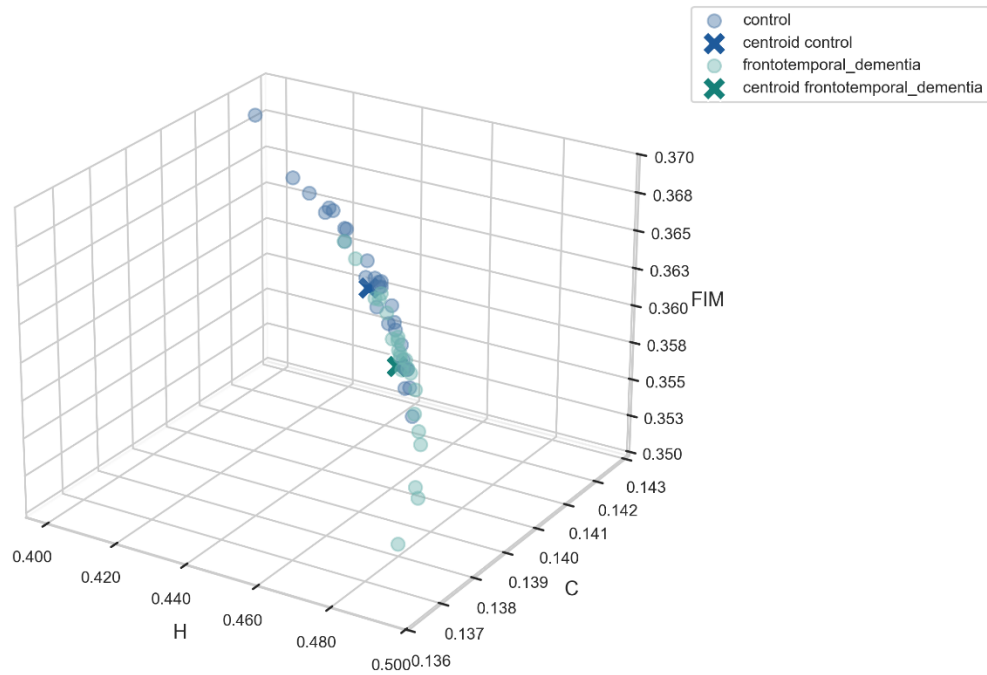


Figure 5. Lateralized_v1 HCFIM projection for control versus frontotemporal dementia.

The frontotemporal dementia broadband projection provides the most visually interpretable contrast among the tested topographic summaries, while still showing considerable group overlap. Patient recordings show higher entropy (H) and lower statistical complexity (C) and FIM relative to controls after regional projection onto the scalp layout (Figure 6).

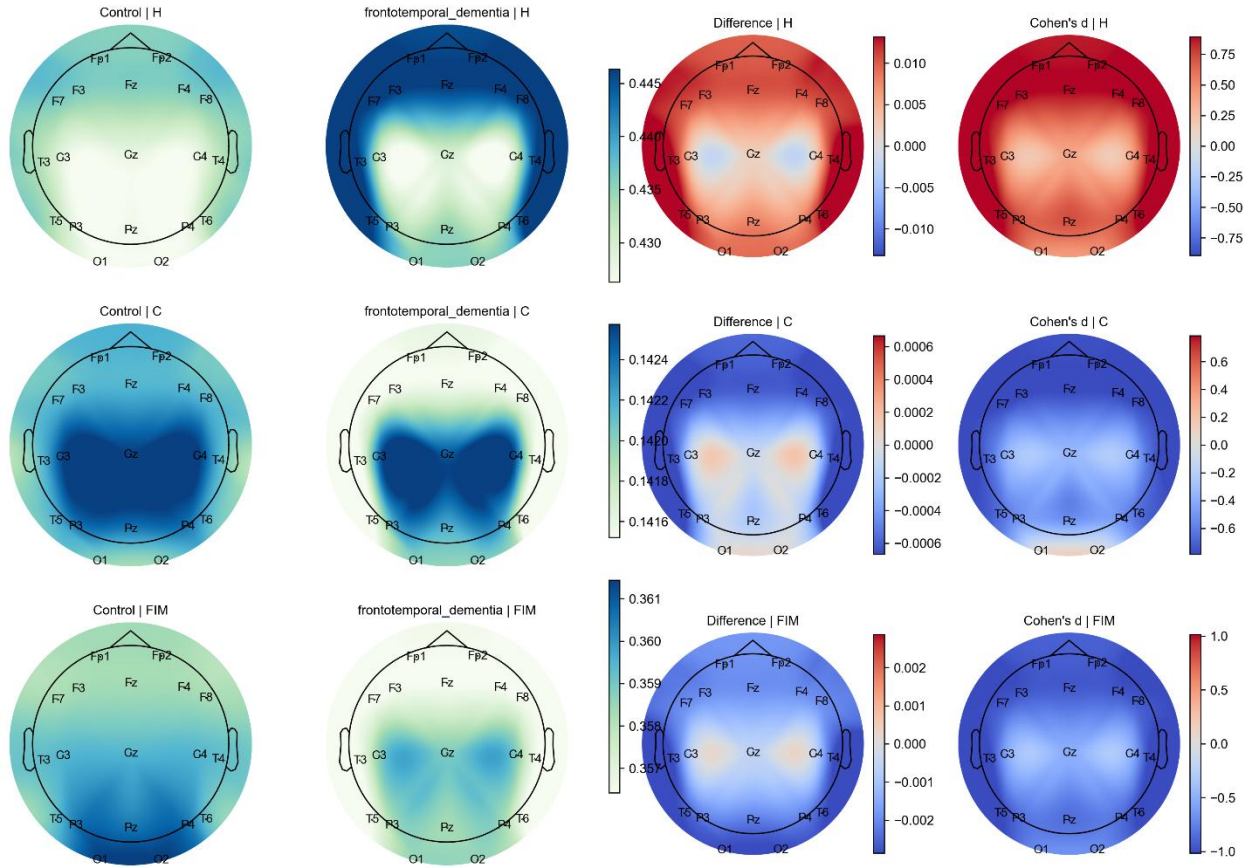


Figure 6. HCFIM regional topographic projection for control versus frontotemporal dementia using Permutation entropy ($m = 4$, $\tau = 1$), Euclidean disequilibrium and the broadband signal. Regional HCFIM values were assigned to their corresponding 10-20 electrodes for visualization only.

3.4 Classification Performance

For Alzheimer's disease, the best regional classifier combined PSD and HCFIM features with an Extra Trees model, reaching ROC AUC = 0.818, balanced accuracy = 0.803, sensitivity = 0.778 and specificity = 0.828. In the lateralized_v1 schema, adding lateral-difference descriptors increased ROC AUC to 0.832 and sensitivity to 0.833, although specificity decreased to 0.759. Thus, lateralized differences improved the screening-oriented sensitivity profile but did not dominate the regional model across all metrics.

For frontotemporal dementia, the best regional ROC AUC was obtained by an HCFIM-only small multilayer perceptron (ROC AUC = 0.810), while the best balanced accuracy was obtained by a regional PSD + HCFIM Extra Trees model (balanced accuracy = 0.740; ROC AUC = 0.775). In the lateralized_v1 schema, the best ROC AUC was lower (0.741 for PSD + HCFIM Logistic Regression), indicating that the lateralized representation was informative geometrically but did not improve downstream classification in this sample.

Across both diagnostic comparisons, models based exclusively on lateral deltas were not competitive. The most consistent evidence therefore supports HCFIM as a complementary representation to PSD, with lateralized features treated as a hypothesis-generating extension rather than as a universal improvement. Bootstrap confidence intervals were wide and overlapped across regional and lateralized schemas; consequently, small differences in ROC AUC should not be interpreted as statistically established improvements.

Comparison	Schema	Feature set / model	ROC AUC	Conf. Intervals	Balanced acc.	Sens. / Spec.
Control vs AD	Regional	PSD + HCFIM / Extra Trees	0.818	0.696 - 0.925	0.803	0.778 / 0.828
Control vs AD	Lateralized_v1	PSD + HCFIM + lateral deltas / Extra Trees	0.832	0.721 - 0.927	0.796	0.833 / 0.759
Control vs FTD	Regional	HCFIM / MLP small	0.810	0.677 - 0.926	0.705	0.652 / 0.759
Control vs FTD	Regional	PSD + HCFIM / Extra Trees	0.775	0.634 - 0.901	0.740	0.652 / 0.828
Control vs FTD	Lateralized_v1	PSD + HCFIM / Logistic ElasticNet	0.741	0.592 - 0.865	0.675	0.696 / 0.655

Table 3. Best subject-level classification results from out-of-fold predictions.

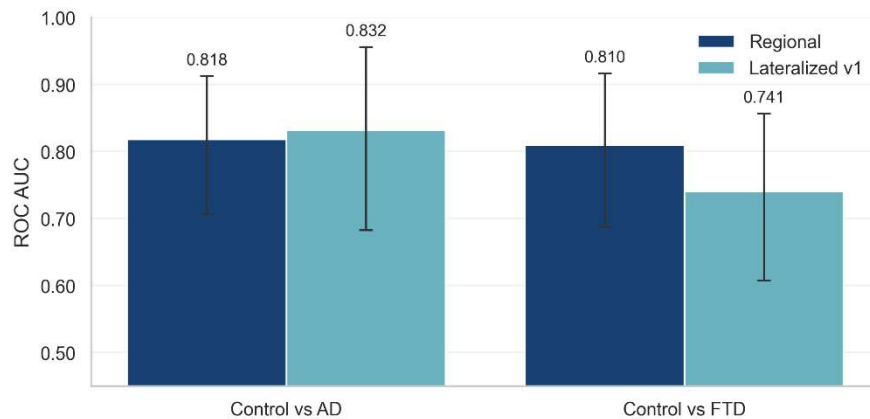


Figure 7. Best ROC AUC comparison between regional and lateralized_v1 schemas by diagnostic task.

4. Discussion

4.1 Principal Findings

This study supports the use of HCFIM geometry as a reproducible, interpretable representation for resting-state EEG in a data-limited neurodegenerative screening setting. The central finding is not that a single classifier solves Alzheimer’s disease or frontotemporal dementia detection, but that subject-level EEG signals can be projected into stable information-theoretic spaces where disease-related shifts become quantifiable by region, brain hemisphere, band, entropy family, distance and FIM displacement.

The results also suggest that the HCFIM representation captures information that is not reducible to PSD alone. For Alzheimer’s disease, PSD + HCFIM features yielded the highest regional classifier performance and the best lateralized screening-oriented sensitivity. For frontotemporal dementia, HCFIM-only features produced the highest regional ROC AUC, while PSD + HCFIM produced the best balanced accuracy. However, the confidence intervals around these estimates were broad and overlapping, so these differences should be interpreted as exploratory evidence of complementarity rather than as proof that one schema or classifier is statistically superior.

4.2 Disease-Specific Geometry

The strongest Alzheimer's disease geometry emerged in the temporal alpha band in the regional schema. This finding is consistent with the broader resting-state EEG literature in which alpha rhythm alterations and loss of organized neural dynamics are repeatedly associated with cognitive decline (Jeong, 2004; Cassani et al., 2018; Babiloni et al., 2021). However, the observed HCFIM displacement was modest; therefore, the result should be interpreted as a discriminative statistical-geometric pattern rather than as a standalone biomarker.

For frontotemporal dementia, the temporal broadband HCFIM displacement was larger and became more localized to the left temporal region in the lateralized schema. This is neurophysiologically plausible given the temporal and frontal involvement typical of frontotemporal dementia syndromes, but the classifier results caution against overclaiming: the more localized geometry did not translate into superior lateralized classification performance in this small cohort.

4.3 Regional Versus Lateralized Features

The lateralized_v1 experiment is best interpreted as an ablation study. It increased dimensionality and revealed meaningful left-right hypotheses, particularly for frontotemporal dementia, but it did not provide uniform gains. In Alzheimer's disease, lateralized deltas slightly improved ROC AUC and sensitivity at the cost of specificity. In frontotemporal dementia, the regional schema remained preferable for classification.

Therefore, the current evidence favors a conservative modeling strategy: report regional HCFIM as the primary reproducible result, and present lateralized_v1 as an exploratory extension for hypothesis generation and future validation. This distinction strengthens the paper because it separates methodological discovery from operational model selection.

4.4 Classifier Capacity and Data-Limited Learning

The comparison with neural-network models is informative. A small multilayer perceptron achieved the highest regional ROC AUC for frontotemporal dementia using HCFIM-only features, but this result must be interpreted cautiously because the dataset contains only 52 subjects for that binary comparison. More complex models can exploit high-dimensional feature spaces but are also more vulnerable to variance and optimistic performance estimates.

For this reason, the interpretation emphasizes leakage-aware subject-level validation, feature selection inside the cross-validation pipeline and out-of-fold prediction storage. The methodological claim is stronger when performance is presented together with feature-family comparisons, geometric separability and explicit limitations, rather than as a competition among classifiers.

4.5 Limitations

Five limitations must be stated explicitly. First, the study used a single public dataset with 88 subjects and no independent external cohort. Second, the analysis relied on preprocessed resting-state derivatives, so results may depend on preprocessing choices made before feature extraction. Third, the feature search space was broad relative to the sample size: multiple entropy families, distances, frequency bands, cortical schemas, feature sets and classifiers were explored. Although validation was performed at subject level and feature selection was kept inside the cross-validation pipeline, the reported best configurations may still reflect selection optimism. The results should therefore be interpreted as exploratory and hypothesis-generating until confirmed by permutation testing, multiplicity-aware analysis and independent cohorts.

Fourth, age and MMSE severity were not modeled as covariates in the primary HCFIM analyses. Although controls, AD and FTD groups differ by diagnostic category, part of the observed information-geometric displacement may correlate with cognitive severity or demographic structure rather than diagnosis alone. Future work should evaluate HCFIM coordinates with age- and MMSE-adjusted models, severity correlations and stratified sensitivity

analyses. Fifth, the study does not validate device-specific Contec or Emotiv models. Device adaptation, operational JSON outputs and Nouha report generation belong to the translational software layer, not to the scientific claim of this manuscript. Consequently, the framework should not be interpreted as a diagnostic system. Its appropriate framing is preclinical, experimental screening and methodological exploration of information-theoretic EEG geometry.

5. Conclusions

This study presents a reproducible information-theoretic framework for characterizing resting-state EEG dynamics in Alzheimer's disease, frontotemporal dementia, and healthy controls under data-limited conditions. Rather than positioning HCFIM as a standalone clinical classifier, the work demonstrates how EEG signals can be transformed into probability distributions and represented geometrically through entropy, statistical complexity, and the Fisher Information Measure. This subject-level representation enables intergroup differences to be inspected as regional shifts in an informational space before classification, extending the analysis beyond conventional spectral power features.

The main evidence supports regional HCFIM features as the central methodological and scientific contribution. In Alzheimer's disease, the combination of PSD and HCFIM features was associated with better classification than either feature family alone in this dataset, while lateralized delta-based differences provided a small screening-oriented gain in ROC AUC and sensitivity. In frontotemporal dementia, HCFIM features captured an informative regional geometry, with temporal broadband and left-temporal patterns suggesting plausible disease-specific structure, although lateralization did not consistently improve classifier performance. Given the broad search space and overlapping confidence intervals, these findings should be framed as exploratory evidence that HCFIM complements PSD by capturing nonlinear and probabilistic aspects of EEG organization.

Overall, the study contributes a reproducible EEG geometry workflow, a leakage-aware validation strategy, and a comparative analysis of spectral, HCFIM, regional, and lateralized feature spaces. The findings support HCFIM as a promising methodological contribution for data-limited EEG screening research, particularly as a framework for exploring disease-specific neurophysiological organization. However, the results should be interpreted as exploratory and preclinical. Future work should validate the selected information theory methods configurations in independent cohorts before translation into operational screening pipelines.

Declaration of Competing Interest

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments: This research was funded by ANID National Doctoral Scholarship N° 21221884.

References

- Babiloni, C., et al. (2021). Measures of resting state EEG rhythms for clinical trials in Alzheimer's disease: recommendations of an expert panel. <https://pmc.ncbi.nlm.nih.gov/articles/PMC8647863/>
- Bandt, C., & Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, 88, 174102. <https://doi.org/10.1103/PhysRevLett.88.174102>
- Bang, J., Spina, S., & Miller, B. L. (2015). Frontotemporal dementia. *The Lancet*, 386(10004), 1672-1682. [https://doi.org/10.1016/S0140-6736\(15\)00461-4](https://doi.org/10.1016/S0140-6736(15)00461-4)
- Cai, G., Shi, Y., Ma, J., Zhang, X., & Wu, W. (2026). Multi-dimensional EEG analysis reveals distinct neurophysiological patterns in Alzheimer's and frontotemporal dementia. *Journal of NeuroEngineering and Rehabilitation*, 23, 81. <https://doi.org/10.1186/s12984-026-01897-9>
- Cassani, R., Estarellas, M., San-Martin, R., Fraga, F. J., & Falk, T. H. (2018). Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment. *Disease Markers*, 2018, 5174815. <https://doi.org/10.1155/2018/5174815>
- Dauwels, J., Vialatte, F., & Cichocki, A. (2010). Diagnosis of Alzheimer's disease from EEG signals: Where are we standing? *Current Alzheimer Research*, 7(6), 487-505. <https://doi.org/10.2174/156720510792231720>
- Deng, B., Cai, L., Li, S., Wang, R., Yu, H., Chen, Y., & Wang, J. (2017). Multivariate multi-scale weighted permutation entropy analysis of EEG complexity for Alzheimer's disease. *Cognitive Neurodynamics*, 11(3), 217-231. <https://doi.org/10.1007/s11571-016-9418-9>
- Echegoyen, I., Lopez-Sanz, D., Martinez, J. H., Maestu, F., & Buldu, J. M. (2020). Permutation entropy and statistical complexity in mild cognitive impairment and Alzheimer's disease: An analysis based on frequency bands. *Entropy*, 22(1), 116. <https://doi.org/10.3390/e22010116>
- Jeong, J. (2004). EEG dynamics in patients with Alzheimer's disease. *Clinical Neurophysiology*, 115(7), 1490-1505. <https://doi.org/10.1016/j.clinph.2004.01.001>
- Martin, M. T., Plastino, A., & Rosso, O. A. (2006). Generalized statistical complexity measures: Geometrical and analytical properties. *Physica A: Statistical Mechanics and its Applications*, 369(2), 439-462. <https://doi.org/10.1016/j.physa.2005.11.053>
- 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>
- Mlinaric, T., Van Den Kerchove, A., Barinaga, Z. I., & Van Hulle, M. M. (2026). EEG-based classification of Alzheimer's disease and frontotemporal dementia using functional connectivity. *Scientific Reports*, 16(1), 1-13. <https://doi.org/10.1038/s41598-026-35316-9>
- Morabito, F. C., Labate, D., La Foresta, F., Bramanti, A., Morabito, G., & Palamara, I. (2012). Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer's disease EEG. *Entropy*, 14(7), 1186-1202. <https://doi.org/10.3390/e14071186>
- NEMAR/OpenNeuro. A dataset of EEG recordings from Alzheimer's disease, frontotemporal dementia and healthy subjects. <https://nemar.org/dataset/on004504>
- Rojas-Pescio, H., Beishon, L., Panerai, R. B., & Chacón, M. (2024). Statistical Complexity Analysis of Neurovascular Coupling with Cognitive Stimulation in Healthy Participants. *Journal of Cognitive Neuroscience*, 36(9), 1995-2010.

Rojas-Pescio, H., Beishon, L., Panerai, R. B., & Chacón, M. (2025). Cognitive brain lateralization through neurovascular coupling in healthy subjects: A statistical complexity analysis. *Physiological Reports*, 13(15), e70492.

Rosso, O. A., Larrondo, H. A., Martin, M. T., Plastino, A., & Fuentes, M. A. (2007). Distinguishing noise from chaos. *Physical Review Letters*, 99(15), 154102. <https://doi.org/10.1103/PhysRevLett.99.154102>

Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chetelat, G., Teunissen, C. E., Cummings, J., & van der Flier, W. M. (2021). Alzheimer's disease. *The Lancet*, 397(10284), 1577-1590. [https://doi.org/10.1016/S0140-6736\(20\)32205-4](https://doi.org/10.1016/S0140-6736(20)32205-4)

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., et al. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups. *Alzheimer's & Dementia*, 7(3), 280-292. <https://doi.org/10.1016/j.jalz.2011.03.003>

Sun, J., Wang, B., Niu, Y., Tan, Y., Fan, C., Zhang, N., Xue, J., Wei, J., & Xiang, J. (2020). Complexity analysis of EEG, MEG, and fMRI in mild cognitive impairment and Alzheimer's disease: A review. *Entropy*, 22(2), 239. <https://doi.org/10.3390/e22020239>

Vignat, C., & Bercher, J.-F. (2003). Analysis of signals in the Fisher-Shannon information plane. *Physics Letters A*, 312(1-2), 27-33. [https://doi.org/10.1016/S0375-9601\(03\)00570-X](https://doi.org/10.1016/S0375-9601(03)00570-X)

Appendices

These appendices document the exploratory methodological space evaluated during the HCFIM EEG research phase. They are intentionally more exhaustive than the main manuscript. Their purpose is to make the work reproducible, auditable and clear about which analyses support the main claim and which analyses were exploratory.

Appendix A. Entropy and Probability-Transform Inventory

The HCFIM framework does not operate directly on the raw EEG time series. Each signal window is first mapped into a finite probability distribution. Entropy, statistical complexity and Fisher Information Measure are then computed over that distribution.

$$EEG\ window \rightarrow probability\ transform \rightarrow P \rightarrow H, C, FIM$$

This appendix lists the probability transformations implemented or explored. Only transformations that define a comparable finite probability distribution should be interpreted as direct HC/HFIM/HCFIM embeddings.

A.1 Direct Probability Domains

Family	Probability domain	Main parameters	HCFIM status
Shannon finite-state entropy	Generic finite probability vector	number of states N	Direct
Permutation entropy	Bandt-Pompe ordinal-pattern probabilities	$m = 3, 4, 5; \tau = 1$	Direct
Rényi ordinal entropy	Bandt-Pompe ordinal-pattern probabilities with Rényi entropy	$\alpha = 0.8, 1.2; m = 3; \tau = 1$	Direct
Tsallis ordinal entropy	Bandt-Pompe ordinal-pattern probabilities with Tsallis entropy	$q = 0.8, 1.2; m = 3; \tau = 1$	Direct
Dispersion entropy	Dispersion symbol-pattern probabilities	$c = 3-5$; embedding dimension selected per experiment	Direct

Family	Probability domain	Main parameters	HCFIM status
Distribution entropy	Histogram of distances between embedded vectors	bins = 16	Direct if binning is fixed
Gridded entropy	Occupancy probabilities in a normalized phase-space grid	grid size fixed per experiment	Direct
Increment entropy	Quantized sign-amplitude increment words	m = 3; quantization level fixed	Direct
Phase entropy	Angular-sector probabilities	number of sectors fixed	Direct
Slope entropy	Slope-symbol word probabilities	angular thresholds fixed	Direct
Spectral entropy	Normalized spectral power distribution	frequency bins fixed	Direct
Wavelet entropy	Relative wavelet-energy distribution	wavelet level fixed	Direct
Normality entropy	Symbolic distribution derived from local normality classes	c = 3; m = 3	Direct after class definition

A.2 Entropy Families

The direct entropy inventory can be grouped as follows:

- **Global finite-state entropy:** Shannon entropy.
- **Ordinal entropies:** permutation entropy and ordinal Rényi/Tsallis variants.
- **Symbolic amplitude entropies:** dispersion, increment, slope and normality entropy.
- **Embedding-distance entropy:** distribution entropy.
- **Time-frequency entropies:** spectral and wavelet entropy.
- **Phase-space occupancy entropy:** gridded entropy.

For all direct probability domains, the finite state space must remain fixed within a given experiment. Otherwise, entropy, complexity and FIM values become non-comparable across subjects or groups.

A.3 Parameter Traceability

The parameter set used to generate each HCFIM point must be stored with the result:

- entropy_family
- probability_transform
- embedding_dimension
- delay
- bins_or_symbols
- entropy_parameter
- distance_measure
- frequency_band
- region
- epoch_quality_subset

This traceability is essential because different probability transformations can place the same physiological signal in different regions of the HC, HFIM and HCFIM spaces.

Appendix B. Derived Information-Theoretic and Algorithmic-Complexity Descriptors

Some measures used in EEG nonlinear analysis do not naturally define the same probability simplex used by entropy-complexity methods. Approximate Entropy, Sample Entropy, Fuzzy Entropy, Lempel-Ziv Complexity and compression-based complexity are important descriptors, but they require an explicit additional construction before they can be projected into HC, HFIM or HCFIM coordinates.

For this reason, the paper should refer to them as:

- derived information-theoretic / algorithmic-complexity descriptors
- unless a probability domain and disequilibrium term are explicitly defined.

B.1 Regularity-Derived Domains

Descriptor	Classical interpretation	Derived probability domain used for geometry
Approximate Entropy	Regularity and self-similarity of embedded patterns	Histogram of Chebyshev distances between embedded vectors
Sample Entropy	Regularity estimated without self-matches	Histogram of Chebyshev distances for $m+1$ embedded vectors
Fuzzy Entropy	Noise-robust regularity through fuzzy similarity	Histogram of fuzzy similarity values between centered embedded vectors
Kolmogorov/correlation entropy	Correlation structure across embedding radii	Normalized increments of the correlation integral across radii

These domains are derived from the internal quantities used by regularity measures. They are not identical to the scalar ApEn/SampEn/FuzzyEn values. The derived domain is used to obtain a probability vector suitable for HCFIM projection.

B.2 Algorithmic-Complexity Domains

Descriptor	Classical interpretation	Derived probability domain or feature
Lempel-Ziv binary complexity	Phrase-generation complexity after binary symbolization	LZ76 phrase-length distribution
Lempel-Ziv quantile complexity	Phrase-generation complexity after quantile symbolization	Phrase-length distribution over quantile symbols
Lempel-Ziv ordinal complexity	Phrase-generation complexity after ordinal symbolization	Phrase-length distribution over ordinal words
Compression complexity	Compressibility of a symbolic representation	zlib/bz2/lzma compressed-length ratio

Lempel-Ziv variants can be projected into HCFIM only after the phrase inventory or phrase-length distribution is normalized into a probability vector. Compression complexity is better treated as a supplementary scalar feature unless the encoding output is explicitly converted into a probability domain.

B.3 Multiscale Families

Multiscale family	Role in the study	Recommended location
Multiscale Entropy	Scale-wise regularity profile	Appendix or future extension
Multiscale Permutation Entropy	Scale-wise ordinal probability geometry	Appendix if selected

Multiscale Fuzzy Entropy	Scale-wise fuzzy regularity	Appendix
Multiscale Dispersion Entropy	Scale-wise dispersion-symbol geometry	Appendix if selected

Multiscale features are scientifically relevant for physiological complexity, but they increase the search space substantially. For the present OpenNeuro analysis, they should be included only if the final reproducible notebook shows stable subject-level gains.

Appendix C. Probability Distances and Complexity-Boundary Curves

C.1 Distances used in the HCFIM Search

Distance		Role	Main use
Jensen-Shannon	Symmetric divergence from the equiprobable distribution		Main HC/HCFIM distance
Wootters	Geometric/statistical distance between probability distributions		Main HC/HCFIM distance
Euclidean	Direct vector-space distance to the uniform distribution		Comparator distance

C.2 Additional Implemented Distances

The implementation also included Kullback-Leibler divergence, Hellinger-Euclidean distance, total variation divergence and SDL complexity. These were considered exploratory and should not be emphasized in the main manuscript unless selected by the final notebook.

C.3 Minimum and Maximum Complexity Curves

For HC planes, minimum and maximum complexity curves can be generated when the probability space is finite, and the complexity definition is based on a disequilibrium term relative to the equiprobable distribution. The curves must be generated separately for each:

- Entropy family
- Entropy parameter
- State-space dimension
- Distance measure
- Embedding configuration

Therefore, the Cmin/Cmax curves for Shannon, Rényi and Tsallis entropy are not interchangeable. The same applies when changing the ordinal dimension m , the number of bins or the probability transform.

HFIM plots should not be forced into the same boundary interpretation. Fisher Information Measure captures local gradients over the ordered probability vector and does not define the same feasible envelope as statistical complexity.

Appendix D. Feature Extraction, Hyperparameter Grids and Classifier Inventory

D.1 Frequency Bands

The core bands used for EEG feature extraction were:

Band	Range
delta	0.5-4 Hz
theta	4-8 Hz
alpha	8-13 Hz

beta	13-30 Hz
broadband	0.5-40 Hz

Low gamma between 30 and 40 Hz was implemented during operational development but is not part of the primary scientific claim.

D.2 Cortical Schemas

The regional schema grouped electrodes into frontal, temporal, central, parietal, occipital.

The lateralized schema grouped electrodes into frontal_left, frontal_right, temporal_left, temporal_right, parietal_left, parietal_right, occipital_left, occipital_right, central_midline_extended

If modernized labels are used, the following mapping must be declared T3 $\rightarrow$ T7, T4 $\rightarrow$ T8, T5 $\rightarrow$ P7, T6 $\rightarrow$ P8.

D.3 Subject-Level Aggregations

Epoch-level features were summarized per subject using mean, standard deviation, median, interquartile range, p25, p75. The classifier input was always the subject-level feature matrix.

D.4 HCFIM Search Grid

The HCFIM candidate search explored:

Component	Values explored
Entropy/transform	Permutation, Rényi, Tsallis, ApEn-derived, SampEn-derived, Fuzzy-derived, K2/correlation-derived, LZ ordinal phrase-derived
Ordinal dimensions	m = 3, 4, 5 for Permutation; m = 3 for Rényi/Tsallis variants
Delay	$\tau = 1$
Rényi parameters	alpha = 0.8, alpha = 1.2
Tsallis parameters	q = 0.8, q = 1.2
Derived-domain bins	bins = 16
Distances	Jensen-Shannon, Wootters, Euclidean
Bands	delta, theta, alpha, beta, broadband
Regions	regional and lateralized schemas
QC subsets	valid_epoch, clean_epoch

D.5 Feature Families

Feature set	Description
PSD only	Relative bandpower, ratios and spectral entropy
HCFIM only	Entropy, complexity and FIM features only
PSD + HCFIM	Combined spectral and geometric representation
Lateralized HCFIM	Selected HCFIM features from hemispheric cortical regions

D.6 Core Classifiers

The core tabular classifiers were Logistic Regression with elastic-net regularization, SVM with RBF kernel, Random Forest, Extra Randomized Trees, Gradient Boosting. Each model was embedded in a leakage-aware pipeline including median imputation, scaling where appropriate and feature selection inside cross-validation.

D.7 Appendix Classifiers

Additional classifiers explored during method development were LDA with automatic shrinkage, PLS-DA, Nystroem(RBF) + LDA, Gaussian Process Classifier with RBF kernel, calibrated LDA, calibrated Logistic Regression, Histogram Gradient Boosting, Gaussian Naive Bayes, XGBoost regularized (when available), LightGBM regularized (when available) and small and medium multilayer perceptrons.

These models should be reported in supplementary tables. Given the small sample size, MLP and high-capacity boosted-tree results should be interpreted as comparative baselines, not as the primary evidence of the method.

Appendix E. Leakage Audit and Subject-Level Validation

The principal validation risk in EEG epoch-based studies is leakage across windows from the same subject. To control this risk, the following rules were applied:

1. Epochs were never treated as independent participants.
2. Train/test partitions were created at subject level.
3. Feature selection was performed inside the validation pipeline.
4. Scaling and imputation were fitted only on training folds.
5. Hyperparameter tuning, when used, was nested or restricted to training data.
6. Out-of-fold predictions were stored for model-level summaries.
7. Threshold selection was performed on out-of-fold probabilities.
8. Confidence intervals were computed over subject-level predictions.

E.1 Primary Metrics

The following metrics were reported ROC AUC, balanced accuracy, sensitivity, specificity, F1 score and confusion matrix. Because the intended downstream use is screening-oriented, sensitivity was considered especially important. However, specificity and balanced accuracy were reported jointly to avoid selecting models that simply overcall pathology.

E.2 Confidence Intervals and Significance

Bootstrap confidence intervals were computed when out-of-fold predictions were available. These intervals quantify uncertainty in subject-level predictive estimates but do not by themselves correct for the large number of model and feature configurations explored. Formal permutation testing under subject-label exchangeability and multiplicity-aware sensitivity analyses are required before making confirmatory claims about the superiority of any selected configuration.

E.3 Stability Checks

Feature and model stability were assessed by comparing valid_epoch versus clean_epoch subsets; PSD-only versus HCFIM-only versus PSD + HCFIM; regional versus lateralized aggregation; selected feature frequency across folds; and direction of group displacement in H, C and FIM.

Appendix F. Out-of-Fold Classifier Curves and Confusion Matrices

Detailed classifier visualizations are provided as appendix material to keep the main manuscript focused on the HCFIM geometry and topographic feature projections. Each panel summarizes out-of-fold predictions at subject level.

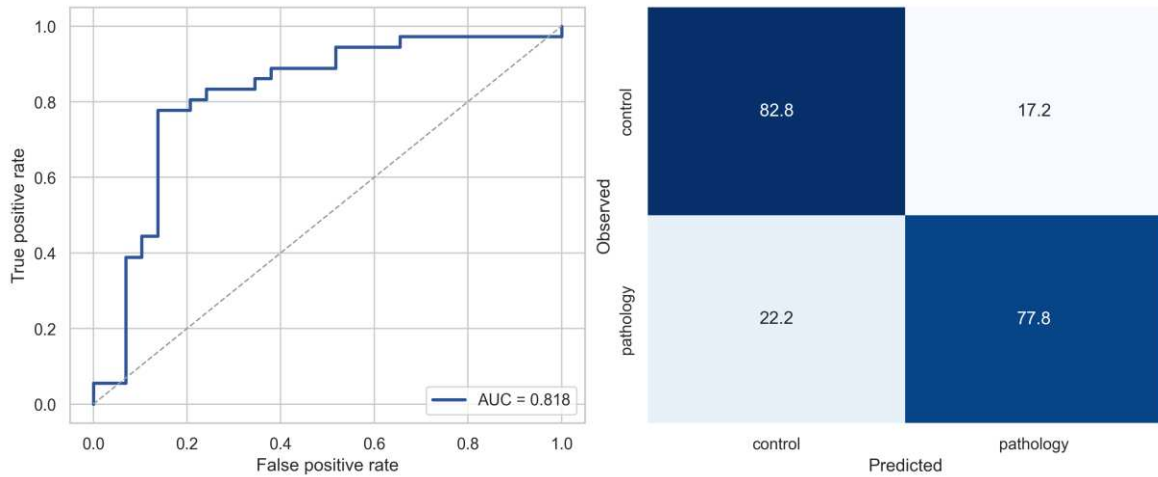


Figure F1. Out-of-fold classification summary for the best regional control versus Alzheimer's disease model.

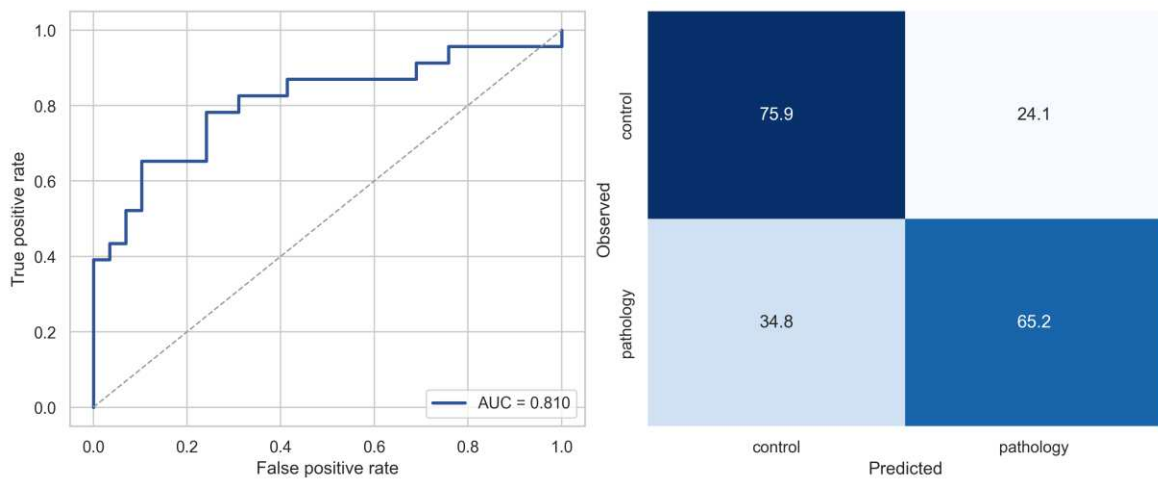


Figure F2. Out-of-fold classification summary for the best regional control versus frontotemporal dementia model.

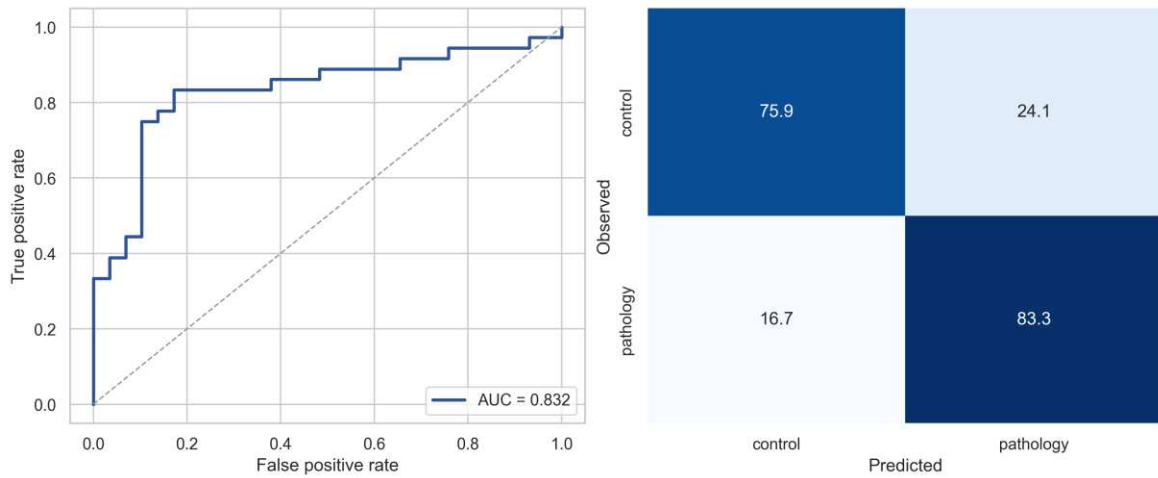


Figure F3. Out-of-fold classification summary for the best lateralized_v1 control versus Alzheimer's disease model.

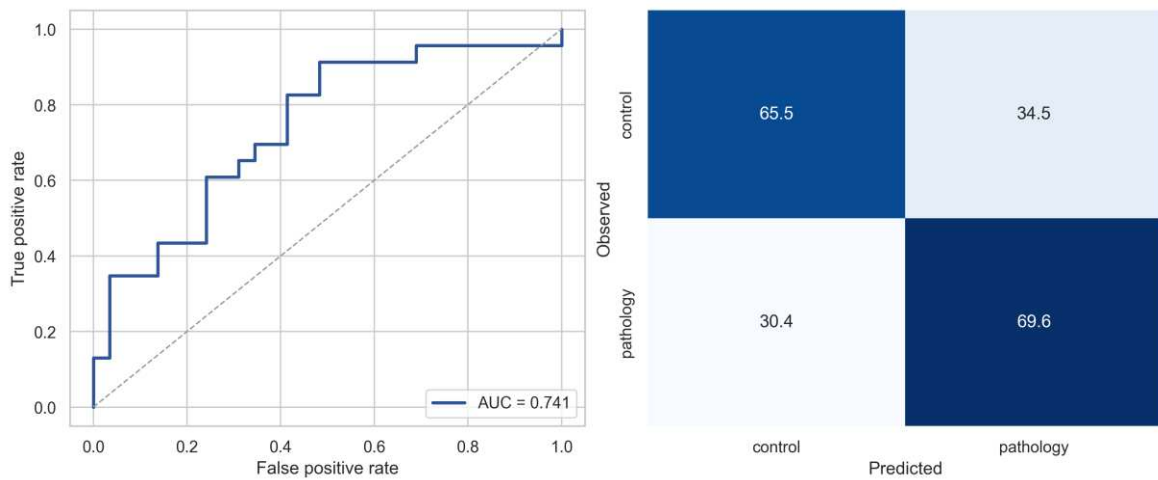


Figure F4. Out-of-fold classification summary for the best lateralized_v1 control versus frontotemporal dementia model.