(Online) ISSN 2712-0562
Original Research
Sustainable Engineering and Innovation
Vol. 8, No. 1, 2026, pp.73-92
https://doi.org/10.37868/sei.v8i1.id736

Morlet wavelet–based olfactory-evoked EEG features for random forest classification of normal, aMCI, and Alzheimer's disease

Nabila A. Alsharif
Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq

Corresponding author E-mail: nabila_alsharif@coadec.uobaghdad.edu.iq

Received Sep. 29, 2024
Revised Dec. 25, 2025
Accepted Jan. 12, 2026
Online Feb. 17, 2026

Abstract

Olfactory impairment and abnormal frontal EEG oscillations are recognized as early markers of Alzheimer's disease (AD). Using a publicly available olfactory EEG dataset of 35 subjects spanning normal cognition, amnestic mild cognitive impairment (aMCI), and AD, each with MMSE scores and demographics, stimulus-locked epochs from four electrodes (Fp1, Fz, Cz, Pz) were processed with wavelet-based time–frequency analysis. Band-limited power ratios (delta, theta, alpha, beta) were computed as log-transformed post-odor/baseline values and aggregated to subject-level features. Statistical analyses revealed graded attenuation of odor-evoked frontal (Fp1) band-power ratios across groups, with significant differences in several band–odor combinations. PCA of Fp1 features showed partial separation of diagnostic categories, while multi-channel features offered weaker discrimination. Random forest classifiers trained on Fp1-only features achieved 66.7% test accuracy, outperforming the four-channel model (55.6%), with moderate sensitivity, specificity, and precision. These findings highlight that compact frontal wavelet-derived band-power ratios during olfactory stimulation carry diagnostically relevant information for distinguishing Normal, aMCI, and AD. The transparent pipeline, combining time–frequency processing, subject-level aggregation, and multiclass classification, offers a scalable framework that can be extended to larger cohorts or integrated with multimodal biomarkers.

Keywords: EEG, Alzheimer's disease, Amnestic mild cognitive impairment (aMCI), Olfactory stimulation, Morlet wavelet, Multiclass classification

1. Introduction

Alzheimer's disease (AD) and amnestic mild cognitive impairment (aMCI) form a continuum of progressive cognitive decline. Identifying early biomarkers is crucial because neurodegenerative changes may precede clinical symptoms by years. Electroencephalography (EEG), a non-invasive technique that measures cortical electrical activity, offers excellent temporal resolution to detect such early abnormalities [1]. The olfactory system is among the earliest sites affected in AD. Pathological changes begin in the olfactory bulb and primary olfactory cortices long before memory complaints emerge. Therefore, odor-evoked EEG responses represent a compelling candidate for detecting subtle functional impairments [2].

Traditional spectral methods (e.g., Fourier transform) assume stationarity, which is not suitable for short, transient odor-related responses. The continuous wavelet transform (CWT)-particularly with the complex Morlet mother wavelet-provides superior joint time–frequency resolution and is widely used in cognitive EEG analysis [3, 4].

The primary objective of this study is to investigate whether odor-evoked EEG oscillations can provide clinically useful markers for distinguishing healthy elderly individuals from those with amnestic mild cognitive impairment and Alzheimer's disease. Specifically, we use complex Morlet wavelet analysis to characterize group differences in delta, theta, alpha, and beta band-power dynamics during an olfactory oddball paradigm and extract wavelet-based features from frontal and central electrodes located at left frontal pole (Fp1), frontal midline (Fz), central midline (Cz), and parietal midline (Pz) according to the international 10–20 system. We then evaluate the diagnostic value of these features using a random forest classifier. In addition, we compare a compact frontal model based solely on Fp1 with a broader four-channel model including Fp1, Fz, Cz, and Pz to examine the trade-off between model simplicity and classification performance.

2. Literature review

Wavelet-based EEG analysis has played a central role in neurodegenerative research, particularly because it captures the transient and non-stationary nature of brain oscillations with high temporal precision. In this section, we review studies closely related to the present work, emphasizing papers that integrate EEG analysis with Morlet wavelet time frequency representations and classification models.

Panda et al. proposed an automated epileptic EEG detection approach using wavelet-based feature extraction followed by support vector machine (SVM) classification. Their pipeline decomposed EEG signals using the discrete wavelet transform (DWT), then computed statistical features (notably energy/entropy/standard deviation) as inputs to the classifier. The study reported promising performance for distinguishing abnormal from normal EEG activity, positioning wavelet-derived features as an effective representation for EEG classification tasks [5].

Uyulan and Erguzel conducted an in-depth comparison of various time–frequency feature extraction methods for mental task classification, emphasizing the use of Morlet wavelets. Their work applied Morlet wavelet transforms (MWT) on EEG signals recorded during mental tasks (e.g., arithmetic and imagery tasks) and evaluated the effectiveness of several feature extraction methods, including wavelet-based and other frequency domain approaches. They reported that Morlet wavelet-based features provided the best classification accuracy when combined with support vector machines (SVMs), highlighting the strength of wavelet transforms in capturing time–frequency characteristics of EEG signals. This paper is significant for our study as it underscores the efficacy of Morlet wavelet transforms in improving classification tasks involving mental state analysis [6].

Chattu and Prasad focused on artifact handling using wavelet processing (including Morlet wavelet-based preprocessing) combined with learning based decision making. The study frames the workflow as artifact reduction followed by classification, and reports performance gains after preprocessing, with a decision tree-based classifier achieving an accuracy around 88% (with accompanying precision, recall, and F measure reported), illustrating how wavelet denoising can be paired with machine learning classifiers to improve EEG decision performance [7].

Buriro et al. evaluated an alcoholism-detection pipeline that combines multichannel EEG with wavelet scattering transform (WST) feature extraction and supervised classification. Using a publicly available UCI alcoholism EEG dataset, they extracted WST features from 1.0-s EEG records (20 male subjects with alcoholism vs 20 male healthy subjects) and compared classifiers under two validation schemes. In record-wise 10-fold cross-validation, WST features fed to an SVM achieved perfect discrimination of alcoholic vs control records, while a 1D-CNN achieved similarly strong performance; under independent subject-wise evaluation, the best mean performance was achieved by WST features combined with linear discriminant analysis (LDA). They

also reported that features from occipital and parietal regions carried the strongest discriminative signal. Methodologically, this paper is directly relevant because WST is a CNN-like, fixed wavelet feature extractor that is commonly implemented with Morlet wavelets, linking Morlet-based time–frequency processing to practical EEG classification performance [8].

Al-Qaysi et al. presented a hybrid motor-imagery EEG framework that combines wavelet-based image representations with deep feature learning and conventional classification. They generated scalograms using multiple mother wavelets, used these scalograms for transfer learning with VGG-16 as a feature extractor, and final class assignment is produced by selecting one of six downstream classifiers; the authors report that the combination "VGG 16 features + neural network classifier" (with the Amor mother wavelet) yields the best performance across two motor imagery datasets, demonstrating a practical template for Morlet style scalogram driven classification [9].

Abdubrani et al. targeted driver-fatigue detection and used Morlet time–frequency analysis as the core representation for classification. After preprocessing (including artifact handling) and extracting features from selected EEG channels/bands, they proposed an enhancement of the Morlet mother wavelet to improve time–frequency resolution, then evaluated multiple classifiers—specifically ANN (Levenberg–Marquardt training), k-NN, and SVM—with the ANN achieving the highest reported accuracy among the compared models. This paper is especially relevant because it treats Morlet design/parameterization as a performance lever rather than a fixed preprocessing step [10].

Li et al. addressed the lower signal amplitude and higher noise typical of ear-EEG SSVEP recordings by integrating Morlet wavelet transforms with a lightweight classification approach. They proposed Morlet-ROCKET (Morlet transform + RandOm Convolutional KErnel Transform), and compared it against established alternatives, including FBCCA and a transformer-based method. Their study supports the idea that Morlet-based time–frequency processing can be paired with modern classifiers (including efficient kernel-based feature models) to improve EEG classification performance under challenging acquisition conditions [11].

Parihar and Swami investigated dementia-related EEG discrimination and used wavelet-based time–frequency feature extraction to support multi-group classification (Alzheimer's disease, frontotemporal dementia, and controls). Their reported modeling emphasizes supervised learning with strong performance using random forest, and they describe a subject-aware evaluation strategy (leave-one-subject-out) to better reflect generalization to unseen individuals. This work is closely related to our study context and reinforces the value of wavelet-derived EEG features combined with robust classification and evaluation design [12].

Usgaonkar et al. presented an end-to-end Morlet-based deep learning framework for multi-class EEG state classification (expert meditators, novice meditators, and control). After artifact reduction, signals are decomposed into specific frequency bands (theta, alpha, gamma), Morlet wavelet transforms are used to generate time-frequency images, and a CNN is trained for multi-class classification. Reported accuracies vary by band, with the highest in the theta band, illustrating a common and highly transferable design: Morlet time frequency image formation followed by CNN-based classification [13].

Taken together, these studies establish a strong methodological and conceptual foundation for combining Morlet wavelet analysis with machine-learning classifiers in the search for EEG biomarkers of cognitive impairment. Wavelet-based features capture the transient, non-stationary structure of neural oscillations, while random forests provide a flexible and robust framework for handling small samples, noisy measurements, and non-linear relationships between spectral patterns and diagnostic status.

Nevertheless, relatively few studies have specifically examined odor-evoked EEG responses using wavelet-derived band-power features within a multiclass classification framework that simultaneously distinguishes healthy elderly individuals, amnestic aMCI, and Alzheimer's disease. Existing work has often focused on resting-state EEG, binary contrasts (e.g., AD vs. controls), or univariate group comparisons rather than explicit

predictive models, and many analyses rely on proprietary datasets with limited reproducibility. The present study addresses these gaps by leveraging a publicly available olfactory EEG dataset, extracting subject-level Morlet band-power features at a small set of frontal and central electrodes, and systematically comparing a compact Fp1-only model with a four-channel model within a random forest framework.

3. Dataset

This study uses a publicly available EEG dataset of odor-evoked brain activity in elderly individuals spanning the cognitive continuum from normal aging to amnestic mild cognitive impairment (aMCI) and Alzheimer's disease (AD) [14, 15]. Participants were recruited from the memory clinic of Ziaeian Hospital in Tehran, Iran, and underwent neurological examination together with a standardized neuropsychological assessment prior to inclusion. Individuals with poor EEG recording quality, a history of major neurological events, or evidence of olfactory dysfunction were excluded. The final sample comprised 35 elderly participants: 15 cognitively normal individuals, 7 patients with aMCI, and 13 patients with a clinical diagnosis of AD. The mean age of the cohort was 70.97 ± 8.58 years, and a little over half of the sample were women. For each participant, Mini-Mental State Examination (MMSE) scores and basic demographic information were available and were used for descriptive analyses and exploratory regression.

All participants completed the same olfactory oddball paradigm. In the original dataset, each odor trial was coded in a binary odor variable, where odor = 0 denotes lemon and odor = 1 denotes rose. Each subject underwent 120 odor trials, with a 2-second odor presentation followed by an 8-second no-odor interval (total trial duration 10 seconds). Two odorants were used: lemon, presented as the frequent "standard" stimulus with probability 0.75, and rose, presented as the infrequent "deviant" stimulus with probability 0.25. The same pseudo-random sequence of standards and deviants was applied across all participants, facilitating direct comparison of both trial-wise and averaged responses between diagnostic groups.

EEG was recorded with a multi-channel system using electrodes placed according to the international 10–20 system, including frontal and central scalp locations over regions implicated in olfactory and cognitive processing. Electrode impedances were maintained below 15 kΩ during acquisition to ensure signal quality. Signals were band-pass filtered between 0.5 and 40 Hz and segmented into epochs extending from 1 second before odor onset to 2 seconds after odor onset. Artifacts associated with eye blinks were identified and removed using independent component analysis (ICA), and residual noisy trials were rejected following visual inspection. The resulting dataset, therefore, consists of preprocessed, artifact-reduced odor-evoked EEG epochs to standard and deviant stimuli for each participant, accompanied by demographic variables and MMSE scores, which together form the basis for the statistical modelling and classification analyses reported in this study.

4. Materials and methods

Electrophysiological responses to olfactory stimulation were analyzed using a structured and reproducible workflow designed to capture stimulus-locked oscillatory dynamics and to evaluate their diagnostic value in differentiating healthy aging, amnestic mild cognitive impairment (aMCI), and Alzheimer's disease (AD). This section details the EEG preprocessing, feature extraction, and statistical procedures applied uniformly to all participants.

The dataset includes older adult participants categorized into three diagnostic groups: cognitively normal controls (Normal), aMCI, and AD. EEG was recorded at a sampling rate of 200 Hz using four standard 10–20 electrodes positioned at Fp1, Fz, Cz, and Pz. Each trial captures odor delivery (lemon or rose) within defined pre-stimulus and post-stimulus windows, allowing the study of transient oscillatory changes associated with odor processing. EEG was recorded from four scalp electrodes (Fp1, Fz, Cz, Pz) positioned according to the international 10–20 system. A schematic of this four-electrode montage is shown in Figure 1.

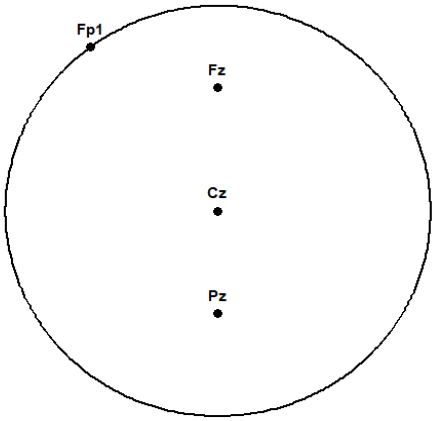


Figure 1. Schematic scalp map of the four electrodes (Fp1, Fz, Cz, Pz)

To ensure high-quality, artifact-free epochs, all trials flagged as noisy in the accompanying metadata were removed prior to analysis, and only clean, stimulus-locked trials were retained. No additional temporal filtering was applied in the present analysis; all filtering and artifact-reduction steps (including independent component analysis, ICA) refer to the dataset's original preprocessing. This conservative choice was motivated by methodological work showing that inappropriately designed high-pass and band-pass filters can substantially distort ERP and time–frequency estimates, particularly around sharp onsets and offsets [16]. This strategy was adopted to preserve the native morphology of odor-evoked responses and to avoid introducing filter-related artifacts. To characterize how oscillatory activity evolves around odor onset, the EEG signals were decomposed into the time–frequency domain using the complex Morlet continuous wavelet transform (CWT). This approach is well-suited to short, transient responses because it provides a controlled compromise between temporal and spectral resolution and allows both slow and fast components to be examined on the same scale.

For a real-valued EEG signal $x(t)$, the continuous wavelet transform at scale $a > 0$ and time b is defined as:

$$W_x(a,b) = \int_{-\infty}^{\infty} x(t)\ \psi^*_{a,b}(t)\, dt$$

where $\psi^*_{a,b}(t)$ denotes the complex conjugate of the scaled and shifted mother wavelet

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\,\psi\left(\frac{t-b}{a}\right)$$

The mother wavelet $\psi(t)$was chosen as a complex Morlet wavelet of the form

$$\psi(t) = \pi^{-1/4} \exp(i\,\omega_0 t) \exp\left(-\frac{t^2}{2}\right)$$

with dimensionless central frequency $\omega_0 = 6$. This parameterisation is widely used in EEG time–frequency analysis because it yields a reasonable balance between temporal precision and frequency resolution over the frequency range of interest. At each time–scale combination (a, b), the complex wavelet coefficient $W_x(a,b)$ was obtained, and the wavelet power was defined as:

$$P_x(a,b) = |\,W_x(a,b)\,|^2$$

Scales were then converted to their corresponding pseudo-frequencies to facilitate interpretation in conventional EEG bands.

From the resulting time–frequency maps, band-limited power was extracted by averaging. $P_x(a,b)$ within four canonical frequency ranges: delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), and beta (12–30 Hz). This procedure produced time-resolved estimates of odor-related changes in slow and fast oscillations for each channel and each trial. In the next step of the analysis, these band-specific trajectories were summarized into

baseline-corrected measures of spectral displacement, which were then used as inputs for the statistical comparisons and classification models [17].

For each channel, frequency band, and trial, wavelet power was averaged within a pre-odor baseline window and a post-odor window aligned to stimulus onset. Let P_{base} denote the mean baseline power and P_{post} the mean post-stimulus power. Odor-evoked modulation was indexed using a log-ratio measure defined as:

$$R = \log_{10}\left(\frac{P_{\text{post}} + \varepsilon}{P_{\text{base}} + \varepsilon}\right)$$

where ε is a small numerical constant added to avoid instabilities when power is close to zero. Positive values of R indicate an increase in band-limited power relative to baseline, whereas negative values correspond to suppression. This proportional change measure is standard in neural time–frequency analysis and helps normalise for between-subject differences in absolute power [18].

Within each participant, log-ratio values were averaged across all clean trials to obtain stable subject-level estimates for each band and channel. If $R_{s,cbk}$ is the log-ratio for the subject, s, channel c, band b, and trial $k = 1, \dots, n_s$, the subject-level feature is defined as:

$$\bar{R}_{s,cb} = \frac{1}{n_s}\sum_{k=1}^{n_s} R_{s,cbk}$$

These subject-level summaries formed the basis of group-level statistics, ANOVA testing, PCA visualization, and machine-learning classification [18].

Diagnostic group differences were examined primarily at the Fp1 electrode, given its proximity to orbitofrontal regions that receive dense olfactory input. For each frequency band and odor condition (odor = 0 denotes lemon and odor = 1 denotes rose), one-way ANOVA was applied to the subject-level log-ratio features. With k groups, total sample size N, and sums of squares SS_{between} and SS_{within}, the F-statistic is defined as

$$F = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{SS_{\text{between}}/(k-1)}{SS_{\text{within}}/(N-k)}$$

which follows an $F_{k-1,\,N-k}$ distribution under the null hypothesis of equal means. Standard significance testing at $\alpha = 0.05$ was used, with attention to effect directions and consistency across bands rather than to p-values alone. Visualization included group means, boxplots, violin plots, and heatmaps to emphasize distributional differences beyond central tendency.

Before interpreting ANOVA results, residual normality and homogeneity of variances were assessed using Shapiro–Wilk tests and visual inspection of Q–Q plots and residual vs fitted plots [18, 19]. Potential outliers were checked using boxplots and standardized z-scores; no extreme cases were removed solely on the basis of statistical outlier status [20].

To characterize the multivariate structure of the spectral features, principal component analysis was applied to the standardized subject-level log-ratios. PCA provides orthogonal linear combinations (principal components) that successively maximize explained variance, and is a standard way to summarize correlated EEG features [21]. The first two components, which captured the majority of variance in the Fp1 feature set, were used to visualize separation among diagnostic groups and to compare true versus predicted labels in the classification analyses. We used a random forest (RF) classifier, an ensemble tree-based method introduced by Breiman [22], which combines predictions from many decision trees built on bootstrap samples of the training data and random subsets of predictors. Each tree votes for a class label, and the RF aggregates these votes by majority rule, yielding a flexible non-parametric classifier that can capture nonlinear interactions while remaining relatively robust to noisy features and overfitting.

For a feature vector x, each tree b = 1,…,B yields a class prediction $\hat{y}^{(b)}(x)$; the forest prediction is the majority vote:

$$\hat{y}(x) = mode\{\hat{y}^{(1)}(x), \dots, \hat{y}^{(B)}(x)\}$$

Two models were trained:

1. An Fp1-only model (4 features: delta, theta, alpha, and beta log-ratios at Fp1).
2. A four-channel model (16 features: the same four bands at Fp1, Fz, Cz, and Pz).

In this study, the RF was trained on the subject-level $\log_{10}$(post/baseline) band-power features derived from odor-evoked EEG at Fp1 (and, in a secondary analysis, from all four electrodes Fp1, Fz, Cz, Pz). The diagnostic label (Normal, aMCI, AD) was used as the response. Performance was evaluated on a held-out test set using standard confusion-matrix-based metrics. For a given class, with true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), the measures were defined as:

$$\text{Sensitivity} = TP / (TP + FN)$$
$$\text{Specificity} = TN / (TN + FP)$$
$$\text{Precision} = TP / (TP + FP)$$
$$F1 = 2 \times \text{Precision} \times \text{Sensitivity} / (\text{Precision} + \text{Sensitivity})$$

Overall accuracy was defined as = (TP + TN) / (TP + FP + TN + FN). These metrics are standard in multi-class classification and have been widely discussed in the methodological literature [23]. However, accuracy and F1 can be misleading on imbalanced datasets [24]. Class-wise and macro-averaged values were computed from the confusion matrices reported in the Results. These quantities were computed separately for each diagnostic class in a one-vs-rest manner, and macro-averaged across the three classes to obtain the summary indices reported in Tables 4 and 6. The overall analysis pipeline, from raw olfactory EEG and MMSE data to statistical analysis and RF classification, is summarized in Figure 2. We first imported stimulus-locked epochs and excluded noisy trials, then applied Morlet time-frequency analysis to derive band-limited power features before aggregating them at the subject level.

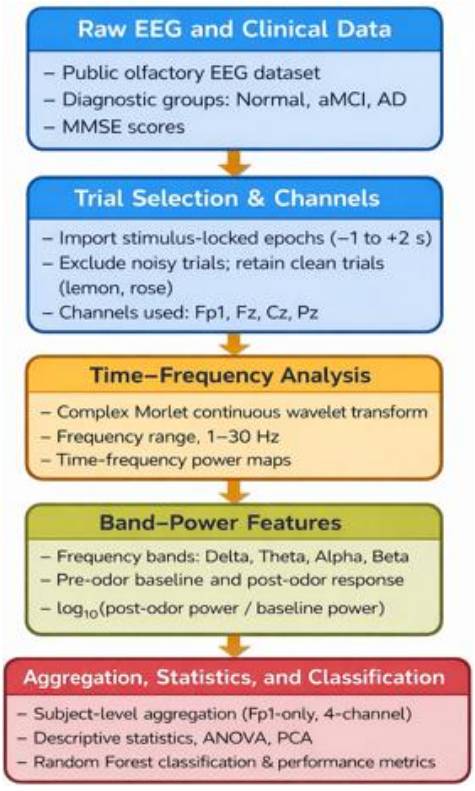


Figure 2. End-to-end analysis pipeline for olfactory EEG and MMSE data

5. Results

This section reports the main empirical findings of the study. We first summarize odor-evoked band-power changes at the frontal pole (Fp1) using descriptive statistics. We then describe the morphology of single-trial responses and their time–frequency characterization, followed by group-wise comparisons of frontal spectral displacement. Finally, we examine the multivariate structure of the spectral features and evaluate their diagnostic performance in the classification models.

The descriptive statistics in Table 1 provide an initial overview of odor-evoked displacement at the frontal pole (Fp1). Across all four bands, median log-power ratios are positive in the three groups, indicating that odor presentation generally increases frontal oscillatory power relative to baseline. However, the magnitude and band profile of this displacement differ systematically by diagnosis. Healthy participants exhibit the largest mean increases in delta and theta, with moderate modulation in alpha and beta. aMCI subjects show consistently smaller displacements across all bands and often lie close to zero, suggesting weakened odor-evoked responses. AD patients fall between Normal and aMCI in the slow bands but display the highest mean values in alpha and beta, consistent with a combination of reduced slow-band enhancement and a broader, less selectively tuned frontal activation. Taken together, these descriptive gradients already hint at an orderly progression in frontal odor-evoked dynamics from Normal through aMCI to AD, which is examined more formally in the subsequent group-comparison analyses.

Table 1. Descriptive statistics of Fp1 log-power ratios (delta, theta, alpha, beta) across groups

Group	Band	Mean	SD	Median	Min	Max
Normal	Delta	0.1221	0.1836	0.0937	-0.0754	0.7526
Normal	Theta	0.0936	0.1633	0.0841	-0.3059	0.5538
Normal	Alpha	0.0833	0.1375	0.0864	-0.2589	0.4036
Normal	Beta	0.0633	0.1224	0.0677	-0.2843	0.3111
AMCI	Delta	0.0398	0.0301	0.0411	-0.0365	0.1017
AMCI	Theta	0.0436	0.0421	0.0340	-0.0142	0.1319
AMCI	Alpha	0.0406	0.0316	0.0371	-0.0919	0.0922
AMCI	Beta	0.0547	0.0362	0.0620	-0.0198	0.1293
AD	Delta	0.0774	0.1265	0.0534	-0.0576	0.4293
AD	Theta	0.0801	0.1195	0.0630	-0.1935	0.3815
AD	Alpha	0.1132	0.1550	0.0703	-0.4474	0.5096
AD	Beta	0.1006	0.1675	0.0735	-0.5707	0.4503

The use of representative single-trial epochs offers a clear and understandable perspective on the manner in which odor-evoked EEG responses evolve throughout the cognitive continuum. The odor-locked waveforms from Fp1, Fz, Cz, and Pz seen in one Normal participant, one aMCI participant, and one AD person are shown in Figure 3. Every trace is time-aligned to the beginning of the odor, and it displays the raw morphology of the scalp response at the single-trial level. This is done before any averaging is done over individual repetitions or participants.

According to the time-domain waveforms shown in Figure 3, there seems to be a consistent decline in the odor-evoked responses across all of the groups. A clearly time-locked deflection can be seen in the Fp1 and midline traces of the Normal individual after the odor has begun to be detected. This deflection has a rather quick beginning and a shape that is consistent across channels. The matching deflections are still apparent in the aMCI individual, but they seem to be less in amplitude and more temporally distributed. This indicates that the person's reaction to the odor is less efficient and less precisely timed. There is a significant loss of organized, stimulus-locked oscillatory activity at the single-trial level in the AD patient, which is consistent with the fact that recognized odor-evoked transients are considerably attenuated and partially masked by continuing background

activity. This complementary time–frequency characterization of the identical epochs is motivated by the discrepancies in waveforms that have been observed.

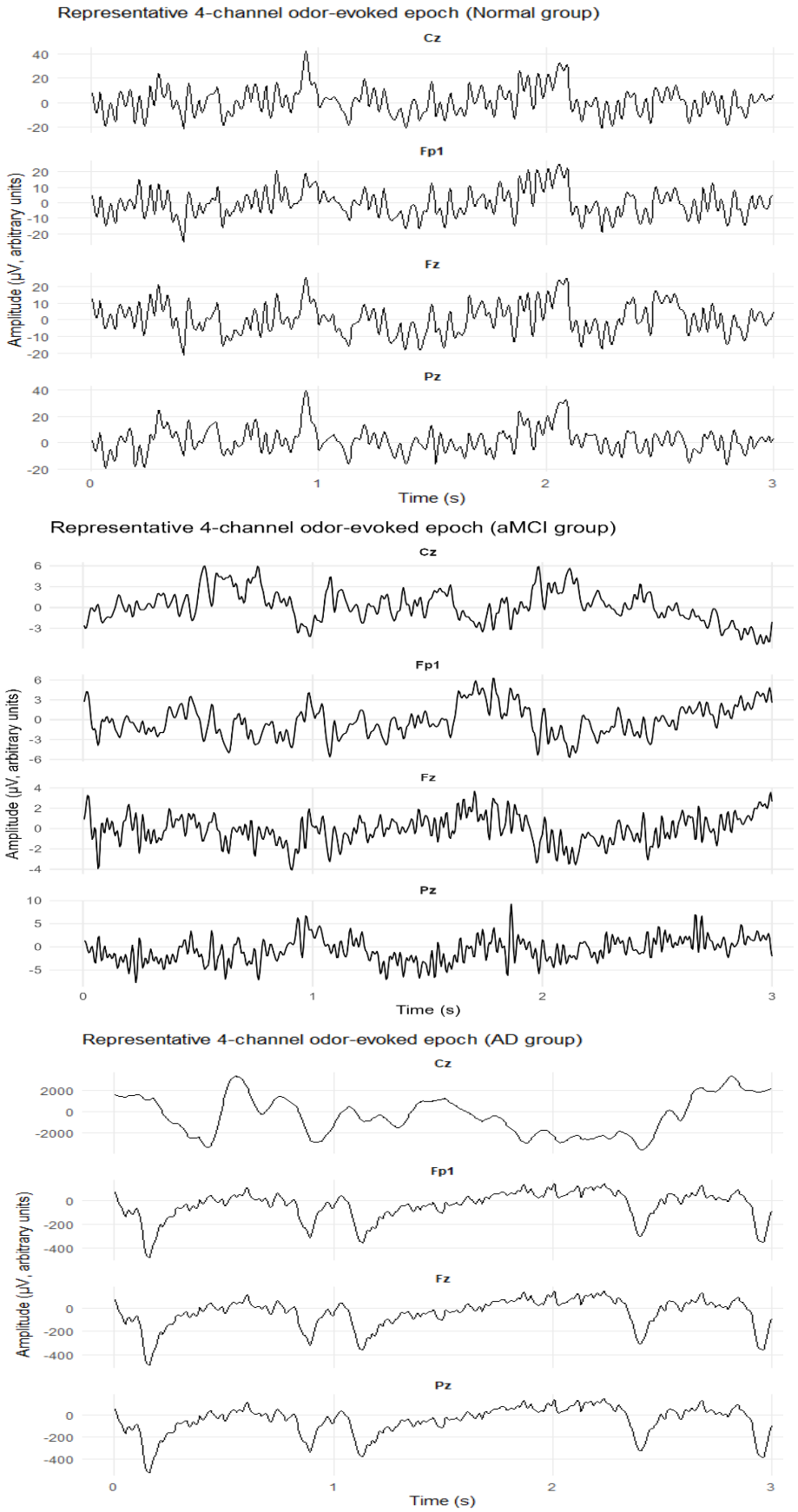
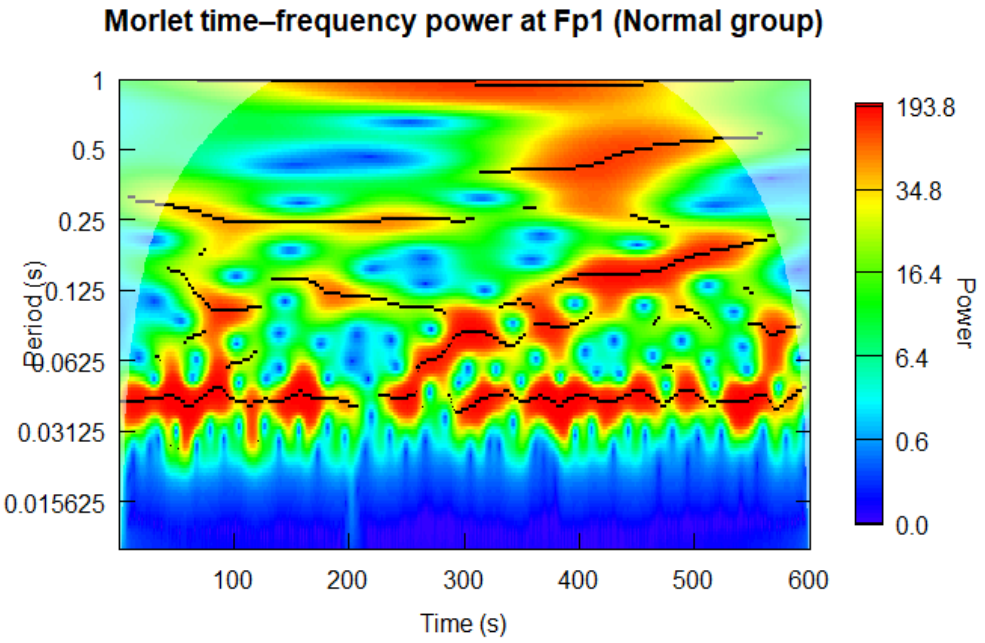
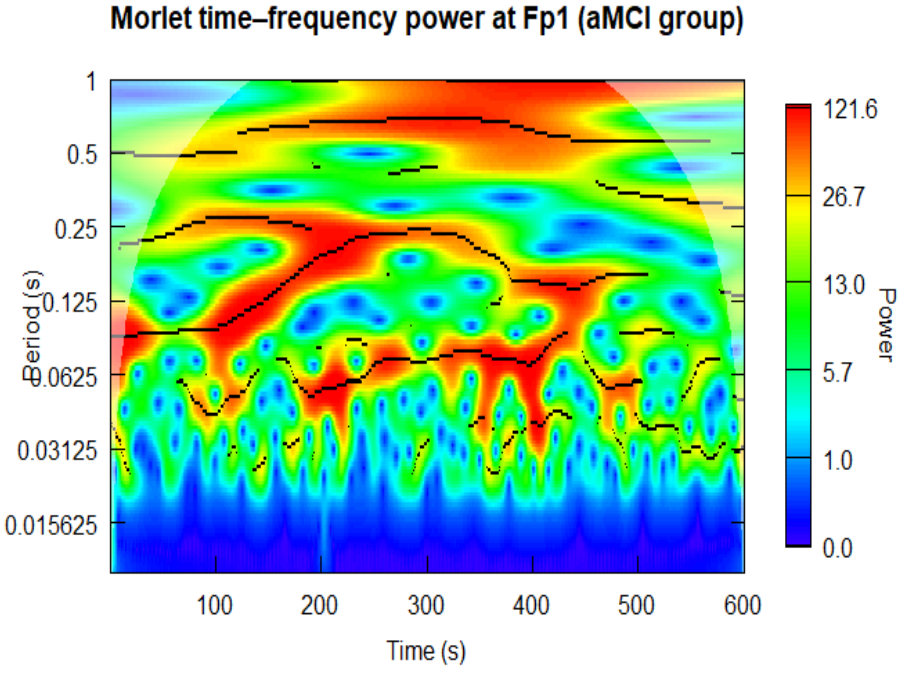


Figure 3. Three panels (Normal, aMCI, AD), each panel shows a 4-channel time-series

To characterize these differences in the time–frequency domain, Figure 4 shows the corresponding Morlet wavelet maps for the same representative epochs. In the Normal subject, the Fp1 map displays a compact, high-amplitude increase in theta–alpha band power that is tightly confined to the expected post-stimulus window, with additional, more modest modulation at higher frequencies. In the representative aMCI epoch, this activation is weaker and more temporally blurred, suggesting a less precisely organized pattern of odor-evoked oscillatory activity. In the representative AD epoch, only a faint, patchy increase in slow-frequency power remains, and higher-frequency enhancement is largely absent.

Taken together, the waveform patterns in Figure 3 and the wavelet profiles in Figure 4 indicate that odor-evoked oscillatory responses are strongest and most stereotyped in healthy elderly participants, intermediate in aMCI, and markedly degraded in AD. This qualitative progression supports the hypothesis that both phase-locked (time-aligned) and induced (non–phase-aligned) components of the olfactory EEG response are progressively disrupted with increasing cognitive impairment and motivates the subsequent quantitative analyses of band-power features across groups.





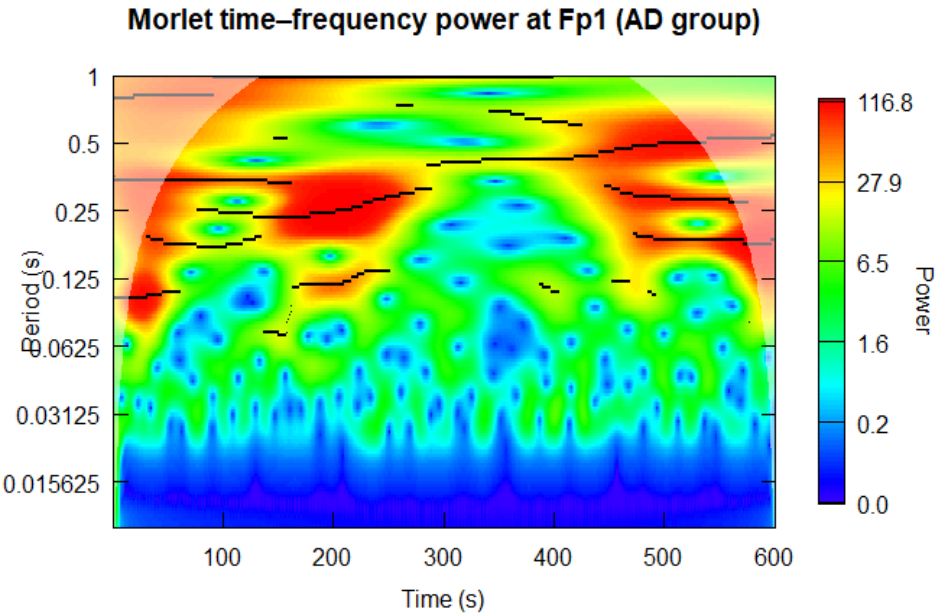


Figure 4. Three panels (Normal, aMCI, AD), each panel shows the Fp1 Morlet scalogram

Because Fp1 was our primary channel of interest, time–frequency maps are shown only for Fp1, even though the underlying epochs included four channels.

Frontal band-power differences across diagnostic groups are summarized from several angles. The heatmap in Figure 5 provides an overview of mean Fp1 band-power displacement (log10(post/baseline)) across groups, frequency bands, and the two odor conditions (lemon and rose). Warmer tiles (yellow- orange) indicate stronger power increases from baseline, whereas darker tiles indicate little or no change. Under the lemon condition (Odor = 0), AD patients show the largest mean enhancement in delta and beta bands, with Normal subjects typically intermediate and aMCI participants closest to baseline. Under the rose condition (Odor = 1), the pattern shifts: Normal subjects exhibit the most consistent slow-band (delta–theta–alpha) enhancements, while aMCI again remains modest and AD responses are comparatively attenuated.

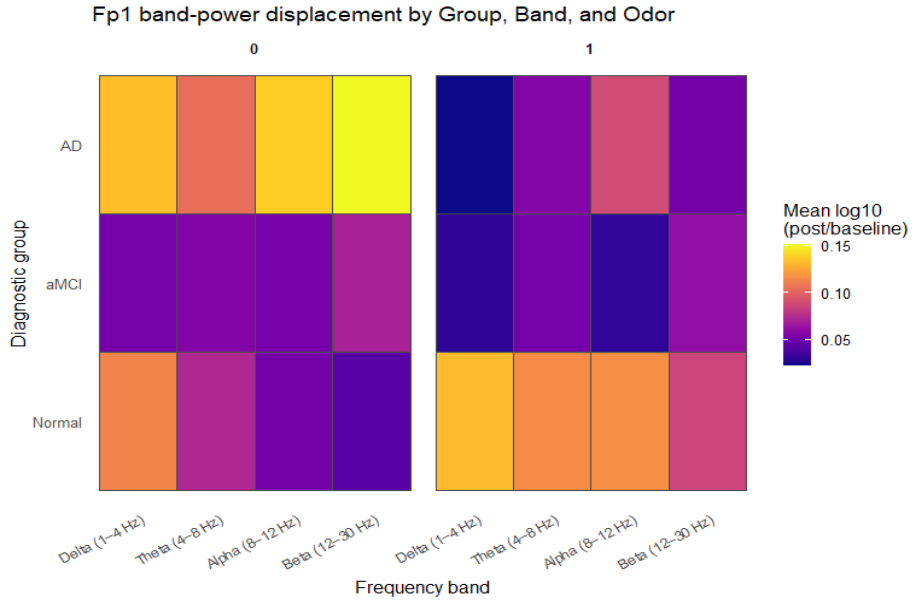


Figure 5. Mean log-power ratios by group and odor

Subject-level boxplots in Figure 6 allow inspection of central tendency and variability around the zero line (no net change from baseline). For lemon trials (odor = 0), median delta, theta, alpha, and beta power are all positive,

with AD usually showing the highest medians and the broadest interquartile ranges, Normal in between, and aMCI closest to zero. For rose trials (odor = 1), group medians move closer together, but Normal participants generally retain slightly higher band-power increases than aMCI and AD, especially in the delta and theta ranges.

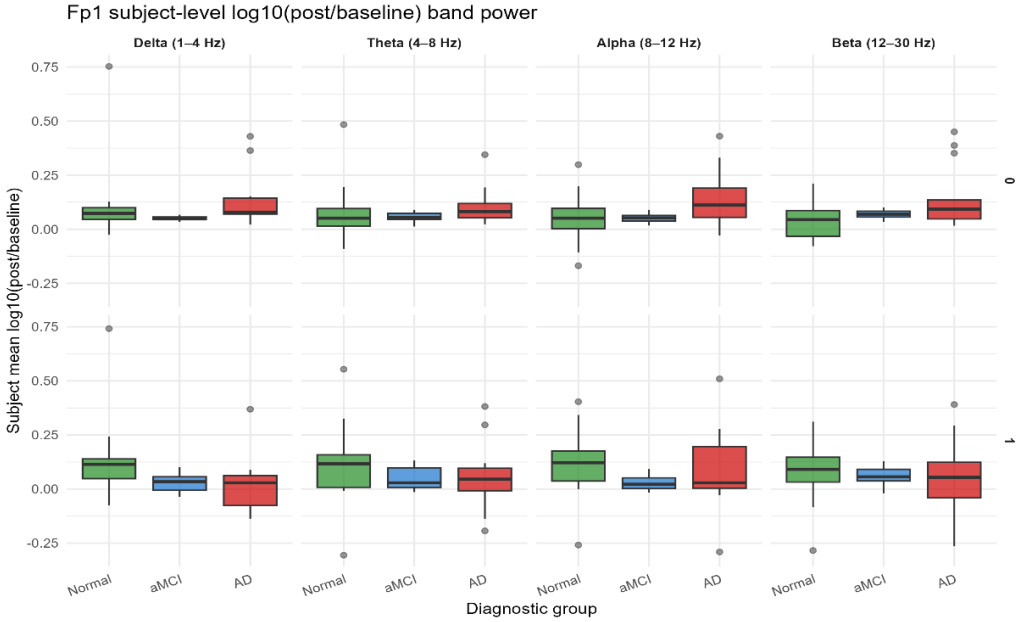


Figure 6. Boxplots of Fp1 band-power across groups

The violin plots in Figure 7 display the full distributions of subject-level band power and highlight both overlap and systematic shifts between groups. For lemon trials (odor = 0), the bulk of the AD distributions is shifted upward relative to Normal and aMCI, indicating stronger frontal power enhancement despite considerable between-subject variability. For rose trials (odor = 1), the Normal distributions are displaced upward across most bands, whereas aMCI and AD are more concentrated around zero, illustrating that group differences coexist with substantial within-group variability.

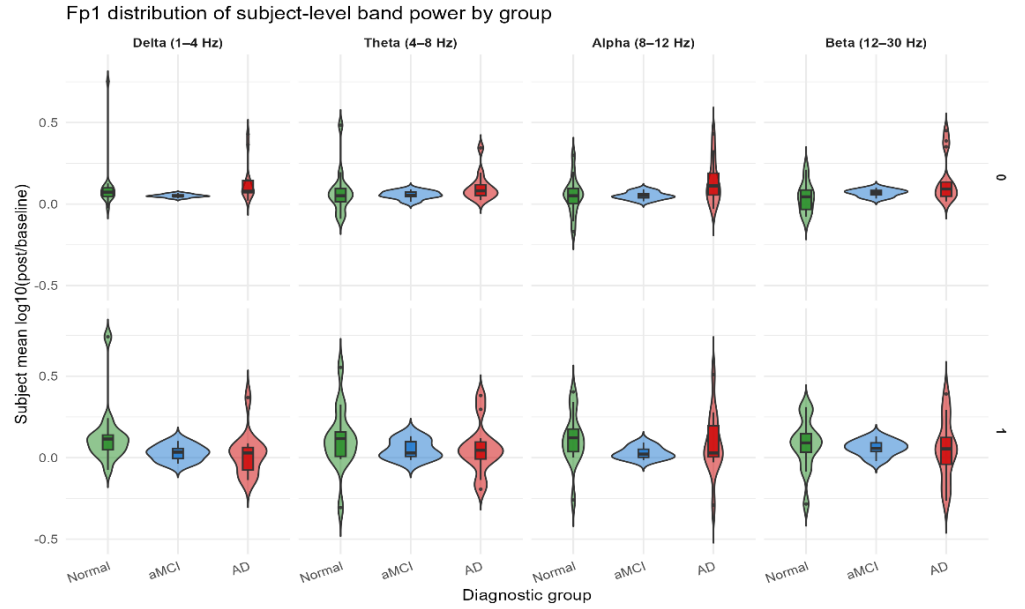


Figure 7. Violin plots of Fp1 band-power across groups

Table 2 formally assess these differences. Beta power in the lemon condition (odor = 0) shows a statistically significant group effect, whereas other bands exhibit non-significant but directionally consistent trends. The

alignment between descriptive gradients and ANOVA patterns supports the view that odor-related frontal dynamics change progressively across the cognitive continuum, even when individual tests do not reach conventional significance thresholds.

For each band-by-odor ANOVA model, we inspected assumption diagnostics. Shapiro–Wilk tests applied to model residuals and visual inspection of Q–Q plots indicated only mild deviations from normality, mainly in the extreme tails, with no pattern suggesting gross misspecification. Plots of residuals versus fitted values showed no strong evidence for systematic heteroscedasticity or non-linear structure. Boxplots and standardized z-scores highlighted a small number of observations with $|z| > 3$, but refitting the ANOVA models after excluding these points did not materially change group means or the pattern of p-values. In line with these diagnostics, we interpret the ANOVA results as reasonably robust summaries of group differences in odor-evoked band-power, while acknowledging that deviations from ideal Gaussian assumptions may slightly attenuate nominal p-value accuracy.

Table 2. ANOVA results for delta, theta, alpha, beta band-power at Fp1

Band	Odor	df1	df2	F	p
Delta (1–4 Hz)	0	2	32	0.78	0.466
Delta (1–4 Hz)	1	2	32	2.27	0.120
Theta (4–8 Hz)	0	2	32	0.56	0.578
Theta (4–8 Hz)	1	2	32	0.61	0.547
Alpha (8–12 Hz)	0	2	32	2.48	0.100
Alpha (8–12 Hz)	1	2	32	0.71	0.501
Beta (12–30 Hz)	0	2	32	3.72	0.035
Beta (12–30 Hz)	1	2	32	0.18	0.835

The PCA map of Fp1 odor-evoked band-power features, shown in Figure 8, does not split the three diagnostic groups into sharply separated clusters, but it does reveal a clear ordering along the first principal component. Many AD cases are displaced toward one side of the axis, consistent with stronger slow-band enhancement and weaker alpha modulation, whereas Normal participants cluster toward the opposite side, reflecting more modest spectral change. aMCI subjects are scattered between these two regions, which visually supports the idea of a gradual transition from normal aging through amnestic aMCI to AD rather than three isolated categories.

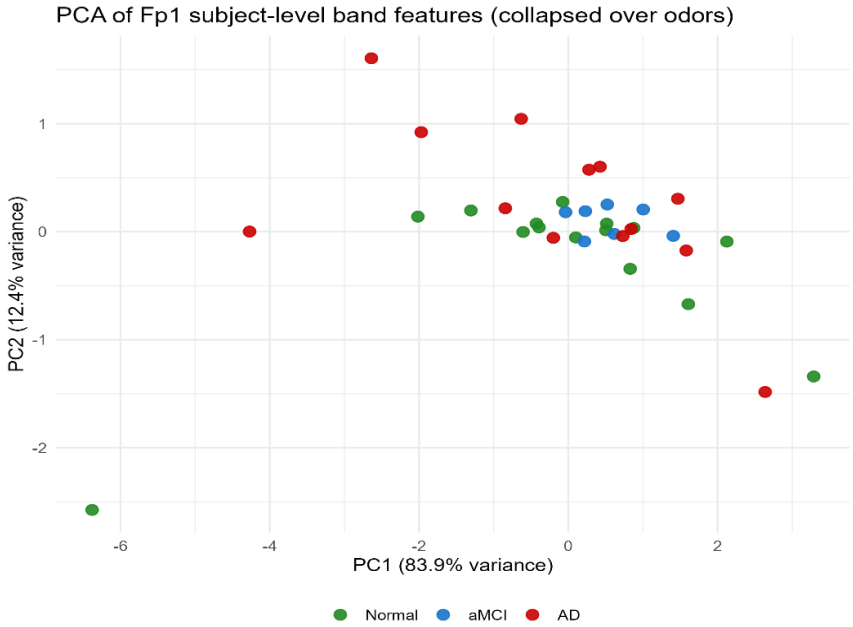


Figure 8. Principal component analysis of Fp1 subject-level odor-evoked band-power features across diagnostic groups

In Figure 9, the same feature space is viewed through the lens of classifier performance. The PCA projection of the held-out test subjects shows that correctly classified individuals tend to lie near the center of their group's region, while misclassified cases appear in the border zones where the groups overlap. This pattern indicates that the behavior of the random forest model is largely constrained by the underlying multivariate geometry of the Fp1 spectral features, and is therefore unlikely to be driven by overfitting or spurious noise.

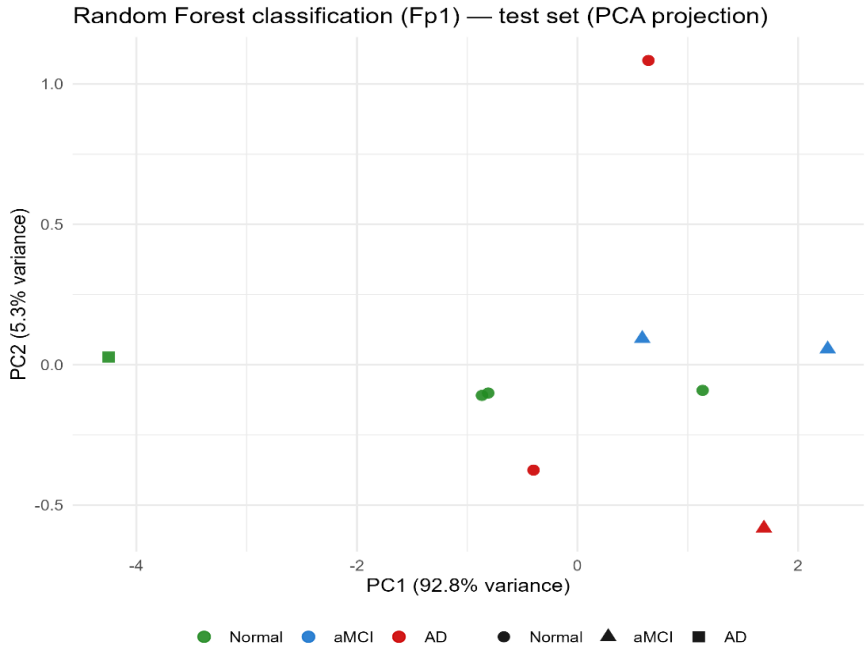


Figure 9. Random forest classification of Fp1 subject-level features on the held-out test set, shown in the PC1–PC2 space

Based on the four band-power features extracted from Fp1, the random forest classifier achieved a test-set accuracy of 66.7%. Class-wise metrics are summarized in Table 3, and the corresponding confusion matrix is reported in Table 4. For Normal subjects, both sensitivity and precision were 0.75, indicating that most Normal cases were correctly identified with relatively few false alarms. All aMCI cases in the test set were correctly classified (sensitivity = 1.00; F1 = 0.80); however, this estimate is based on only two aMCI test subjects and should be interpreted as preliminary and high-variance (unstable). In contrast, detection of AD was weaker (sensitivity ≈ 0.33 and F1 = 0.40), which is consistent with the overlap in frontal spectral profiles between some AD participants and the other groups when only this compact Fp1 feature set is used.

Table 3. Test-set classification metrics for the Fp1-only random forest model

Group	Sensitivity	Specificity	Precision	F1	
Normal	0.75	0.80	0.75	0.75	
aMCI	1.00	0.86	0.67	0.80	
AD	0.33	0.83	0.50	0.40	Overall accuracy (66.7%)

Table 4. Confusion matrix for the Fp1-only random forest model (test set)

	Predicted		
True	Normal	aMCI	AD
Normal	3	0	1
aMCI	0	2	0
AD	1	1	1

Table 5 reports macro-averaged metrics across the three diagnostic categories. The macro-sensitivity of about 0.69, macro-specificity around 0.83, and macro-F1 close to 0.65 indicate a reasonably balanced profile rather than dominance by a single group. Overall, these results suggest that odor-evoked band-power displacement at a single frontal sensor already carries diagnostically informative structure, especially for distinguishing aMCI from both Normal aging and manifest AD.

Table 5. Macro-averaged test-set metrics for the Fp1-only random forest model

	Sensitivity	Specificity	Precision	F1
Macro-average row	0.69	0.83	0.64	0.65

Expanding the feature space from four Fp1 bands to all four electrodes resulted in a more complex but not necessarily more accurate model. The four-channel random forest achieved a test-set accuracy of 55.6% (Table 6), which is lower than the 66.7% obtained with Fp1 alone. Class-wise metrics show that performance is now more uneven: for Normal subjects, the F1 score is 0.50, for aMCI it increases to 0.67, and for AD it reaches 0.57. Overall accuracy decreased, with improved AD detection but reduced Normal performance, yielding a more uneven class-wise profile. The corresponding confusion matrix is reported in Table 7. Given the small test set (n = 9) and the small aMCI count (n = 2), class-wise estimate -especially for aMCI- should be interpreted as high-variance.

Table 6. Test-set classification metrics for the 4-channel random forest model

Group	Sensitivity	Specificity	Precision	F1	
Normal	0.50	0.60	0.50	0.50	
aMCI	0.50	1.00	1.00	0.67	
AD	0.67	0.67	0.50	0.57	Overall accuracy 55.6%

Table 7. Confusion matrix for the four-channel random forest model (test set)

	Predicted		
True	Normal	aMCI	AD
Normal	2	0	2
aMCI	1	1	0
AD	1	0	2

To summarise these results across classes, Table 8 reports macro-averaged indices. The macro-sensitivity of about 0.56, macro-specificity around 0.76, and macro-F1 of approximately 0.58 confirm a drop in discriminative performance compared with the Fp1-only classifier, even though specificity remains relatively high. Taken together, the class-wise and macro-level metrics suggest that adding channels introduces additional variability that is not tightly aligned with the diagnostic labels.

Table 8. Macro-averaged metrics for the four-channel random forest model

	Sensitivity	Specificity	Precision	F1
Macro-average row	0.56	0.76	0.67	0.58

The variable-importance profile in Figure 10 helps to clarify this finding. The highest-ranked predictors are delta power at Pz, followed by alpha and delta power at Fp1 and delta and beta components at Cz and Fz, with additional alpha and theta bands at Pz also contributing. In contrast, many of the remaining channel–band combinations (not shown among the top features) have negligible importance. This pattern indicates that diagnostically useful information is concentrated in a relatively small subset of fronto-central and parietal bands, whereas most other multichannel features mainly add noise or redundancy rather than improving separation between Normal, aMCI, and AD.

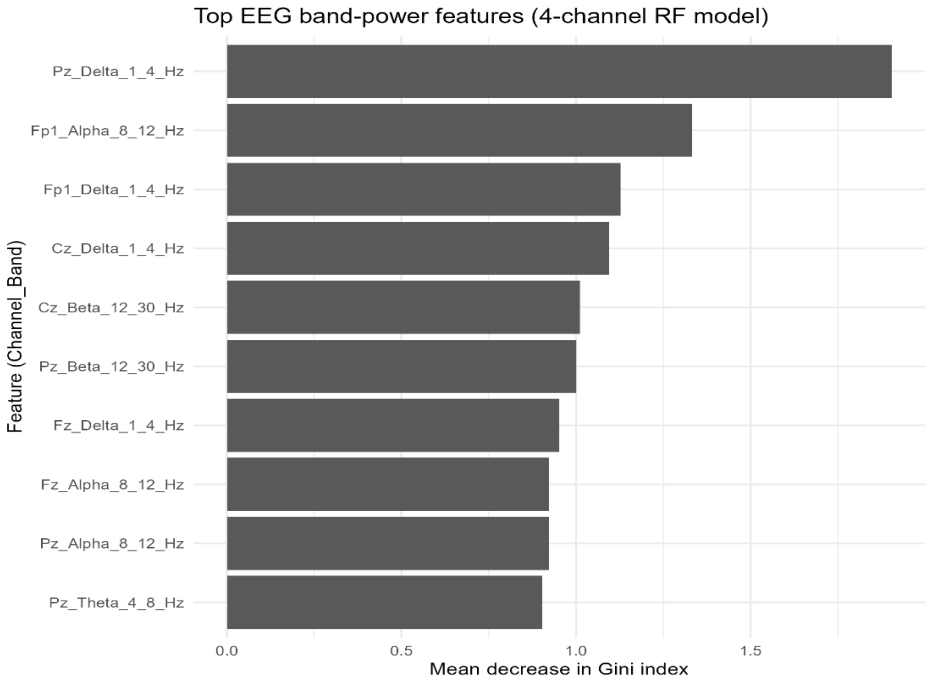


Figure 10. Variable importance of EEG band-power features in the 4-channel random forest model (mean decrease in Gini index)

The geometry of the four-channel feature space, as illustrated in Figure 11, reinforces the modest classification performance of the multichannel model. Each point represents a test subject projected onto the first two principal components, with marker shape indicating the true diagnosis and marker color the class predicted by the random forest. Correct classifications occur where shape and color match, but several subjects display mismatched symbols, reflecting errors. The points from Normal, aMCI, and AD are spread along PC1 with substantial overlap rather than forming compact, well-separated clusters. This diffuse configuration is consistent with the lower accuracy of the four-channel classifier and suggests that, once all channels are included, the additional variability they introduce is only weakly aligned with the diagnostic boundaries.

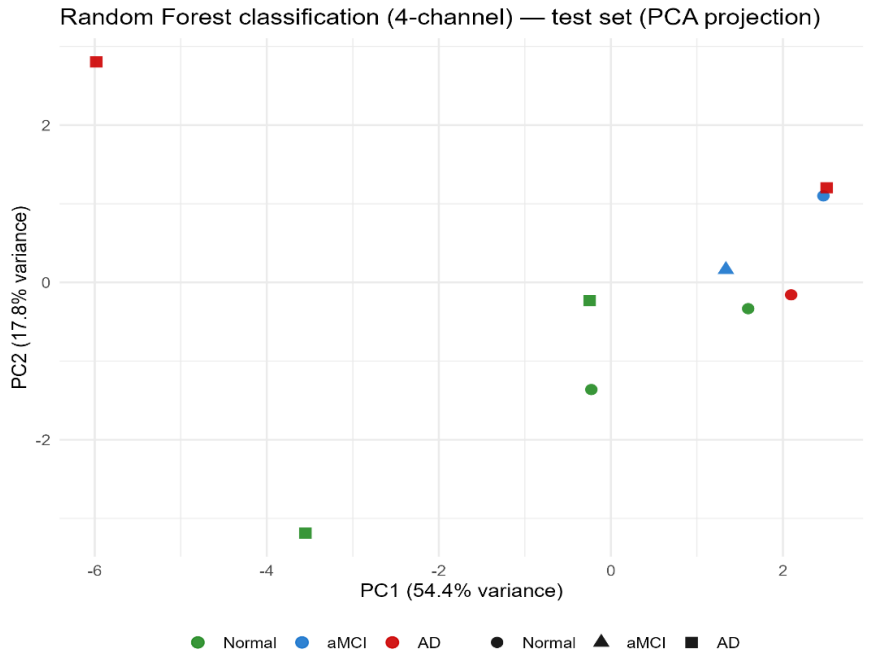


Figure 11. Random forest classification using four-channel EEG band-power features on the held-out test set (PC1–PC2 projection)

6. Discussion

This study set out to build and evaluate a fully documented statistical pipeline for analyzing odor-evoked EEG responses across the cognitive spectrum from Normal aging to amnestic MCI and Alzheimer's disease. Starting from raw data, we implemented a reproducible sequence of preprocessing, single-trial wavelet analysis, subject-level feature aggregation, and group-wise inference, and then used those features in a straightforward random forest classifier. The combined results support two main messages: first, that odor-evoked frontal oscillatory activity differs systematically across diagnostic groups; and second, that even compact band-power summaries of these responses carry diagnostically relevant information, although not at a level that would justify stand-alone clinical use.

The single-trial time-domain and time–frequency figures provide an intuitive entry point to these findings. Representative four-channel epochs show that Normal participants tend to exhibit coherent odor-locked deflections that are preserved across Fp1, Fz, Cz, and Pz, whereas aMCI and AD traces look progressively more fragmented and less clearly time-locked to the stimulus. The corresponding Morlet scalograms at Fp1 illustrate the same pattern in the time–frequency domain: organized, band-limited power modulation during odor presentation in Normal subjects, with more diffuse and attenuated responses in aMCI, and the least structured activity in AD. These visual patterns are qualitative by design, but they align with the idea that olfactory stimulation elicits reproducible frontal oscillatory responses that degrade as neurodegeneration advances.

The subject-level band-power summaries formalize this impression. By baselining each trial and then averaging log10(post/baseline) power across trials per subject, channel, band, and odor, we obtain interpretable measures of how much each frequency band is displaced by the odor relative to its own pre-stimulus baseline. Group summaries and boxplots at Fp1 reveal graded shifts rather than all-or-none differences: in several band–odor combinations, Normal subjects cluster at one end of the distribution, AD subjects at the other, with aMCI occupying an intermediate position. This pattern is consistent with a continuum view of cognitive impairment rather than three isolated groups. Importantly, the analysis respects the within-subject structure by aggregating trials before comparing groups, which keeps the inferential unit at the subject level, as is appropriate for clinical interpretation.

The multivariate perspective offered by PCA adds another layer of insight. When Fp1 band-power features are collapsed over odors and projected into two principal components, participants tend to separate into partially distinct clusters corresponding to Normal, aMCI, and AD. The separation is not perfect, but it is visually apparent and stronger for the Fp1-only feature set than for the four-channel model. This suggests that, in this dataset, odor-evoked frontal activity at Fp1 captures a relatively coherent summary of the group structure, and that adding a small number of additional midline and parietal channels does not automatically improve discriminability. From a statistical modelling perspective, this illustrates a common theme in high-dimensional data: more features do not always translate into better separation, especially when the added variables introduce noise or weakly informative variation.

The Random Forest experiments quantify how much of this structure can be harnessed for prediction. Even with a minimalist feature set based solely on Fp1, the model achieved a test-set accuracy of about two-thirds, clearly above the one-third chance level expected under three equally likely classes. Macro-averaged F1-scores were in the moderate range, indicating that performance was reasonably balanced across groups rather than being driven by a single "easy" class. Extending the feature space to the four-channel model produced classification performance of similar order and slightly weaker group separation in PCA, reinforcing the impression that the most informative signal in this paradigm is concentrated at Fp1. For a statistical audience, the key point is not that random forest is uniquely optimal, but that standard tree-based methods can extract clinically relevant signal from a small, carefully engineered feature set derived from a transparent pipeline.

The exploratory regression linking Fp1 alpha displacement to MMSE demonstrates how the same features can be used to study continuous clinical outcomes rather than discrete diagnoses. While the present sample is not

large enough to make strong claims about effect sizes, the model serves as a template for how wavelet-derived EEG summaries can be regressed on cognitive scores while adjusting for age and gender. This dual-use classification and regression highlights the flexibility of the proposed pipeline for different inferential goals.

Taken together, the results show that odor-evoked EEG band-power features can be handled in a way that is both statistically principled and practically implementable. The study does not try to push accuracy to its absolute maximum; instead, it focuses on documenting each step from raw data to figures and tables, making it clear what was done, why it was done, and how the resulting features behave under standard multivariate and machine-learning tools.

7. Limitations

Several limitations need to be acknowledged before any translational conclusions are considered. First, the dataset is relatively modest in size and comes from a single center, with group membership defined by existing clinical diagnoses and MMSE scores. This design is appropriate for a methodological demonstration but limits generalizability. Sampling variability and local clinical practice may influence both the cognitive labels and the EEG characteristics, and we did not include an external validation cohort.

Second, the EEG montage is intentionally sparse, restricted to four electrodes (Fp1, Fz, Cz, Pz). This choice simplifies the analysis, reduces the dimensionality of the feature space, and makes the pipeline easier to reproduce in settings where dense arrays are not available. However, it also means that potentially informative temporal, parietal, or olfactory-related sources are not captured. More complex spatial patterns, phase-based measures, or connectivity indices are beyond the scope of the current work and could change the conclusions about which regions or frequency bands are most informative.

Third, the sample sizes for the three diagnostic groups are modest (N = 15 Normal, 7 aMCI, 13 AD), which limits statistical power and increases uncertainty in group comparisons. Because multiple ANOVAs were conducted across frequency bands and odor conditions on this modest sample, we emphasized consistent qualitative patterns and effect sizes rather than strict reliance on individual p-values. The Random Forest classifier was evaluated using repeated cross-validation with a fixed resampling seed and a simple train/test split; these choices reduce variance but do not eliminate optimism in performance estimates. Together, the small sample and single site mean that the reported classification metrics should be viewed as preliminary and require validation in independent, larger cohorts.

Fourth, the random forest models rely on a single random split into training and test sets, supplemented by repeated cross-validation within the training data. Although this is a standard and defensible strategy, it does not exhaust all possible sources of variability in model performance. With a larger dataset, one could consider nested cross-validation, more extensive hyper-parameter tuning, or comparisons with alternative classifiers (for example, penalized multinomial regression, gradient boosting, or linear discriminant analysis). In the current sample, reporting a single held-out test set provides a realistic but somewhat noisy estimate of predictive performance.

Fifth, the study does not explicitly model or control for several potential confounders, such as medication use, vascular risk factors, anosmia severity, or comorbid neurological conditions. These factors may influence both odor perception and EEG responses. The regression of Fp1 alpha on MMSE adjusts for age and gender only, which is a minimal adjustment set. As a result, the observed associations between EEG features and cognitive status should be interpreted as correlational rather than causal.

Finally, the analysis is tailored to one specific experimental paradigm (short odor bursts with two odorants, lemon and rose, separated by a no-odor interval) and one specific preprocessing strategy (stimulus-locked epochs, −1 to +2 s, with noisy trials excluded). Different stimulation protocols, longer recording windows, or alternative artefact-handling approaches could yield different time–frequency dynamics and potentially

different classification behavior. The present work, therefore, establishes a proof of concept rather than a universal template.

8. Conclusions and future directions

This work demonstrates a complete, reproducible pipeline for analyzing odor-evoked EEG responses in the context of cognitive impairment, from raw data to wavelet-based features, group-wise comparisons, and supervised classification. Using only four scalp electrodes and log-ratio band-power features derived from Morlet wavelets, we showed that odor-evoked frontal oscillations differ systematically across Normal, aMCI, and AD groups and can support above-chance discrimination between these diagnoses. The combination of subject-level summaries, visualization (single-trial epochs, time–frequency maps, heatmaps, PCA), and Random Forest models provides a coherent statistical narrative that is accessible to both clinical and methodological audiences.

At the same time, the results underline that classification performance remains moderate and that substantial work is still needed before EEG-based olfactory paradigms can be considered for diagnostic decision-making. In our view, the main contribution of this study lies less in the specific numerical values of accuracy or F1, and more in the transparent integration of wavelet time–frequency analysis with standard statistical tools and machine-learning methods. The pipeline can be extended in several directions, including richer EEG montages, alternative feature sets, more flexible classifiers, and more detailed modelling of clinical covariates.

In summary, odor-evoked EEG band-power features contain meaningful information about cognitive status, and a carefully designed statistical workflow can reveal this information in a way that is both interpretable and reproducible. The present findings provide a methodological basis for future studies that integrate rigorous statistical approaches with clinically relevant questions, supporting larger cohorts and longitudinal monitoring of cognitive impairment. These findings suggest that olfactory-evoked frontal EEG oscillations, summarized through a transparent wavelet–Random Forest pipeline, carry diagnostically relevant information and could motivate larger, prospective studies aimed at developing affordable, non-invasive tools for tracking neurodegenerative processes.

Future research should focus on validating these findings in independent cohorts, exploring richer odor paradigms and additional EEG features (including connectivity and phase-based measures), and embedding the resulting models into longitudinal designs that track cognitive trajectories over time. Such developments may eventually contribute to affordable, non-invasive tools for early detection and monitoring of neurodegenerative processes in routine clinical settings.

Declaration of competing interest

The authors declare that they have no known financial or non-financial competing interests in any material discussed in this paper.

Funding information

The author declares that they have received no funding from any financial organization to conduct this research.

References

[1] J. Jeong, “EEG dynamics in patients with Alzheimer’s disease,” *Clin. Neurophysiol.*, vol. 115, no. 7, pp. 1490–1505, 2004. https://doi.org/10.1016/j.clinph.2004.01.001

[2] T. Kovács, “Mechanisms of olfactory dysfunction in aging and neurodegenerative disorders,” *Ageing Res. Rev.*, vol. 3, no. 2, pp. 215–232, 2004. https://doi.org/10.1016/j.arr.2003.10.003

[3] C. Tallon-Baudry and O. Bertrand, “Oscillatory gamma activity in humans and its role in object representation,” *Trends Cogn. Sci.*, vol. 3, no. 4, pp. 151–162, 1999. https://doi.org/10.1016/S1364-6613(99)01299-1

[4] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bull. Am. Meteorol. Soc.*, vol. 79, no. 1, pp. 61–78, 1998. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

[5] R. Panda, P. S. Khobragade, P. D. Jambhule, S. N. Jengthe, P. R. Pal, and T. K. Gandhi, "Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure detection," in *Proc. 2010 Int. Conf. Syst. Med. Biol. (ICSMB)*, IIT Kharagpur, India, 2010. https://doi.org/10.1109/ICSMB.2010.5735413

[6] C. Uyulan and T. T. Erguzel, "Analysis of time–frequency EEG feature extraction methods for mental task classification," in *Proc. 2017 Int. Conf. Intell. Syst. Comput. Vis. (ISCV)*, 2017, pp. 168–174.

[7] P. Chattu and C. V. P. R. Prasad, "Removing artifacts in EEG data based on wavelets and neural networks," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 22, pp. 5470–5478, 2021.

[8] A. B. Buriro *et al.*, "Classification of alcoholic EEG signals using wavelet scattering transform-based features," *Comput. Biol. Med.*, vol. 139, p. 104969, Dec. 2021. https://doi.org/10.1016/j.compbiomed.2021.104969

[9] Z. T. Al-Qaysi, A. Al-Saegh, A. F. Hussein, and M. A. Ahmed, "Wavelet-based hybrid learning framework for motor imagery classification," *Iraqi J. Electr. Electron. Eng.*, vol. 19, no. 1, pp. 47–56, 2023.

[10] R. Abdubrani, M. Mustafa, and Z. L. Zahari, "Enhancement of Morlet mother wavelet in time–frequency domain in electroencephalogram (EEG) signals for driver fatigue classification," in *Advances in Intelligent Manufacturing and Mechatronics*, Lecture Notes in Electrical Engineering, vol. 988, pp. 151–161, 2023. https://doi.org/10.1007/978-981-19-8703-8_13

[11] X. Li *et al.*, "The classification of SSVEP-BCI based on ear-EEG via random convolutional kernel transform with Morlet wavelet," *Discov. Appl. Sci.*, vol. 6, p. 149, 2024. https://doi.org/10.1007/s42452-024-05816-2

[12] A. Parihar and P. D. Swami, "Analysis of EEG signals with the use of wavelet transform for accurate classification of Alzheimer's disease, frontotemporal dementia and healthy subjects using machine learning models," *Fusion: Pract. Appl.*, vol. 14, no. 2, pp. 43–55, 2024. https://doi.org/10.54216/FPA.140203

[13] S. Usgaonkar, D. R. Edla, and R. R. Reddy, "A meditation-based brain state classification framework: an integrated Morlet wavelet transforms and CNN approach with EEG signals," *Multimed. Tools Appl.*, 2025. https://doi.org/10.1007/s11042-025-21065-w

[14] M. J. Sedghizadeh *et al.*, "Brain electrophysiological recording during olfactory stimulation in mild cognitive impairment and Alzheimer's disease patients: an EEG dataset," *Data Brief*, vol. 47, p. 109289, 2023. https://doi.org/10.1016/j.dib.2023.109289

[15] M. J. Sedghizadeh, H. Aghajan, and Z. Vahabi, "Brain electrophysiological recording during olfactory stimulation in mild cognitive impairment and Alzheimer's disease patients: an EEG dataset," *Mendeley Data*, V1, 2023. https://doi.org/10.17632/sgzbgwjfkr.1

[16] A. Widmann, E. Schröger, and B. Maess, "Digital filter design for electrophysiological data—a practical approach," *J. Neurosci. Methods*, vol. 250, pp. 34–46, Jul. 2015. https://doi.org/10.1016/j.jneumeth.2014.08.002

[17] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bull. Am. Meteorol. Soc.*, vol. 79, no. 1, pp. 61–78, 1998. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

[18] M. X. Cohen, *Analyzing Neural Time Series Data: Theory and Practice*. Cambridge, MA: MIT Press, 2014.

[19] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3–4, pp. 591–611, 1965. https://doi.org/10.1093/biomet/52.3-4.591

[20] P. E. Greenwood, "ANOVA model diagnostics including QQ-plots," in *Intermediate Statistics with R (LibreTexts)*, 2022.

[21] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philos. Trans. A Math. Phys. Eng. Sci.*, vol. 374, no. 2065, p. 20150202, 2016. https://doi.org/10.1098/rsta.2015.0202

[22] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001. https://doi.org/10.1023/A:1010933404324

[23] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009. https://doi.org/10.1016/j.ipm.2009.03.002

[24] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, 2020. https://doi.org/10.1186/s12864-019-6413-7