Non-linear and time-domain sleep qEEG features predict CSF protein damage markers in early Alzheimer's disease

Anna Michela Gaeta[a,1], Lorena Gallego Viñarás[b,c,1], Ferran Barbé[d,e], Pablo M. Olmos[f,c], Reinald Pamplona[g], Farida Dakterzada[h], Arrate Muñoz-Barrutia[b,c,*], Gerard Piñol-Ripoll[i,j]

[a]*Servicio de Neumología, Hospital Universitario Severo Ochoa, Leganés, Spain*
[b]*Departamento de Bioingeniería, Universidad Carlos III de Madrid, Leganés, Spain*
[c]*Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain*
[d]*Group of Translational Research in Respiratory Medicine, Hospital Universitari Arnau de Vilanova and Santa Maria, IRBLleida, , Lleida, , , Spain*
[e]*Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Madrid, Spain*
[f]*Departamento de Teoría de la Señal y las Comunicaciones, Universidad Carlos III de Madrid, Leganés, Spain*
[g]*Department of Experimental Medicine, University of Lleida-Biomedical Research Institute of Lleida (UdL-IRBLleida), Lleida, Spain*
[h]*Unitat Trastorns Cognitius, Clinical Neuroscience Research, IRBLleida-Santa Maria Lleida University Hospital, Lleida, Spain*
[i]*Institut de Recerca Biomèdica de Lleida (IRBLleida), Lleida, Spain*
[j]*Alzheimer's Disease and Other Cognitive Disorders Unit, Neurology Service, Hospital Clínic de Barcelona, Fundació de Recerca Clínic - Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, Barcelona, Spain*

Abstract

Introduction:

In preclinical Alzheimer's disease (AD), oxidative stress induces non-enzymatic

*Corresponding author
Address: Escuela Politécnica Superior, Avda. de la Universidad, 30, 28911 Leganés (Madrid) España
Telephone: +34 91 624 82 25
E-mail: mamunozb@ing.uc3m.es

[1]Anna Michela Gaeta and Lorena Gallego Viñarás contributed equally to this work.

protein damage—detected as cerebrospinal fluid (CSF) biomarkers—and disrupts sleep-related networks, altering sleep electroencephalographic patterns. Due to the invasiveness of CSF sampling, quantitative electroencephalography (qEEG) is proposed as a non-invasive alternative for predicting oxidatively modified protein levels via Machine Learning (ML).

Methods: Forty-two mild-to-moderate AD patients underwent polysomnography (PSG). qEEG features were extracted. CSF protein oxidation markers levels —glutamic semialdehyde, aminoadipic semialdehyde, Nε-carboxyethyl-lysine, Nε-carboxymethyl-lysine, and Nε-malondialdehyde-lysine —were assessed by gas chromatography/mass spectrometry, and ML models trained to predict CSF biomarker levels.

Results: qEEG features from slow-wave sleep (SWS) and rapid eye movement (REM) sleep, particularly over frontal and central regions, yielded $R^2 > 0.9$ and RMSE < 0.1 for biomarker prediction.

Conclusion: qEEG is a non-invasive, scalable tool for detecting AD-related oxidative stress, with potential implications for early diagnosis and risk stratification.

Keywords: Alzheimer's disease, sleep qEEG, CSF protein oxidation-derived markers, oxidative stress, entropy, skewness, maximum value, variance, ensemble methods

---

1. Introduction

In preclinical Alzheimer's disease (AD), pathological changes advance silently before cognitive symptoms emerge, creating a neurotoxic environment that disrupts neuronal network integrity [1, 2, 3, 4, 5, 6, 7]. Identifying biomarkers that capture early disease alterations is crucial as timely intervention may delay or prevent progression[8, 9].

Oxidative stress—stemming from an imbalance between pro- and antioxidants—is a major contributor to early mechanisms preceding overt AD pathology [3, 4, 5, 6, 7]. It triggers excess reactive oxygen species (ROS) production that damages key macromolecules, disrupting cellular homeostasis and fostering amyloid-β (Aβ) production and tau hyperphosphorylation[6][6]. Due to ROS instability, their effects are measured through stable oxidative modifications [10].

Proteins are particularly vulnerable, resulting in irreversible structural alterations and functional impairment [11]. Non-enzymatic oxidative stress biomarkers—including protein carbonyls, advanced glycation end-products (AGEs), and advanced lipoxidation end-products (ALEs)—indicate irreversible protein damage [12, 10].

Specific non-enzymatic oxidatively modified proteins — including aminoadipic semialdehyde (AASA), glutamic semialdehyde (GSA), Nε-malondialdehyde-lysine (MDAL), Nε-carboxymethyl- lysine (CML) and Nε-carboxyethyl-lysine (CEL) — are significantly elevated in AD brains compared to aged controls. These biomarkers reflect increased oxidative damage and disrupted protein homeostasis in AD, which contribute to neurodegeneration and impaired functional net-

works [13, 14].

By promoting AD neuropathology, oxidative stress affects brain regions and networks involved in sleep-wake regulation, [15, 16, 17], leading to decreased neuronal firing thresholds and synaptic dysfunction [6, 18]. These pathological changes manifest as altered oscillatory patterns in sleep electroencephalogram (EEG), particularly during slow wave sleep (SWS) and rapid eye movement (REM) sleep, which are crucial for synaptic homeostasis and cognition [17, 16, 19, 15].

Although oxidative stress biomarkers are not yet validated for clinical diagnosis, they provide valuable insight into early molecular damage and may serve as complementary indicators of biological vulnerability in individuals at risk for AD [20]. However, their widespread use is hindered by the invasive nature of cerebrospinal fluid (CSF) sampling and the limited reliability of oxidized protein markers in peripheral biofluids [21, 22, 23].

In this context, sleep EEG emerges as a promising, non-invasive, and cost-effective tool for assessing early oxidative stress-related processes. By capturing neurophysiological alterations associated with oxidative damage, quantitative EEG (qEEG) overcomes the limitations of brain tissue sampling [13]and the impracticality of routine CSF analysis while maintaining high specificity for early AD-related changes [20, 23]. Integrating qEEG with computational approaches, including Machine Learning (ML), could offer a powerful strategy for early detection, risk stratification, and monitoring of AD progression.

qEEG enables automated, objective analysis of brain connectivity and net-

work dynamics. Traditional sleep qEEG analyses, based on spectral power and sleep stage scoring, provide a limited perspective on neurophysiological changes [24]. In contrast, non-linear dynamics and time-domain qEEG measures offer a more comprehensive characterization of brain network complexity and functional organization[24, 25, 26]. These advanced metrics exhibit greater sensitivity, capturing subtle preclinical neural alterations that conventional methods often overlook[25].

We, therefore, hypothesize that specific non-linear and time-domain sleep EEG features serve as quantitative indicators of neural activity, indirectly reflecting non-enzymatic oxidative stress-related biochemical processes and network vulnerability, even in the absence of cognitive impairment or neuropathology.

ML techniques, known for their ability to detect complex patterns and automate decision-making, are well-suited for EEG analysis. By handling high-dimensional data and capturing intricate neural signatures, ML holds significant potential for improving the diagnostic accuracy and efficiency of qEEG-based biomarkers in AD [27, 28].

Despite the growing use of qEEG and ML in AD research [29, 30, 31, 32], these approaches have not yet been extensively applied to predicting novel AD-related CSF biomarkers, particularly those associated with non-enzymatic protein oxidative damage. Addressing this gap could provide critical insights into early oxidative stress-related neurodegeneration.

This study seeks to overcome the limitations of invasive and costly CSF biomarker acquisition by exploring the potential of noninvasive EEG-based methods for early

oxidative risk stratification, as illustrated in Figure 1. By identifying electrophysiological correlates of oxidative stress, we aim to enable more personalized, sleep-based prevention strategies for individuals at risk of AD.

Accordingly, this study evaluates the predictive power of various quantitative sleep EEG features for main AD-related CSF protein oxidation markers—including GSA, AASA, CEL, CML, and MDAL—leveraging ML-based modelling to enhance biomarker prediction accuracy and reliability.

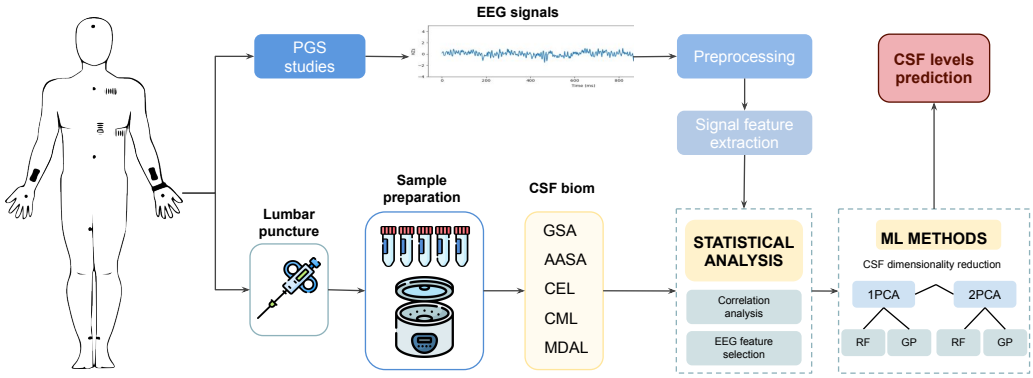


Figure 1: Workflow diagram illustrating the study process. EEG signals were extracted from PSG recordings and analyzed to extract temporal and nonlinear features. Simultaneously, CSF biomarker levels were measured from patient samples. Statistical analysis was performed to examine associations between EEG features and biomarker levels. Machine learning models, including Principal Component Analysis (PCA), Random Forest (RF), and Gaussian Process (GP), were then applied to predict protein levels based on EEG-derived features.

2. Materials and Methods

2.1. Sample and study design

This is an ancillary study from NCT02814045, a prospective trial designed to evaluate the evolution of AD patients with and without Obstructive Sleep Apnea

(OSA) after one year of follow-up. The study was conducted in line with the Declaration of Helsinki and was approved by the human ethics care ethics committee of the Hospital Arnau de Villanova (CE-1218). General recruitment and screening procedures are previously described [33]. In short, eligible patients were recruited from the Cognitive Disorders Unit at the Hospital Universitari Santa Maria (Lleida, Spain), between 2014 and 2018.

The subject involved in the analysis met the following criteria: (1) drug naïve; (2) aged over 60 years; (3) diagnosed with AD according to the NIA-AA criteria [34]; (4) diagnosed with mild-moderate cognitive impairment (an early stage of memory or other cognitive ability loss) using the Mini-Mental State Examination (MMSE$\geq$20) [35]; and 5) having signed informed consent along with the responsible caregiver (and/or legal representative if different from the responsible caregiver). Exclusion criteria for this study were: (1) history of sleep disorders; (2) comorbidities as cancer, severe depression, severe renal or hepatic insufficiency, severe cardiac or respiratory failure; (3) history of mental disorders according to DSM-V-TR™ criteria; (4) history of untreated (or treated by less than 3 months before the screening visit) vitamin B12 or folate deficiency; (5) history of untreated thyroid disease; (6) current use of medication under investigation or beta-blockers, antidepressants, neuroleptics, hypnotics use for less than 15 days before the conduction of polysomnography (PSG); (7) Magnetic Resonance Image (MRI) evidence of hydrocephalus, stroke, space-occupying lesion, or any clinical relevant central nervous system disease besides AD; (8) excessive alcohol intake (>280 gr/week); and (9) visual and/or communication disturbances

affecting the compliance with the study procedures.

Eligible patients undergoing overnight PSG, blood, and CSF samples were collected to determine the Apolipoprotein E genotypes (APOE) and the levels of the biomarkers, respectively. Sociodemographic and clinical variables were collected, including age, sex, ethnicity, anthropometric data, clinical history, and sleep-related symptoms.

2.2. Sample collection

Fasting blood and CSF samples were obtained between 8:00 and 10:00 a.m. to minimize circadian-related fluctuations. CSF was collected in polypropylene tubes and centrifuged at 2000×g for 10 minutes at 4°C to remove cells and other insoluble components. Blood samples were drawn into EDTA tubes, and plasma was separated by centrifuging the blood at 1500 rpm for 20 minutes. The buffy coat was isolated for DNA extraction and subsequent APOE genotyping. All samples were stored at -80°C until further analysis. The samples were provided with the assistance of the IRBLleida Biobank (B.0000682) and PLATAFORMA BIOBANCOS PT17/0015/0027 and were subsequently processed for biomarker analysis.

2.3. APOE genotyping

APOE genotype was performed with a Maxwell® RCS blood DNA kit (Promega, USA) from 2 mL of a peripheral blood sample, using 20 microliters of DNA. Participants were dichotomized according to APOE genotype as a homozygous or heterozygous carrier (APOEε4+) or not carrier (APOEε4-) [36].

2.4. AD core biomarker measurement

Commercial kits (Innotest® β-Amyloid1-42; Innotest® hTAU Ag; and Innotest® Phospho-TAU181P, Fujirebio-Europe, Gent, Belgium) were used to evaluate the main AD biomarkers (Aβ42, t-tau, and p-tau). Analyses were performed by board-certified laboratory technicians masked to clinical data in one round of experiments using one batch of reagents. Intra-assay coefficients of variation were lower than 10%–13% for internal quality control samples (two per plate). The amounts of product were expressed as pg/ml. Concentrations of Aβ42 below 600 pg/ml, t-tau exceeding 425 pg/ml, and p-tau above 65 pg/ml were classified as positive or indicative of abnormality [37].

2.5. Oxidative damage markers measurement

A total of five biomarkers of non-enzymatic protein modifications indicative of oxidative stress were detected and quantified, as previously described [13, 37]. These included markers of protein oxidation— glutamic semialdehyde (GSA) and aminoadipic semialdehyde (AASA); glycoxidation—Nε-carboxyethyl- lysine (CEL) and Nε-carboxymethyl- lysine (CML); and lipoxidation—Nε-malondialdehyde-lysine (MDAL). All were identified as trifluoroacetic acid methyl ester (TFAME) derivatives in acid- hydrolyzed, delipidated, and reduced CSF protein samples using gas chromatography-mass spectrometry (GC/MS).

Briefly, protein samples (500 μg) were delipidated twice by adding methanol: chloroform (2:1, v/v). Following centrifugation, the methanol phase containing proteins was precipitated with cold trichloroacetic acid (TCA) and reduced with

500 mM sodium borohydride in 0.2 M borate buffer (pH 9.2) after overnight incubation at room temperature. The samples were then reprecipitated with TCA, and a mixture of deuterated internal standards (IS) for each marker was added, including [2H8]Lys (d8-Lys), [2H5]5-hydroxy-2-aminovaleric acid (d5-HAVA) as a stable derivative for GSA, [2H4]6-hydroxy-2-aminocaproic acid (d4-HACA) as a stable derivative for AASA, [2H4]CEL (d4-CEL), [2H4]CML (d4-CML), and [2H8]MDAL (d8-MDAL). Proteins were then hydrolyzed with hydrochloric acid (HCl) at 155°C and dried under vacuum.

TFAME derivatives were prepared by adding methanolic-HCl followed by trifluoroacetic anhydride incubation. The derivatized protein hydrolysates were resuspended in dichloromethane and transferred to autosampler vials for SIM-GC/MS analysis. Reference materials from NIST and quality control samples were included every five samples. Four microliters were injected in splitless mode into an Intuvo 9000 GC system equipped with an HP-5MS UI capillary column and coupled to a 5977 MSD (Agilent Technologies, Barcelona, Spain). The injector temperature was maintained at 250ºC.

The oven temperature program was as follows: 110°C for 5 min, ramped to 150°C and held for 2 min, then to 240°C for 5 min, followed by 300°C for 15 min, and held at 300°C for an additional 5 min. A 3-minute post-run at 110°C completed the cycle, yielding a total runtime of 52 min. Data acquisition was performed in selected ion monitoring (SIM) mode across an m/z range of 30–650. The MSD source temperature was set at 250°C, and the transfer line temperature at 310°C. Helium served as the carrier gas. Detected ion pairs were: Lys and d8-

Lys, m/z 180 and 187, respectively; HAVA (GSA) and d5–HAVA (d5-GSA), m/z 280 and 285, respectively; HACA (AASA) and d4–HACA (d4-AASA), m/z 294 and 298, respectively; CEL and d4-CEL, m/z 379 and 383, respectively; CML and d4-CML, m/z 392 and 396, respectively; and MDAL and d8-MDAL, m/z 474 and 482, respectively. The concentration of each marker was expressed as micromoles of GSA, AASA, CEL, CML, or MDAL per mole of lysine.

2.6. Polysomnography

Polysomnography (PSG) was performed overnight using the Philips Respironics Alice 6 LDx system. Each recording included two bilateral EOG signals, six EEG channels from the central (C3-A2, C4-A1), frontal (F3-A2, F4-A1), and occipital (O1-A2, O2-A1) regions, as well as chin EMG, respiratory effort from chest and abdomen, airflow, pulse oximetry, ECG, body position, snoring, and leg movements.

Manual scoring and diagnosis were conducted by expert physicians in accordance with American Academy of Sleep Medicine (AASM) guidelines, which included annotating artefact segments due to movements or poor sensor contact that hinder physiological signal interpretation [38].

Sleep quality measures included total sleep time (TST), sleep efficiency (SE), and arousal index (AI). TST was calculated by subtracting sleep latency and wakefulness after sleep onset from total recording time. SE was the percentage of TST relative to time in bed (TIB), defined as the period from lights-off to lights-on. AI indicated the number of arousals per hour of sleep [38].

Additional PSG measures included time spent awake (W), which is defined as the duration and percentage of TIB spent awake. Latency to stage 1 (N1) of non-rapid eye movement (NREM) sleep was measured from lights off until the first N1 episode. Latency to REM sleep was similarly defined. N1, N2, and N3 sleep durations were quantified in minutes and as percentages of TST, with N3 also recognized as SWS. REM sleep duration was similarly measured [38].

For respiratory parameters, apnea was defined as a reduction in breathing amplitude to below 10% of baseline for over 10 seconds, while hypopnea involved a decrease in airflow by at least 30% for more than 10 seconds, leading to arousal or oxygen desaturation (a decrease in oxygen saturation greater than 3%). The apnea-hypopnea index (AHI) was calculated as the number of apnea and hypopnea events per hour of sleep. Cumulative sleep time percentage with oxygen saturation (SpO2) $< 90\%$ was defined as CT90, and average SpO2 was the mean oxygen saturation during the study. Obstructive sleep apnea (OSA) was defined as an AHI ≥ 15 events/hour [38].

2.7. Signal preprocessing

A signal preprocessing phase is needed to homogenize the overall signal information retrieved from the different EEG recordings and mitigate noise effects during the analysis phase, as previously performed [39]. To preprocess and analyze signal data offline, the MATLAB® signal processing toolbox was employed.

The initial dataset comprised overall recorded PSG signals with annotations provided by clinical experts, following AASM guidelines as described above [38].

As EEG has been considered the PSG signal of interest, the EEG signals from the six channels were used: both-sided central (C3-A2, C4-A1), frontal (F3-A2; F4-A1), and occipital (O1-A2; O2-A1) signals.

EEG signals were monopolar since all of them resulted from the difference between the signal recorded in a specific brain region (C-central, F-frontal, O-occipital) and the reference electrode (A-earlobe or M-mastoid). All of them were digitally acquired with an initial sampling rate (fs_0) ranging from 200 to 500 Hz.

The preprocessing phase was broken down into four primary steps:

- **Resampling**: The EEG signals were resampled for consistency across channels and subjects, addressing variations in sampling rates to an output rate (fs_f) of 500 Hz following the AASM guidelines [38].

- **Filtering**: Bandpass filtering was applied to all signals, with cutoff frequencies set according to AASM guidelines [38]. This ensured consistent spectral content within 0.3 – 35 Hz while preserving essential EEG frequencies. Specifically, frequency cuts removed the DC signal (0 Hz) while retaining Delta (0.5 - 4 Hz), Theta (4 – 8 Hz), Alpha (8 – 13 Hz), and Beta (13 – 32 Hz) bands. A fifth-order digital Butterworth filter was used for this process, chosen for its "maximally flat" response in the passband, which minimizes ripple effects despite a slower transition band [40].

- **Segmentation**: Signals were segmented based on sleep stages annotated by clinicians from hypnograms, resulting in four separate databases representing each sleep stage: N1, N2, N3, and REM.

- **Artifact removal**: Segments containing artifacts were removed based on annotations provided by experts for each session and their respective timestamps.

2.8. Signal features extraction

EEG signals were analyzed to extract seven features from time-domain and nonlinear features subsets, employing various signal-processing techniques and mathematical models [41, 24, 42, 43, 44].

For each signal segment annotated as a different sleep stage, the parameters were assessed as the mean of two-minute windows with a 50% overlap.

The final set of features was computed as the mean of the segments of each sleep stage (N1, N2, N3, and REM) for the whole overnight recording and through all EEG channels.

- **Time-domain features** provide insights into signal characteristics such as amplitude, variability, and distribution over time, helping to identify physiological dynamics and anomalies. Variance (indicating data distribution around the mean) and skewness measured signal dispersion and asymmetry, while kurtosis and maximum value represented peakedness and amplitude extremes [24, 42, 44, 43].

- **Non-linear parameters** analysis examines complex interactions within electrophysiological signals, highlighting their chaotic behavior and non- stationarity [44, 24]. Sample Entropy quantifies complexity based on sequence similarity probability, where lower values indicate regularity and higher

values reflect complexity. Lempel Ziv complexity measures data intricacy through compression-based entropy, with higher values denoting increased complexity [44, 42, 41].

This approach aimed to extract informative characteristics that describe signal dynamics and correlations [41, 24]. Appendix A provides detailed definitions and corresponding mathematical formulas for the time-domain and non-linear features extracted from EEG signals.

3. Statistical analysis

Descriptive statistics were obtained for various variables of relevance, defining the demographic and clinical characteristics of the patient group. Depending on the nature of the variable under examination, mean, median, standard deviation, or frequencies are estimated. The listwise deletion method was used to cope with missing data.

The associations between CSF biomarkers (GSA, AASA, CEL, CML, MDAL) and overall EEG features of both-sided central, frontal, and occipital signals and during N1, N2, N3, and REM sleep stages have been assessed using Spearman's rank test. We also calculated adjusted correlations controlling for age, gender, and ApoE4 status. All tests were two-tailed, and p-values<0.05 were considered statistically significant. Initially, we investigated the interrelationships among CSF biomarkers (GSA, AASA, CEL, CML, MDAL). Correlation matrices were utilized for this analysis.

The study uncovered a substantial correlation among the targeted CSF biomarkers for prediction, indicating that predictions made independently would likely lack meaningful coherence. Hence, a combined approach to predict all CSF biomarkers is deemed appropriate. Given the constraints of the limited database, this insight is expected to significantly enhance the predictions' accuracy.

Following that, we conducted EEG feature selection, retaining those features that showed a significant correlation (both unadjusted and adjusted) with at least one CSF biomarker. Descriptive statistics and Sperman's test were carried out using the R statistical software (version 3.3.1).

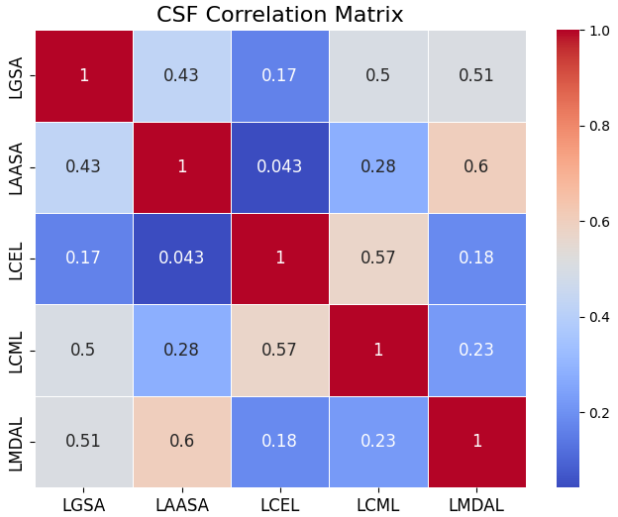


Figure 2: Heatmap and correlation matrix of cerebrospinal fluid (CSF) oxidative stress biomarkers. The matrix illustrates the pairwise Pearson correlation coefficients among glutamic semialdehyde (GSA), aminoadipic semialdehyde (AASA), Nϵ-carboxyethyl-lysine (CEL), Nϵ-carboxymethyl-lysine (CML), and Nε-malondialdehyde-lysine (MDAL). Warmer colours indicate stronger positive correlations, while cooler colours represent weaker or negative correlations. The observed intercorrelations suggest shared pathological mechanisms and justify the integration of these biomarkers into the principal component analysis (PCA) to reduce dimensionality and enhance predictive modeling accuracy in Alzheimer's disease (AD) research.

4. Machine Learning methods for the prediction of CSF oxidative damage biomarkers

Subsequently, various methods were explored to predict CSF levels from the retained EEG features. These methods were performed using Python (version 3.10.12). The libraries employed were Scikit-Learn for the Random Forest implementation and GPy in the Gaussian Process model.

4.1. Dimensionality reduction

To address the interdependencies among CSF biomarkers (Figure 2), we integrated them using Principal Component Analysis (PCA), a statistical technique used to reduce the dimensionality of data while retaining as much variability as possible [45]. For our analysis, we extracted either one principal component (1PC), capturing 54.89% of the variance, or two principal components (2PCs), explaining 78.46% of the variance, with each CSF biomarker contributing with distinct weights.

Similarly, EEG data preprocessing involved handling missing values using Multiple Imputation by Chained Equations (MICE), which imputes values iteratively based on inter-variable relationships [46, 47]. EEG features were also combined through PCA to reduce dimensionality and noise.

Using this biomarker and EEG components, models were trained for prediction, and predicted components were transformed back to their original scales via inverse matrix operations, yielding final values for each CSF biomarker.

4.2. Machine Learning approaches

To address this regression problem, we utilized two machine learning techniques: Random Forest (RF) and Gaussian Process (GP), each applied to the previously pre-processed database. The RF model, a supervised ensemble learning method, combines multiple decision trees trained independently to create a robust predictive framework [48]. It was employed to predict the one or two principal components (1PC and 2PCs) derived from the CSF biomarker data, using the EEG extracted features. This approach was implemented both with EEG features transformed through PCA and with the original EEG features without dimensionality reduction, enabling comparison of the model's performance across different input representations. On the other hand, the GP models a probability distribution over possible functions that fit a given set of data points, incorporating uncertainty into predictions through a covariance (kernel) function [49]. This method was also used to predict both 1PC and 2PCs, enabling direct performance comparisons with RF and providing an additional layer of interpretability due to its ability to measure uncertainty. In this model, the selected kernel was the Radial Basis Function (RBF) with Automatic Relevance Determination (ARD), enabling the assessment of each EEG feature's importance during prediction.

5. Results

5.1. Characteristics of the cohorts

We initially recruited a consecutive series of 144 individuals with a recent diagnosis of mild to moderate AD. Out of these, 128 patients completed a full PSG

Data	number (%) or mean [SD]
Sociodemographic data	
Women	21 (50%)
Age, years	74.71 [69.32; 80.11]
BMI, kg/m^2	27.96 [22.61; 33.30]
Comorbidities	
Hypertension	24 (57.14%)
Diabetes mellitus	7 (16.67%)
Dyslipidemia	21 (50%)
Heart diseases	5 (11.90%)
Stroke	2 (4.76%)
Cognition	
MMSE	23.31 [20.96; 25.66]
Sleep-related symptoms	
ESS	5.98 [2.27; 9.68]
Snoring	1.36 [0.15; 2.56]
Not restorative sleep	0.17 [-0.32; 0.66]
Witnessed apneas	0.19 [-0.32; 0.70]
Insomnia	0.33 [-0.24; 0.91]
Polysomnographic parameters	
TIB, minutes	412.98 [379.02; 446.94]
TST, minutes	262.28 [170.2; 354.36]
Sleep efficiency, %	63.67 [41.53; 85.8]
N1 sleep stage, %	13.27 [5.40; 21.13]
N2 sleep stage, %	25.96 [10.63; 41.29]
N3 sleep stage, %	16.30 [4.32; 28.28]
REM sleep, %	8.05 [1.05; 15.05]
Latency to N1, minutes	36.95 [-12.45; 86.35]
Latency to REM sleep, minutes	156.34 [58.04; 254.65]
AHI events/hour	21.90 [5.30; 38.49]
AI events/hour	38.91 [19.38; 58.44]
Average SpO_2, %	92.74 [90.39; 95.09]
T90%	7.95 [-5.56; 21.46]
ODI events/hour	28.89 [5.96; 51.83]
AD core biomarkers	
Aβ42, pg/ml	510.64 [376.79; 644.50]
t-Tau, pg/ml	550.55 [281.51; 819.59]
p-Tau, pg/ml	81.27 [48.57; 113.96]
Oxidative damage biomarkers	
GSA, μmol/mol of lysine	6556.15 [3884.08; 9228.23]
AASA, μmol/mol of lysine	88.58 [50.61; 126.54]
CEL, μmol/mol of lysine	116.07 [62.44; 169.71]
CML, μmol/mol of lysine	1126.60 [813.04; 1440.17]
MDAL, μmol/mol of lysine	210.68 [-21.64; 443.00]

Table 1: Characteristics of patients with mild to moderate Alzheimer's disease (AD). Key: AHI, apnea-hypopnea index; AI, arousal index; Aβ42, amyloid-beta protein; AASA, aminoadipic semialdehyde; CEL Nε-carboxyethyl-lysine; CML, Nε-Carboxymethyl-lysine; BMI, body mass index; ESS: Epworth sleepiness scale; IQR, interquartile range; p-tau, phospho-tau protein; SD, standard deviation of mean; t-tau, total-tau protein. GSA, glutamic semialdehyde; MDAL, Nε-malondialdehyde-lysine; MMSE, Mini-Mental State Examination; TIB, time in bed; t-tau, total-tau protein; TST, total sleep time; T90, cumulative sleep time percentage with oxyhemoglobin saturation (SpO2) $< 90\%$.

[33]. However, 70 cases were removed because their PSG recordings were no longer available, while several had missing data for some of the CSF biomarkers, resulting in a final sample of 42 patients. Table 1 lists the sample characteristics, which include socio-demographic information, comorbidities, cognitive assessments, CFS biomarker levels, APOE genotyping, self-reported drowsiness, and a summary of extracted polysomnographic parameters.

5.2. EEG features

In Supplementary Material, Table I indicates the average values of the 168 EEG signal features for each EEG channel according to sleep phases. In addition, appendix B offers a list of the overall EEG features computed (Table B.3).

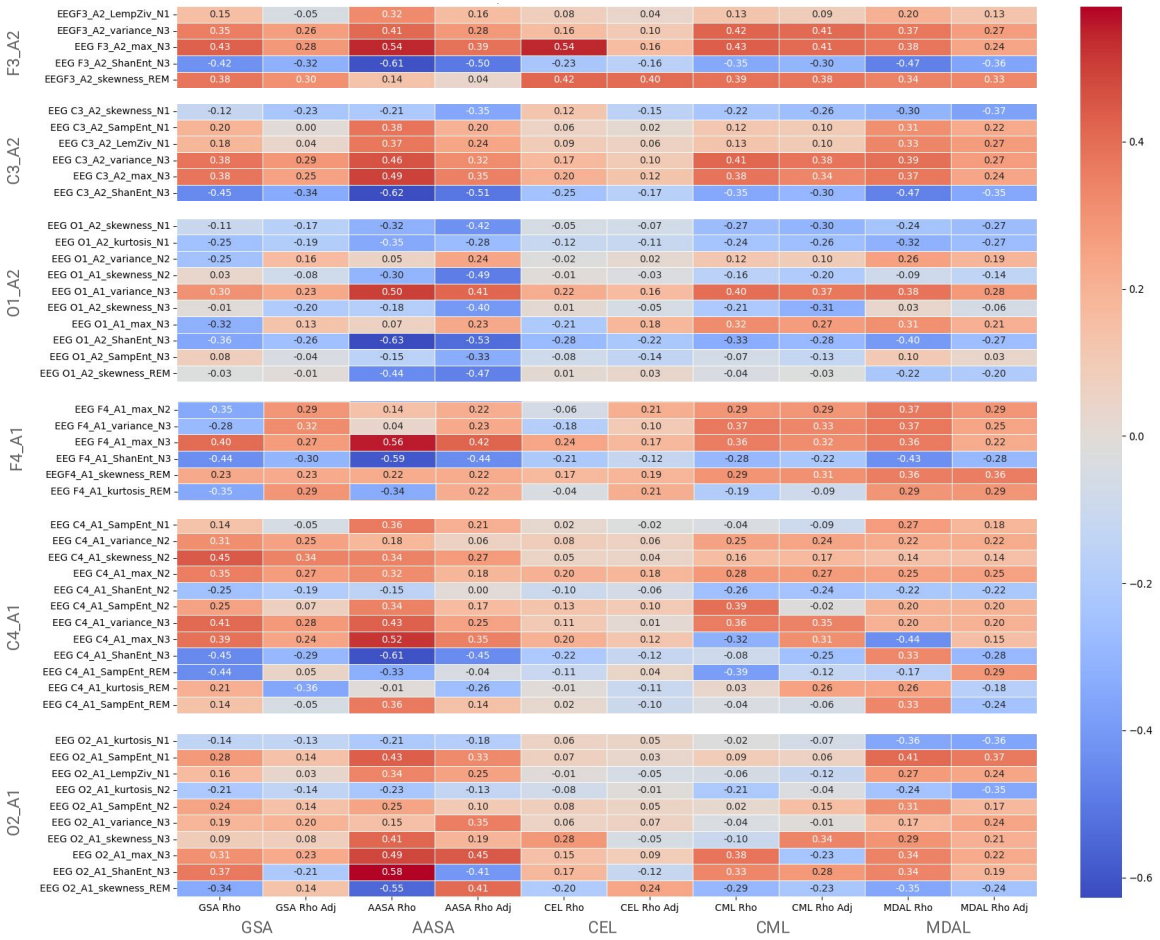


Figure 3: Significant correlations between sleep EEG features and cerebrospinal fluid biomarkers according to sleep stages. The represented values are unadjusted (Rho) or adjusted for age, sex, and ApoE4 status (Rho adjusted). Warmer colours indicate stronger positive correlations, while cooler colours represent weaker or negative correlations. TKey: AASA, aminoadipic semialdehyde; CEL, Nε-carboxyethyl-lysine; CML, Nε-carboxymethyl-lysine; GSA, glutamic semialdehyde; MDAL, Nε-malondialdehyde-lysine; C3_A2, C3-A2 EEG channel; C4_A1, C4-A1 EEG channel; EEG, electroencephalogram; F3_A2, F3-A2 EEG channel; F4_A1, F4-A1 EEG channel; O1_A2, O1-A2 EEG channel; O2_A1, O2-A1 EEG channel; LempelZiv, Lempel-Ziv; max, maximum value; SampEnt, Sample Entropy; ShanEnt, Shannon Entropy.

5.3. Correlations between EEG features and CSF biomarkers

We examined the relationships between qEEG features and CSF biomarkers. Figure 3 presents the statistically significant correlations, both before and after adjustment for potential confounding factors.

Based on predefined selection criteria, the following EEG features were retained for analysis:

- Frontal region (F3-A2; F4-A1): Variance, maximum amplitude, and Shannon Entropy during slow-wave sleep (N3).
- Central region (C3-A2; C4-A1): Variance, maximum amplitude, and Shannon Entropy during N3.
- Occipital region (O1-A2; O2-A1): Shannon Entropy (O1-A2) and variance and maximum amplitude (O2-A1) during N3.
- REM sleep feature: Skewness in the frontal region (F3).

In addition, we examined the correlations among the CSF biomarkers themselves. Figure 2 illustrates these intercorrelations, highlighting strong associations across all biomarkers. This supports their integration into one or two principal components (PCA) to enhance predictive modeling accuracy.

5.4. CSF prediction

The performance metrics used to evaluate these models include R-squared (R^2) and Root Mean Squared Error (RMSE), as reported in Table 2. R^2 quantifies

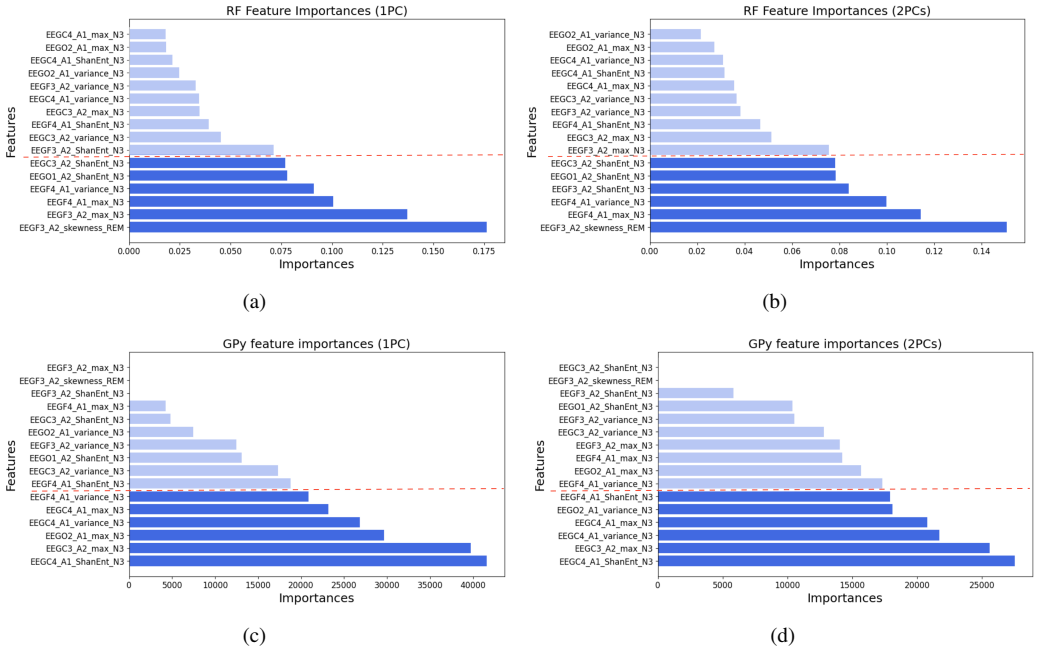


Figure 4: Importance of retained EEG features in cerebrospinal fluid (CSF) biomarker prediction. The top six most relevant features are highlighted for each model. (a) Random Forest (RF) model trained to predict the first Principal Component (1PC) of CSF biomarkers. (b) RF model trained to predict two Principal Components (2PCs) of CSF biomarkers. (c) Gaussian Process (GP) model trained for the prediction of 1PC. (d) GP model trained for the prediction of 2PCs. Features extracted from SWS (N3) and REM sleep, particularly from frontal and central brain regions, played a crucial role in biomarker prediction. These results underscore the relevance of qEEG-derived features in assessing AD-related oxidative stress and the potential integration into non-invasive diagnostic frameworks.

	Random Forest 1PC			Random Forest 2PCs			Gaussian Process 1PC			Gaussian Process 2PCs		
	R^2	RMSE	RMSE (norm)	R^2	RMSE	RMSE (norm)	R^2	RMSE	RMSE (norm)	R^2	RMSE	RMSE (norm)
GSA	0.9415	742.16	0.1195	0.9232	892.68	0.1421	0.9860	362.18	0.0275	0.9742	516.55	0.0748
AASA	0.8201	911.77	0.1469	0.7859	1038.71	0.1648	0.8548	818.68	0.0778	0.8516	864.59	0.1191
CEL	0.9014	700.01	0.1127	0.9185	678.38	0.2407	**0.9865**	**258.86**	**0.0370**	**0.9856**	**284.36**	**0.0479**
MCL	0.9428	580.67	0.0935	0.9571	490.41	0.1099	0.8747	858.81	0.1828	0.8823	812.74	0.1848
MDAL	**0.9984**	**114.86**	**0.0185**	**0.9811**	**415.86**	**0.1678**	0.9841	364.21	0.0529	0.9374	757.32	0.0595

Table 2: Comparison of performance metrics for CSF biomarker prediction across different models. The table presents the R^2 score, Root Mean Squared Error (RMSE), and normalized RMSE for each biomarker, evaluated on the test set. Results are shown for four models: RF with 1PC, RF with 2PCs, GP with 1PC and GP with 2PCs. Higher R^2 values indicate stronger predictive performance, while lower RMSE values reflect greater accuracy. In bold, the best performing results for RF (MDAL) and GP (CEL) are highlighted. Key: Aβ42, amyloid-beta protein; AASA, aminoadipic semialdehyde; CEL, Nε-carboxyethyl-lysine; CML, Nε-Carboxymethyl-lysine; GSA, glutamic semialdehyde; MDAL, Nε-malondialdehyde-lysine; RMSE, Root Mean Squared Error; R^2, R-squared.

how well the regression model explains the variance in the dependent variable, indicating the proportion of variability accounted for by the predictor variables [50]. Its values range from 0 to 1, with higher values reflecting better model performance. In contrast, RMSE measures the average deviation between predicted and observed values, providing an absolute measure of prediction accuracy. Lower RMSE values indicate a superior model fit [50].

In our evaluation, RF and GP models produced comparable performance metrics. Among the RF models, the most stable and consistent performance across all CSF biomarkers was achieved by the model trained directly with the retained EEG variables for predicting a single Principal Component (1PC) of the biomarkers. This model consistently achieved R^2 values above 0.94 for all biomarkers. When considering normalized data, the highest RMSE was 0.1469 (AASA prediction), while all other biomarkers had RMSE values below 0.12. Notably, GSA and MDAL exhibited the lowest prediction error and highest R^2 values, indicating

the strongest predictive reliability.

A second model, trained by merging three principal components for EEG features (capturing 88.69% of the database variance) and one PC for biomarkers, achieved similar overall metrics and errors (data not shown). However, its consistency across biomarkers was slightly reduced, suggesting that direct training with retained EEG variables may offer more robust predictions.

In both 1PC and 2PC models, GSA consistently exhibits the highest weighting in PCA, reinforcing its importance as a key biomarker. Additionally, five EEG features consistently hold the highest weights during RF training, as shown in Figure 4:

- EEGF3_A2_max_N3 (maximum amplitude, frontal region, N3 sleep),
- EEGF3_A2_skewness_REM (skewness, frontal region, REM sleep),
- EEGC3_A2_ShanEnt_N3 (Shannon Entropy, central region, N3 sleep),
- EEGF4_A1_variance_N3 (variance, frontal region, N3 sleep),
- EEGF4_A1_max_N3 (maximum amplitude, frontal region, N3 sleep).

These features also dominate GP modeling for 1PC prediction, suggesting that they provide the most informative contributions in simultaneous biomarker prections.

The GP model performs comparably to the RF model, as illustrated in Figure 5. It achieves $R^2 = 0.98$ for MDAL, GSA, and CEL, but exhibits slightly lower

performance for AASA ($R^2 < 0.86$). A key advantage of GP modeling is its ability to estimate uncertainty. However, the associated error margins are relatively large, especially at higher biomarker levels, indicating lower confidence in predictions. This may stem from the limited dataset size, particularly the scarcity of samples from patients with advanced AD.

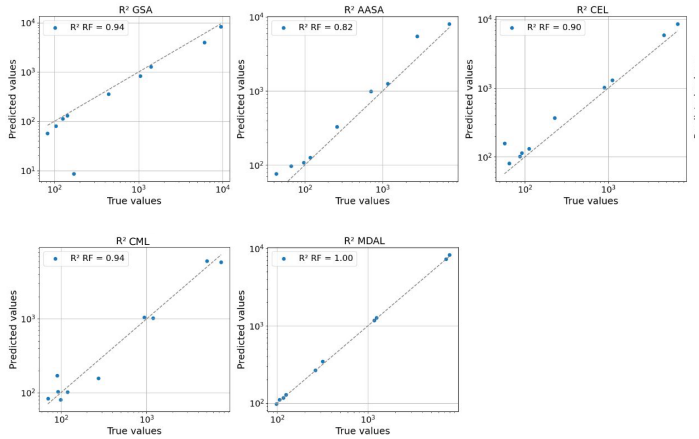
These findings also provide insight into potential uncertainty in the RF model, which does not generate uncertainty estimates. Both models exhibit a trend where higher biomarker levels correlate with increased prediction error, highlighting the challenge of accurately predicting extreme biomarker values.

6. Discussion

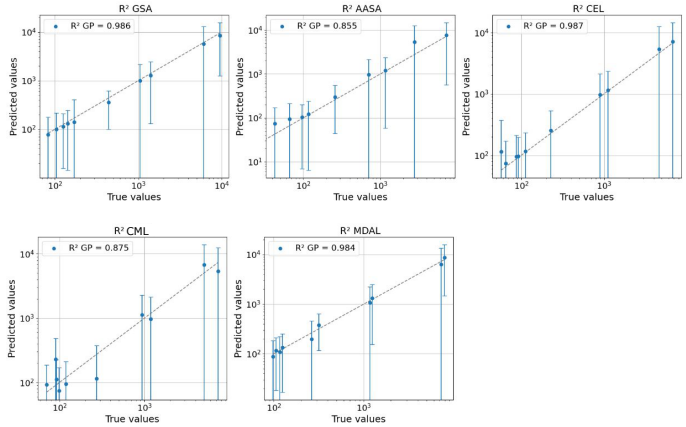
In this study, we developed advanced ML models to assess the predictive capabilities of multiple nonlinear and temporal statistics features extracted from sleep EEG for the levels of CSF protein oxidation markers in early-stage AD. Our findings reveal, for the first time, that maximum amplitude, variance, and Shannon Entropy during SWS, as well as skewness during REM sleep — particularly over frontal brain regions — are robust predictors of non-enzymatic oxidative stress biomarkers. These markers, which include protein oxidation, advanced glycation end-products (AGEs), and advanced lipoxidation end-products (ALEs), are established indicators of irreversible oxidative damage in AD brains.

6.1. Integration of Functional and Molecular Markers

Our results introduce a novel, clinically relevant framework that integrates functional neurophysiological data (sleep EEG) with molecular biomarkers, bridg-



(a)



(b)

Figure 5: R^2 score visualizations for test predictions across all biomarkers, presented on a logarithmic scale to enhance clarity. The two top-performing models are shown: (a) Random Forest with 1 Principal Component (1PC) and (b) Gaussian Process with 1 PC, incorporating uncertainty values associated with each prediction. Each plot compares predicted vs. true values for cerebrospinal fluid (CSF) oxidative stress biomarkers, demonstrating the accuracy and robustness of the models. Higher R^2 values indicate stronger predictive performance. Error bars represent the variance of the estimation, with RF models exhibiting much lower variance than GP models, indicating more stable predictions. Electrode positions are denoted as C3_A2, C4_A1, F3_A2, F4_A1, O1_A2, and O2_A1, corresponding to standard EEG derivations used in the analysis. Key: AASA, aminoadipic semialdehyde; CEL, Nε-carboxyethyl-lysine; CML, Nε-carboxymethyl-lysine; GSA, glutamic semialdehyde; MDAL, Nε-malondialdehyde-lysine; N3, N3 sleep stage; max, maximum value; R^2, R-squared; ShanEnt, Shannon Entropy.

ing two fundamental aspects of AD pathology: network dysfunction and oxidative molecular damage [6, 17]. By leveraging non-invasive sleep EEG alongside CSF biomarker analysis, we enhance the biological specificity of qEEG as an early detection tool.

Recent studies suggest that nonlinear EEG metrics detect AD-related brain alterations that traditional frequency-based analyses often miss [26, 51]. Features such as Shannon Entropy capture declining brain signal complexity, while variance and skewness reflect regional connectivity disruptions characteristics of AD [52, 53, 54]. However, limited research has explored these markers specifically in sleep EEG in AD - a gap this study addresses.

6.2. Bidirectional Relationship Between Sleep and Oxidative Stress

The differentiation of EEG features across SWS and REM sleep highlights the bidirectional relationship between sleep disturbances and oxidative stress [17].

Disruptions in sleep architecture—especially the reduction of SWS and alterations in REM sleep—not only reflect the impact of oxidative damage on thalamocortical and corticocortical circuits involved in sleep regulation but also exacerbate oxidative stress [55, 56, 57].

SWS plays a key role in cellular repair, mitochondrial function, and antioxidant defence [55]. Reduced SWS duration or fragmentation is associated with elevated CSF oxidative stress markers and accelerated AD progression [56]. Similarly, the glymphatic system, which clears neurotoxic waste such as Aβ, operates most effectively during NREM sleep, particularly SWS [58, 59]. Disruptions to

Aβ clearance from SWS impairment or hypoxia (e.g., sleep apnea) may promote Aβ accumulation and neuronal vulnerability [60].

REM sleep also plays a critical role in neurotoxin clearance and antioxidant regulation. Its deprivation increases oxidative stress, impairs cognitive function, and depletes antioxidants, linking reduced REM sleep to lower CSF Aβ levels and cholinergic neurons degeneration in AD [57, 61, 62, 55].

Our findings align with region-specific oxidative damage in AD, showing that qEEG features predictive of oxidative stress markers are most prominent in the frontal brain during SWS and REM sleep [19, 63, 64].

6.3. Strengths and Limitations

Our study has several key strengths. First, we analyzed a rigorously defined cohort of mild to moderate AD patients, ensuring a well-characterized population. Second, we utilized a comprehensive assessment battery, including overnight PSG and CSF biomarker analysis, a combination rarely used in AD research. Third, we included only drug-free participants to eliminate potential medication-induced distortions in REM sleep EEG features [65].

A major strength of our approach is the integration of ML to uncover complex, nonlinear relationships between qEEG features and CSF oxidative stress biomarkers, relationships that traditional linear methods often overlook [66, 67]. Unlike prior AD studies that primarily relied on linear EEG metrics [68, 55], our work incorporates nonlinear and time-domain EEG features, offering a deeper understanding of oxidative stress-related neurophysiological alterations.

Despite these strengths, some limitations must be acknowledged. The primary constraint is the small dataset from a single center, which may limit generalizability. However, our use of advanced ML models helped mitigate this issue by enhancing predictive performance. While a low number of patients with high CSF biomarker levels contributed to increased prediction error (see Figure 5), both RF and GP models demonstrated strong predictive accuracy, consistently identifying key EEG features. Furthermore, GP modeling provided uncertainty estimates, particularly at higher biomarker levels, reinforcing confidence in predictions at lower values, which could improve with larger datasets.

Other limitations include the use of only six EEG channels, restricting detailed topographic analysis, and the absence of a healthy control group due to ethical concerns surrounding CSF collection. Additionally, while nonenzymatic protein biomarkers show promise for early AD risk assessment and disease progression, their clinical validation remains ongoing [37].

6.4. Clinical and Research Implications

This integrative qEEG-CSF biomarker strategy could pioneer a paradigm shift in AD risk stratification and early diagnosis. By demonstrating that non-invasive sleep EEG features reliably predict key oxidative stress biomarkers, we propose a cost-effective alternative to invasive CSF sampling [69].

Moving away from invasive molecular markers toward a non-invasive, sleep-based monitoring system could enable clinicians to identify individuals at risk for AD—where early intervention is critical [8, 9]— as well as track disease progres-

sion in AD patients with heightened oxidative stress-related network damage.

This approach could be particularly valuable for large-scale screening of asymptomatic individuals, allowing for early intervention before clinical symptoms emerge. It can be seamlessly integrated into longitudinal monitoring protocols, complementing traditional biomarkers and neuroimaging methods [9].

Additionally, sleep is a modifiable risk factor for oxidative stress, making this strategy highly relevant for personalized prevention. Targeted lifestyle and sleep interventions - including cognitive behavioral therapy for insomnia (CBT-I), transcranial stimulation, or melatonin therapy - could improve SWS and REM sleep, reduce oxidative damage and potentially delay or prevent symptom onset [19, 70, 5].

Furthermore, EEG-derived features could serve as biomarkers to assess treatment efficacy, offering real-time insights into the effects of oxidative stress-targeting therapies on neural activity [5]. Lastly, our methodology may propel research efforts toward the validation of CSF oxidatively modified protein biomarkers for clinical use without requiring invasive sampling procedures.

6.5. Future Directions

Future research should validate these findings in larger, diverse cohorts, including preclinical AD cases and healthy older adults, to distinguish AD pathology from normal aging [9]. Longitudinal studies are essential to assessing the stability and predictive value of qEEG biomarkers over time and their role in disease progression and treatment response.

Additionally, studies should explore how qEEG-based oxidative stress markers correlate with established AD biomarkers ($A\beta$, p-tau, t-tau) to refine risk models.

AI-driven approaches could automate qEEG feature selection and classification, improving diagnostic accuracy while deep learning techniques enhance model generalizability [71, 72]. High-density EEG should also be incorporated to improve spatial resolution, enabling a more detailed topographic analysis of AD-related network disruptions [73, 74].

Future research should account for factors (e.g., OSA, ApoE) which exacerbate oxidative damage and may affect qEEG-based biomarker accuracy in AD populations [16, 20, 75]. Lastly, intervention-based studies should evaluate whether sleep-targeted therapies [5] can modulate qEEG features, reduce oxidative stress, and delay cognitive decline.

7. Conclusions

Our study demonstrates that advanced nonlinear and time-domain sleep EEG features, combined with an ML, reliably predict CSF protein oxidation-derived markers in early AD. This approach outperforms traditional methods and offers a non-invasive, cost-effective alternative to lumbar puncture sampling.

Future research will focus on developing customized algorithms for different patient populations, refining predictive models, and integrating qEEG biomarkers into early AD detection frameworks. If validated, this strategy could facilitate personalized risk assessment and targeted interventions aimed at reducing oxidative

stress and slowing neurodegeneration in aging populations.

Data availability

The data supporting this study are available upon reasonable request to the corresponding author.

CRediT authorship contribution statement

Anna Michela Gaeta: Conceptualization, Writing – original draft, Investigation, Methodology, Visualization. Lorena Gallego Viñarás: Formal analysis, Writing – original draft, Methodology, Investigation, Data curation, Visualization. Ferran Barbé: Project administration, Writing – review & editing, Investigation, Methodology. Pablo M. Olmos: Writing – review & editing, Formal analysis, Methodology, Investigation, Visualization. Reinald Pamplona: Methodology, Data curation, Writing – review & editing. Farida Dakterzada: Methodology, Data curation, Writing – review & editing. Arrate Muñoz-Barrutia: Project administration, Investigation, Formal analysis, Methodology, Writing – review & editing. Gerard Piñol-Ripoll: Project administration, Conceptualization, Investigation, Methodology, Writing – review & editing.

Declaration of interest

The authors affirm that they have no financial or personal conflicts of interest that could have potentially influenced the outcomes of this study.

Funding

This research project was funded by the Generalitat of Catalonia: Department of Health (PERIS 2019 SLT008/18/00050) and Agency for Management of University and Research Grants (2021SGR00761), the "Fundació La Marató TV3" (464/C/2014), the Institute of Health Carlos III (PI22/01687), and the Diputació de Lleida (PIRS 2021 and PIRS2024) to GP-R. This research was also funded by the Diputació de Lleida (PIRS2021 and PIRS2024), and the Generalitat of Catalonia: Department of Health (PERIS ref. SLT002/16/00250) and Agency for Management of University and Research Grants (2021SGR00990) to R.P. The study has also been partially supported under grants PID2021-123182OB-I00, PID2023-152631OB-I00 and PID2023-152233OB-I00, by Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación, MCIN/AEI/10.13039/501100011033, co-financed by European Regional Development Fund (ERDF), 'A way of making Europe. PMO also acknowledges the support by BBVA Foundation (Beca Leonardo a Investigadores y Creadores Culturales 2024) and by the Comunidad de Madrid under grants IND2022/TIC-23550 and ELLIS Unit Madrid. IRBLleida is a CERCA Programme/Generalitat of Catalonia. FB is supported by the ICREA program, Generalitat of Catalonia.

Acknowledgements

We want to offer our heartfelt appreciation to all the patients and staff at the Hospital Universitari Santa Maria's Sleep and Dementia Unit. The IRBLleida Biobank (B.0000682) and plataforma biobancos PT17/0015/0027/ also assisted

us. All authors have revised and approved the submission of the manuscript. The authors utilized the language model ChatGPT developed by OpenAI to assist in drafting this paper.

8. References

[1] C. Qiu, M. Kivipelto, E. von Strauss, Epidemiology of alzheimer's disease: Occurrence, determinants, and strategies toward intervention, Dialogues in Clinical Neuroscience 11 (2) (2009) 111–128.

[2] C. L. Masters, R. Bateman, K. Blennow, C. C. Rowe, R. A. Sperling, J. L. Cummings, Alzheimer's disease, Nature Reviews Disease Primers 1 (2015) 15056. doi:10.1038/nrdp.2015.56.

[3] I. Ferrer, Hypothesis review: Alzheimer's overture guidelines, Brain Pathology 33 (1) (2023) e13122.

[4] M. Jové, N. Mota-Martorell, P. Torres, V. Ayala, M. Portero-Otin, I. Ferrer, R. Pamplona, The causal role of lipoxidative damage in mitochondrial bioenergetic dysfunction linked to alzheimer's disease pathology, Life 11 (5) (2021) 388.

[5] R. Dhapola, S. K. Beura, P. Sharma, S. K. Singh, D. HariKrishnaReddy, Oxidative stress in alzheimer's disease: current knowledge of signaling pathways and therapeutics, Molecular Biology Reports 51 (1) (2024) 48.

[6] E. Tönnies, E. Trushina, Oxidative stress, synaptic dysfunction, and alzheimer's disease, Journal of Alzheimer's Disease 57 (4) (2017) 1105–1121, pMID: 28059794; PMCID: PMC5409043. doi:10.3233/JAD-161088.

[7] X. Wang, E. K. Michaelis, Selective neuronal vulnerability to oxidative stress in the brain, Frontiers in aging neuroscience 2 (2010) 1224.

[8] C.-C. Tan, J.-T. Yu, L. Tan, Biomarkers for preclinical alzheimer's disease, Journal of Alzheimer's Disease 42 (4) (2014) 1051–1069.

[9] C. R. Jack Jr, J. S. Andrews, T. G. Beach, T. Buracchio, B. Dunn, A. Graf, O. Hansson, C. Ho, W. Jagust, E. McDade, et al., Revised criteria for diagnosis and staging of alzheimer's disease: Alzheimer's association workgroup, Alzheimer's & Dementia 20 (8) (2024) 5143–5169.

[10] J. R. Requena, C.-C. Chao, R. L. Levine, E. R. Stadtman, Glutamic and aminoadipic semialdehydes are the main carbonyl products of metal-catalyzed oxidation of proteins, Proceedings of the National Academy of Sciences 98 (1) (2001) 69–74.

[11] A. Tramutola, C. Lanzillotta, M. Perluigi, D. A. Butterfield, Oxidative stress, protein modification and alzheimer disease, Brain research bulletin 133 (2017) 88–96.

[12] S. Thorpe, J. Baynes, Maillard reaction products in tissue proteins: new products and new perspectives, Amino acids 25 (2003) 275–281.

[13] R. Pamplona, E. Dalfó, V. Ayala, M. J. Bellmunt, J. Prat, I. Ferrer, M. Portero-Otín, Proteins in human brain cortex are modified by oxidation, glycoxidation, and lipoxidation: effects of alzheimer disease and identification of lipoxidation targets, Journal of Biological Chemistry 280 (22) (2005) 21522–21530.

[14] F. C. Quevenco, M. G. Preti, J. M. Van Bergen, J. Hua, M. Wyss, X. Li, S. J. Schreiner, S. C. Steininger, R. Meyer, I. B. Meier, et al., Memory performance-related dynamic brain connectivity indicates pathological burden and genetic risk for alzheimer's disease, Alzheimer's research & therapy 9 (2017) 1–11.

[15] S. Matsumoto, T. Tsunematsu, Association between sleep, alzheimer's, and parkinson's disease, Biology 10 (11) (2021) 1127.

[16] E. N. Minakawa, K. Wada, Y. Nagai, Sleep disturbance as a potential modifiable risk factor for alzheimer's disease, International Journal of Molecular Sciences 20 (4) (2019) 803, pMID: 30781802; PMCID: PMC6412395. doi:10.3390/ijms20040803.

[17] C. Czekus, P. Steullet, A. Orero López, I. Bozic, T. Rusterholz, M. Bandarabadi, K. Q. Do, C. Gutierrez Herrera, Alterations in trn-anterodorsal thalamocortical circuits affect sleep architecture and homeostatic processes in oxidative stress vulnerable gclm-/- mice, Molecular psychiatry 27 (11) (2022) 4394–4406.

[18] A. Ballesteros, A. Summerfelt, X. Du, P. Jiang, J. Chiappelli, M. Tagamets, P. O'Donnell, P. Kochunov, L. E. Hong, Electrophysiological intermediate biomarkers for oxidative stress in schizophrenia, Clinical Neurophysiology 124 (11) (2013) 2209–2215.

[19] S. M. Romanella, D. Roe, E. Tatti, D. Cappon, R. Paciorek, E. Testani, A. Rossi, S. Rossi, E. Santarnecchi, The sleep side of aging and alzheimer's disease, Sleep Medicine 77 (2021) 209–225, epub 2020 May 30, PMID: 32912799; PMCID: PMC8364256. doi:10.1016/j.sleep.2020.05.029.

[20] F. Di Domenico, G. Pupo, E. Giraldo, M. C. Badìa, P. Monllor, A. Lloret, et al., Oxidative signature of cerebrospinal fluid from mild cognitive impairment and alzheimer disease patients, Free Radical Biology and Medicine 91 (2016) 1–9, epub 2015 Dec 8, PMID: 26675344. doi:10.1016/j.freeradbiomed.2015.12.004.

[21] M. Haddad, M. Perrotte, S. Landri, A. Lepage, T. Fülöp, C. Ramassamy, Circulating and extracellular vesicles levels of n-(1-carboxymethyl)-l-lysine (cml) differentiate early to moderate alzheimer's disease, Journal of Alzheimer's Disease 69 (2019) 751–762. doi:10.3233/JAD-181272. URL https://doi.org/10.3233/JAD-181272

[22] A. Klimiuk, M. Maciejczyk, M. Choromańska, K. Fejfer, N. Waszkiewicz, A. Zalewska, Salivary redox biomarkers in different stages of dementia

severity, Journal of Clinical Medicine 8 (6) (2019) 840. doi:10.3390/jcm8060840.
URL https://doi.org/10.3390/jcm8060840

[23] K. T. Kim, Lumbar puncture: considerations, procedure, and complications, encephalitis 2 (4) (2022) 93.

[24] N. V. Thakor, S. Tong, Advances in quantitative electroencephalogram analysis methods, Annual Review of Biomedical Engineering 6 (2004) 453–495, pMID: 15255777. doi:10.1146/annurev.bioeng.5.040202.121601.

[25] J. Jeong, Nonlinear dynamics of eeg in alzheimer's disease, Drug Development Research 56 (2002) 57–66.

[26] J. Jeong, Eeg dynamics in patients with alzheimer's disease, Clinical neurophysiology 115 (7) (2004) 1490–1505.

[27] D. Bzdok, N. Altman, M. Krzywinski, Statistics versus machine learning, Nature Methods 15 (2018) 233–234. doi:10.1038/nmeth.4642.

[28] L. A. Gemein, R. T. Schirrmeister, P. Chrabąszcz, D. Wilson, J. Boedecker, A. Schulze-Bonhage, F. Hutter, T. Ball, Machine-learning-based diagnostics of eeg pathology, NeuroImage 220 (2020) 117021.

[29] A. H. Al-Nuaimi, M. Blūma, S. S. Al-Juboori, C. S. Eke, E. Jammeh, L. Sun, E. Ifeachor, Robust eeg based biomarkers to detect alzheimer's disease, Brain Sciences 11 (8) (2021) 1026.

[30] S.-S. Poil, W. De Haan, W. M. van der Flier, H. D. Mansvelder, P. Scheltens, K. Linkenkaer-Hansen, Integrative eeg biomarkers predict progression to alzheimer's disease at the mci stage, Frontiers in aging neuroscience 5 (2013) 58.

[31] A. M. Gaeta, M. Quijada-López, F. Barbé, R. Vaca, M. Pujol, O. Minguez, M. Sánchez-de-la Torre, A. Muñoz-Barrutia, G. Piñol-Ripoll, Predicting alzheimer's disease csf core biomarkers: a multimodal machine learning approach, Frontiers in Aging Neuroscience 16 (2024) 1369545. doi: 10.3389/fnagi.2024.1369545.

[32] L. Gallego-Viñarás, J. M. Mira-Tomás, A. M. Gaeta, G. Pinol-Ripoll, F. Barbé, P. M. Olmos, A. Muñoz-Barrutia, Alzheimer's disease detection in eeg sleep signals, IEEE Journal of Biomedical and Health Informatics (2024).

[33] A. M. Gaeta, I. D. Benítez, C. Jorge, G. Torres, F. Dakterzada, O. Minguez, R. Huerto, M. Pujol, A. Carnes, M. Dalmases, et al., Prevalence of obstructive sleep apnea in alzheimer's disease patients, Journal of neurology 267 (2020) 1012–1022.

[34] G. M. McKhann, D. S. Knopman, H. Chertkow, et al., The diagnosis of dementia due to alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease, Alzheimer's & Dementia: The Journal of the

Alzheimer's Association 7 (3) (2011) 263–269. doi:10.1016/j.jalz.2011.03.005.

[35] A. Lobo, P. Saz, G. Marcos, et al., Revalidación y normalización del mini-examen cognoscitivo (primera version en castellano del mini-mental status examination) en la población general geriátrica, Medicina Clínica 112 (1999) 767–774. doi:10.1016/S0025-7753(05)70654-4.

[36] L. Zhong, Y. Z. Xie, T. T. Cao, Z. Wang, T. Wang, X. Li, et al., A rapid and cost-effective method for genotyping apolipoprotein e gene polymorphism, Molecular Neurodegeneration 11 (2016) 2, pMID: 26754117; PMCID: PMC4710020. doi:10.1186/s13024-016-0069-4.

[37] F. Dakterzada, M. Jové, J. L. Cantero, R. Pamplona, G. Pinoll-Ripoll, Plasma and cerebrospinal fluid nonenzymatic protein damage is sustained in alzheimer's disease, Redox Biology 64 (2023) 102772.

[38] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, R. M. Lloyd, C. L. Marcus, B. V. Vaughn, The AASM Manual for the Scoring of Sleep and Associated Events: Rules, terminology, and technical specifications Version 2.2, American Academy of Sleep Medicine, 2015.

[39] A. Bandyopadhyay, C. Goldstein, Clinical applications of artificial intelligence in sleep medicine: a sleep clinician's perspective, Sleep and Breathing 27 (1) (2023) 39–55.

[40] M. Shouran, E. Elgamli, Design and implementation of butterworth filter, International Journal of Innovative Research in Science, Engineering and Technology 9 (9) (2020).
URL https://www.ijirset.com/upload/2020/september/Design_Butterworth.pdf

[41] A. Bezbochina, E. Stavinova, A. Kovantsev, P. Chunaev, Enhancing predictability assessment: An overview and analysis of predictability measures for time series and network links, Entropy 25 (11) (2023) 1542.

[42] D. Alvarez, R. Hornero, J. V. Marcos, N. Wessel, T. Penzel, M. Glos, F. Del Campo, Assessment of feature selection and classification approaches to enhance information from overnight oximetry in the context of apnea diagnosis, International journal of neural systems 23 (05) (2013) 1350020.

[43] H. He, B. Wu, Y. Bi, Data Mining for Biomedical Applications, Springer, Cham, 2018. doi:10.1007/978-3-319-66644-0.
URL https://link.springer.com/book/10.1007/978-3-319-66644-0

[44] D. Álvarez, R. Hornero, J. V. Marcos, F. Del Campo, Feature selection from nocturnal oximetry using genetic algorithms to assist in obstructive sleep apnoea diagnosis, Medical engineering & physics 34 (8) (2012) 1049–1057.

[45] I. T. Jolliffe, J. Cadima, Principal component analysis: A review and recent developments, Philosophical Transactions of the Royal Society A: Mathe-

matical, Physical and Engineering Sciences 374 (2065) (2016) 20150202. doi:10.1098/rsta.2015.0202.

[46] J. N. Wulff, L. E. Jeppesen, Multiple imputation by chained equations in praxis: Guidelines and review, Electronic Journal of Business Research Methods 15 (1) (2017) 41–56.
URL http://www.ejbrm.com/volume15/issue1

[47] M. Mera-Gaona, U. Neumann, R. Vargas-Canas, D. M. López, Evaluating the impact of multivariate imputation by mice in feature selection, PLOS ONE 16 (7) (2021) e0254720, erratum in: PLoS One. 2021 Dec 16;16(12):e0261739, PMID: 34320016; PMCID: PMC8318311. doi:10.1371/journal.pone.0254720.

[48] O. Sagi, L. Rokach, Explainable decision forest: Transforming a decision forest into an interpretable tree, Information Fusion 61 (2020) 124–138. doi:10.1016/j.inffus.2020.03.013.

[49] C. E. Rasmussen, Gaussian processes in machine learning, in: Advanced Lectures on Machine Learning, Springer, Berlin, Heidelberg, 2004, pp. 63–71. doi:10.1007/978-3-540-28650-9_4.

[50] D. Chicco, M. J. Warrens, G. Jurman, The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation, PeerJ Computer Science 7 (2021) e623, pMID: 34307865; PMCID: PMC8279135. doi:10.7717/peerj-cs.623.

[51] C. A. Chetty, H. Bhardwaj, G. P. Kumar, T. Devanand, C. A. Sekhar, T. Aktürk, I. Kiyi, G. Yener, B. Güntekin, J. Joseph, et al., Eeg biomarkers in alzheimer's and prodromal alzheimer's: a comprehensive analysis of spectral and connectivity features, Alzheimer's Research & Therapy 16 (1) (2024) 236.

[52] Y. Karaca, M. Moonis, Shannon entropy-based complexity quantification of nonlinear stochastic process: diagnostic and predictive spatiotemporal uncertainty of multiple sclerosis subgroups, in: Y. Karaca, D. Baleanu, Y.-D. Zhang, O. Gervasi, M. Moonis (Eds.), Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different Complex Systems, Academic Press, 2022, pp. 231–245. doi:10.1016/B978-0-323-90032-4.00018-3.

[53] Y. Aoki, R. Takahashi, Y. Suzuki, R. D. Pascual-Marqui, Y. Kito, S. Hikida, K. Maruyama, M. Hata, R. Ishii, M. Iwase, et al., Eeg resting-state networks in alzheimer's disease associated with clinical symptoms, Scientific Reports 13 (1) (2023) 3964.

[54] T. Koenig, D. Studer, D. Hubl, L. Melie, W. Strik, Brain connectivity at different time-scales measured with eeg, Philosophical Transactions of the Royal Society B: Biological Sciences 360 (1457) (2005) 1015–1024.

[55] B. A. Mander, J. R. Winer, W. J. Jagust, M. P. Walker, Sleep: A novel mechanistic pathway, biomarker, and treatment target in the pathology of

alzheimer's disease?, Trends in Neurosciences 39 (8) (2016) 552–566. doi:10.1016/j.tins.2016.05.002.

[56] B. P. Lucey, A. McCullough, E. C. Landsness, C. D. Toedebusch, J. S. McLeland, A. M. Zaza, A. M. Fagan, L. McCue, C. Xiong, J. C. Morris, et al., Reduced non–rapid eye movement sleep is associated with tau pathology in early alzheimer's disease, Science translational medicine 11 (474) (2019) eaau6550.

[57] D. C. Mathangi, R. Shyamala, A. S. Subhashini, Effect of rem sleep deprivation on the antioxidant status in the brain of wistar rats, Annals of Neurosciences 19 (4) (2012) 161–164. doi:10.5214/ans.0972.7531.190405.

[58] L. M. Hablitz, H. S. Vinitsky, Q. Sun, F. F. Stæger, B. Sigurdsson, K. N. Mortensen, et al., Increased glymphatic influx is correlated with high eeg delta power and low heart rate in mice under anesthesia, Science Advances 5 (2019) eaav5447. doi:10.1126/sciadv.aav5447.

[59] L. Xie, H. Kang, Q. Xu, M. J. Chen, Y. Liao, M. Thiyagarajan, J. O'Donnell, D. J. Christensen, C. Nicholson, J. J. Iliff, T. Takano, R. Deane, M. Nedergaard, Sleep drives metabolite clearance from the adult brain, Science 342 (6156) (2013) 373–377. doi:10.1126/science.1241224.

[60] C. R. Kyrtsos, J. S. Baras, Modeling the role of the glymphatic pathway and

cerebral blood vessel properties in alzheimer's disease pathogenesis, PLOS ONE 10 (2015) e0139574. doi:10.1371/journal.pone.0139574.

[61] B. N. Mallick, S. Gulyani, Alterations in synaptosomal calcium concentrations after rapid eye movement sleep deprivation in rats, Neuroscience 73 (3) (1996) 729–736. doi:10.1016/0306-4522(96)00040-5.

[62] L. Ramanathan, S. Gulyani, R. Nienhuis, et al., Sleep deprivation decreases superoxide dismutase activity in rat hippocampus and brain stem, Neuroreport 13 (11) (2002) 1387–1390. doi:10.1097/01.wnr.0000024388.08782.f8.

[63] D. Chylinski, M. Van Egroo, J. Narbutas, V. Muto, M. A. Bahri, C. Berthomier, et al., Timely coupling of sleep spindles and slow waves linked to early amyloid-β burden and predicts memory decline, eLife 11 (2022) e78191. doi:10.7554/eLife.78191.

[64] A. Takeda, M. Wakai, H. Niwa, R. Dei, M. Yamamoto, M. Li, et al., Neuronal and glial advanced glycation end product [nε-(carboxymethyl)lysine] in alzheimer's disease brains, Acta Neuropathologica 101 (2001) 27–35. doi:10.1007/s004010000256.

[65] H. Azami, S. Moguilner, H. Penagos, R. A. Sarkis, S. E. Arnold, S. N. Gomperts, A. D. Lam, Eeg entropy in rem sleep as a physiologic biomarker in early clinical stages of alzheimer's disease, Journal of Alzheimer's Dis-

ease 91 (4) (2023) 1557–1572, pMID: 36641682; PMCID: PMC10039707. doi:10.3233/JAD-221152.

[66] C. H. Chang, C. H. Lin, C. Y. Liu, C. S. Huang, S. J. Chen, et al., Plasma *d*-glutamate levels for detecting mild cognitive impairment and alzheimer's disease: machine learning approaches, Journal of Psychopharmacology 35 (2021) 265–272. doi:10.1177/0269881120972331.

[67] C. A. Lovejoy, A. Abbas, D. Ratneswaran, An introduction to artificial intelligence in sleep medicine, Journal of Thoracic Disease 13 (10) (2021) 6095–6098. doi:10.21037/jtd-21-1569.

[68] A. D'Atri, S. Scarpelli, M. Gorgoni, I. Truglia, G. Lauri, S. Cordone, et al., Eeg alterations during wake and sleep in mild cognitive impairment and alzheimer's disease, iScience 24 (4) (2021) 102386, pMID: 33981973; PMCID: PMC8086022. doi:10.1016/j.isci.2021.102386.

[69] P. Monllor, A. Cervera-Ferri, M. A. Lloret, D. Esteve, B. Lopez, J. L. Leon, A. Lloret, Electroencephalography as a non-invasive biomarker of alzheimer's disease: A forgotten candidate to substitute csf molecules?, International Journal of Molecular Sciences 22 (19) (2021) 10889, pMID: 34639229; PMCID: PMC8509134. doi:10.3390/ijms221910889.

[70] B. P. Lucey, It's complicated: The relationship between sleep and alzheimer's disease in humans, Neurobiology of Disease 144 (2020) 105031, epub 2020 Jul 29. doi:10.1016/j.nbd.2020.105031.

[71] M. H. Kabir, S. Mahmood, A. Al Shiam, A. S. M. Miah, J. Shin, M. K. I. Molla, Investigating feature selection techniques to enhance the performance of eeg-based motor imagery tasks classification, Mathematics 11 (8) (2023) 1921. doi:10.3390/math11081921.

[72] V. Gerla, V. Kremen, M. Macas, D. Dudysova, A. Mladek, P. Sos, L. Lhotska, Iterative expert-in-the-loop classification of sleep psg recordings using a hierarchical clustering, Journal of Neuroscience Methods 317 (2019) 61–70. doi:10.1016/j.jneumeth.2019.01.013.

[73] R. Gordon, E. Rzempoluck, Introduction to laplacian montages, American Journal of Electroneurodiagnostic Technology 44 (2) (2004) 98–102.

[74] S. Stoyell, J. Wilmskoetter, M. Dobrota, D. Chinappen, L. Bonilha, M. Mintz, B. Brinkmann, S. Herman, J. Peters, S. Vulliemoz, M. Seeck, M. Hämäläinen, C. Chu, High-density eeg in current clinical practice and opportunities for the future, Journal of Clinical Neurophysiology 38 (2) (2021) 112–123. doi:10.1097/WNP.0000000000000807.

[75] X. Xiong, Y. Ren, S. Gao, J. Luo, J. Liao, C. Wang, S. Yi, R. Liu, Y. Xiang, J. He, Eeg microstate in obstructive sleep apnea patients, Scientific Reports 11 (1) (2021) 17178.

Appendix A. Definitions and mathematical formulas for quantitative EEG features

This supplementary material provides detailed definitions and corresponding mathematical formulas for the time-domain and non-linear features extracted from EEG signals.

Appendix A.1. Time Domain Statistics

- $Variance\ (Var)$

  Variance measures the dispersion of the EEG signal values around the mean.

$$\text{Var}(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2 \tag{A.1}$$

  where μ is the mean of the signal X, and N is the number of samples [42].

- *Skewness (Skew)*

  Skewness measures the asymmetry of the probability distribution of the EEG signal.

$$\text{Skew}(X) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i - \mu}{\sigma}\right)^3 \tag{A.2}$$

  where σ is the standard deviation of the signal X [42].

- *Kurtosis (Kurt)*

  Kurtosis measures the tailedness of the probability distribution of the EEG

signal[42].

$$\text{Kurt}(X) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i - \mu}{\sigma}\right)^4 - 3 \tag{A.3}$$

- *Maximum Value (max)*

  The maximum value is the highest amplitude observed in the EEG signal [43].

$$\text{max}(X) = \max(X) \tag{A.4}$$

Appendix A.2. Non-Linear Features

- *Shannon Entropy (ShanEnt)*

  Shannon Entropy quantifies the uncertainty or randomness in the EEG signal.

$$\text{ShanEnt}(X) = -\sum_{i=1}^{N} p(X_i)\log p(X_i) \tag{A.5}$$

  where $p(X_i)$ is the probability of the i-th signal value [41].

- *Sample Entropy (SampEnt)*

  Sample Entropy measures the complexity and irregularity of the EEG signal.

$$\text{SampEnt}(X, m, r) = -\log\left(\frac{A}{B}\right) \tag{A.6}$$

  where A is the number of matching templates of length $m + 1$ and B is the number of matching templates of length m within a tolerance r [42].

- *Lempel-Ziv (LempZiv)*

Lempel-Ziv Complexity measures the complexity of the EEG signal based on the number of distinct substrings and their recurrence rates. The formula is given by:

$$\text{LempZiv}(X) = \frac{C(X)}{\log N} \tag{A.7}$$

where $C(X)$ is the number of distinct substrings in the signal X and N is the length of the signal [42].

Appendix B. Quantitative EEG signals features

Sleep Stage	Channel	Feature Type	EEG Features
N1	F3_A2 F4_A1 C3_A2 C4_A1 O1_A2 O2_A1	Time domain Nonlinear	variance, skewness, kurtosis, max ShanEnt, SampEnt, LempZiv
N2	F3_A2 F4_A1 C3_A2 C4_A1 O1_A2 O2_A1	Time domain Nonlinear	variance, skewness, kurtosis, max ShanEnt, SampEnt, LempZiv
N3	F3_A2 F4_A1 C3_A2 C4_A1 O1_A2 O2_A1	Time domain Nonlinear	variance, skewness, kurtosis, max ShanEnt, SampEnt, LempZiv
REM	F3_A2 F4_A1 C3_A2 C4_A1 O1_A2 O2_A1	Time domain Nonlinear	variance, skewness, kurtosis, max ShanEnt, SampEnt, LempZiv

Table B.3: EEG features computed for the EEG F3-A2, F4_A1, C3_A2, C4_A1, O1_A2, O2_A1 channels, segmented by sleep stages. Key: EEG, Electroencephalogram; LempelZiv, Lempel Ziv complexity; max, maximum value; SampEnt, Sample Entropy; ShanEnt, Shannon Entropy. The notation used throughout the manuscript when referring to specific features follows the format: EEG_*'Channel'*_*'EEG Feature'*_*'Sleep Stage'*.