

Review

From Acquisition to Validation: Methodological Dependencies and Reproducibility in EEG-Based Alzheimer's Disease Detection

Ruimin Wang ¹, Takenao Sugi ¹ and Takao Yamasaki ^{2,*}¹ Faculty of Science and Engineering, Saga University, Saga 840-8502, Japan; wangrm@cc.saga-u.ac.jp (R.W.); sugi@cc.saga-u.ac.jp (T.S.)² Department of Neurology, Minkodo Minohara Hospital, Fukuoka 811-2402, Japan

* Correspondence: yamasaki@kyudai.jp

Abstract

Alzheimer's disease (AD) is a progressive neurodegenerative disorder for which early detection and reliable monitoring remain major clinical challenges. Electroencephalography (EEG) combined with machine learning has attracted growing interest as a scalable and non-invasive approach to AD detection, yet reported classification accuracies vary widely across studies and are rarely comparable or clinically translatable. One important reason is that the analytical pipeline—from data acquisition to model validation—involves numerous methodological choices whose inter-stage dependencies and reproducibility implications are rarely made explicit. In this narrative review, we adopt a **methodological chain framework** to make these dependencies explicit, organizing EEG-based AD research into five sequential stages: data acquisition, preprocessing, feature representation, modeling, and validation. Choices at each stage can shape downstream analyses, inflate reported performance, and reduce cross-study comparability in ways that are difficult to detect when stages are assessed independently. These effects are particularly consequential in EEG-based AD research, where cohorts are typically small and biomarkers are subtle. We make three primary contributions: (1) we describe inter-stage methodological dependencies that may contribute to reproducibility problems and performance inflation; (2) we synthesize major sources of methodological variability across representative EEG–AD studies and evaluate their differential impact on spectral, connectivity, and complexity features; and (3) we provide practical, stage-aligned recommendations culminating in a minimum reporting checklist.

Keywords: Alzheimer's disease; electroencephalography (EEG); methodological framework; feature representation; machine learning; validation; reproducibility; data leakage; biomarker; preprocessing



Academic Editors: Loris Nanni and Daniele Giansanti

Received: 1 April 2026

Revised: 8 May 2026

Accepted: 10 May 2026

Published: 13 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the leading cause of dementia worldwide, characterized by gradual cognitive decline in memory, executive function, and visuospatial abilities [1]. The disease develops along a continuum—from preclinical stages to mild cognitive impairment (MCI) and eventually to dementia—during which subtle neural alterations precede overt clinical symptoms [2,3]. As global populations age, the societal and economic burden of AD continues to escalate, underscoring the urgent need for reliable tools for early detection and longitudinal monitoring.

Electroencephalography (EEG), as a non-invasive, low-cost, and temporally precise modality, has emerged as a promising candidate for capturing functional brain alterations associated with AD. Over the past decade, a growing body of work has combined EEG with machine learning and deep learning techniques, reporting promising but highly variable classification performance in distinguishing MCI from cognitively normal individuals [4,5] and in detecting AD across multiple disease stages [5,6]. These advances have fueled interest in EEG as a scalable digital biomarker for population-level screening and disease monitoring.

However, despite these promising results, EEG-based AD detection remains characterized by substantial methodological heterogeneity and limited reproducibility. Methodological variability is distributed across the full analytical pipeline. During acquisition, parameters such as channel density and recording duration can limit which biomarkers are estimated reliably. During preprocessing, decisions including reference scheme, high-pass filter cutoff, and artifact rejection criteria can substantially affect the features used for model training. At the validation stage, many high-performing deep learning studies in EEG-based AD diagnosis have used evaluation pipelines with identifiable forms of data leakage—most commonly segment-level data splitting—which can inflate accuracy estimates [7,8]. When such choices are incompletely reported or inconsistently implemented, reported performance metrics become difficult to compare across studies.

Current review articles have largely organized the literature according to feature categories or classification models [9–15]. A recent review by Yuan and Zhao (2025) [16], for instance, provides a comprehensive account of quantitative EEG biomarkers spanning power spectral analysis, functional and effective connectivity, microstate dynamics, and nonlinear complexity measures, illustrating how each feature class captures distinct aspects of AD-related neural disruption. Such feature-oriented reviews offer an important foundation for the field, clarifying what has been measured, how different biomarkers relate to underlying pathophysiology, and where individual feature classes show diagnostic promise. While such approaches provide a valuable organizational framework, they tend to treat each methodological component in isolation.

In this study, to provide an integrated account of how individual methodological problems—such as small sample sizes, heterogeneous preprocessing, and lack of external validation—may interact and compound across the full pipeline, we adopt a **methodological chain framework** that organizes EEG-based AD research as a sequential process spanning acquisition, preprocessing, feature representation, modeling, and validation. Rather than treating these stages as independent modules, this framework makes explicit the methodological dependencies between them: variability and bias introduced upstream may propagate through the pipeline and influence downstream analyses. Importantly, these dependencies are not rigidly deterministic—downstream approaches can sometimes partially compensate for upstream limitations, and the sensitivity of different feature classes to upstream choices varies considerably. The framework is therefore best understood as an organizing principle for identifying where variability originates and how it may compound, rather than as a strictly one-directional system.

Building on this framework, the present review makes three primary contributions. First, we describe inter-stage methodological dependencies that may contribute to reproducibility problems, limited comparability, and performance inflation in EEG-based AD research. Second, we synthesize major sources of methodological variability across representative studies at each stage, evaluating how they differentially affect spectral, connectivity, and complexity features with concrete study-level examples. Third, we provide practical, stage-specific recommendations for study design, evaluation, and reporting,

supported by data leakage sources, a methodological quality map of representative studies, and a minimum reporting checklist (Box 1).

Box 1. Minimum reporting items for EEG-AI studies in Alzheimer’s disease.

1. **Cohort:** Sample size, demographics, diagnostic criteria, inclusion/exclusion criteria, and study setting (single-center or multi-center).
2. **EEG acquisition and quality control:** Device, channel number, montage/reference, sampling frequency, recording paradigm, duration, and quality control procedures including quantitative quality metrics where available.
3. **Preprocessing:** Filtering parameters (type, cutoffs, order), re-referencing scheme, segmentation (epoch length and overlap), artifact handling method and parameters, downsampling procedure, and whether data-dependent parameters were determined within training folds or globally.
4. **Feature representation:** Feature domains, key parameter settings, and feature-selection strategy, including whether selection was nested within cross-validation; distinguish spectral, connectivity, and complexity features as appropriate.
5. **Modeling and validation:** Prediction unit (segment- or subject-level), segment aggregation strategy if applicable, model type, hyperparameter tuning procedure, cross-validation design (nested or non-nested), and leakage prevention measures adopted.
6. **Evaluation and reproducibility:** Class balance, performance metrics beyond accuracy (area under the curve, balanced accuracy, F1 score), uncertainty estimates (confidence intervals or fold-wise variability), external validation status, and code/data availability.

2. The Methodological Chain Framework

EEG-based AD detection involves five sequential stages: data acquisition, preprocessing, feature representation, modeling, and validation (Figure 1). In this review, we use the term **methodological chain** to emphasize that these stages are interdependent—choices at any stage may shape the statistical structure, interpretability, and reliability of downstream analyses, with the strength of this influence varying by feature class and analytical design. Methodological inconsistency at one or more stages can contribute to non-comparable outcomes even when similar models or features are used, motivating the stage-by-stage analysis that follows.

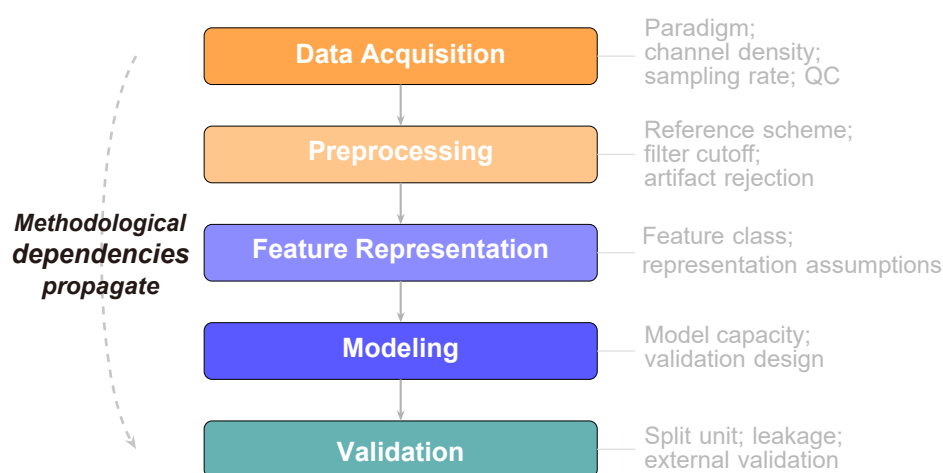


Figure 1. The methodological chain framework for EEG-based AD detection. Each stage shapes the statistical structure available to downstream analyses, and methodological choices introduced upstream may propagate through the pipeline and influence reported performance.

3. Data Acquisition

3.1. Paradigm-Dependent Signal Variability

EEG recordings in AD research are acquired under heterogeneous experimental paradigms, most commonly resting-state and task-based conditions, which probe distinct neurophysiological processes and produce substantially different signal structures [11,17]. An important but underappreciated source of variability is that paradigm heterogeneity is not always visible at the level of study labels: papers nominally reporting “resting-state EEG” may use recording durations ranging from 5 to 30 min, different eyes-open versus eyes-closed conditions, varying instructions regarding mental activity, and different levels of alertness monitoring. These differences can produce variability in alpha power and spectral composition that are unrelated to disease status, yet are rarely controlled for in cross-study comparisons [11,18].

Resting-state EEG is associated with well-established spectral biomarkers of AD: increased delta and theta power and reduced alpha power are consistently reported across studies [18,19]. Its minimal cognitive demands make it particularly suitable for impaired populations and facilitate standardized acquisition across sites [11]. However, resting-state activity is inherently sensitive to vigilance fluctuations, drowsiness, and momentary lapses in wakefulness—sources of latent variability that are difficult to quantify and rarely reported. In AD patients, where vigilance may be compromised, this variability can be particularly pronounced and may be difficult to distinguish from disease-related EEG slowing if uncontrolled.

Task-based paradigms, particularly those using event-related potentials (ERPs), provide process-specific probes of cognitive function. Components such as the P300 and visual evoked potentials have been consistently associated with attentional and memory deficits in AD [15,20,21]. These paradigms can detect subtle dysfunctions not observable in resting-state recordings, but they introduce additional variability from task design, participant compliance, and performance heterogeneity—factors that are especially difficult to control in cognitively impaired populations [17].

Paradigm choice therefore acts as an upstream methodological condition in the methodological chain: it influences which neural processes are reliably observable and which biomarkers can be estimated. Notably, gamma-band activity—which is difficult to interpret reliably in awake resting-state recordings due to myogenic contamination—has been identified as a potentially informative early biomarker in multimodal work [6], illustrating how paradigm selection can influence whether certain disease-relevant signals are reliably accessible or interpretable. This review focuses on resting-state EEG, given its dominant role in the current machine-learning-based AD detection literature; task-based paradigms are discussed where relevant.

3.2. Recording Configuration

Channel density is an important acquisition parameter that is often insufficiently justified relative to the intended analysis. Its impact is feature-dependent in a way that is frequently overlooked: while 19-channel clinical configurations are often adequate for spectral power analyses [11,19]—as illustrated by Siuly et al. [4], who reported high MCI detection accuracy using 19 channels (i.e., Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2)—the same low-density setup may raise validity concerns for dense connectivity-based analyses.

When full-scalp coherence or phase-locking matrices are computed from sparse electrode layouts, spatial undersampling may cause neighboring electrodes to capture signals from overlapping cortical sources; together with volume conduction, this can bias connectivity estimates [22]. In AD research, where disrupted large-scale connectivity is a central

hypothesis, insufficiently justified channel density may weaken the construct validity of connectivity-based biomarkers. For connectivity analyses, 32–64 channel systems are generally preferable [18]; Trinh et al. [5] employed a 32-channel system across six clinical sites and included an independent test set for evaluating connectivity features in a multi-site setting. Source localization typically benefits from high-density recordings (≥ 64 channels), though the required density depends on the inverse model and intended spatial precision. Studies should justify channel density relative to the specific analytical goal, rather than treating it as an incidental hardware detail.

3.3. Sampling Frequency

Sampling frequency influences which temporal structures in the EEG signal can be represented faithfully, but its impact on reproducibility is strongly feature-dependent. For low-frequency spectral measures—delta (0.5–4 Hz), theta (4–8 Hz), and alpha (8–13 Hz) power—128 Hz sampling is generally adequate, since the Nyquist criterion is satisfied with substantial margin [11,19]. Modest sampling rate differences are therefore less likely to be a dominant source of variability for studies focused primarily on low-frequency spectral biomarkers.

The situation differs for phase-based connectivity metrics such as phase-locking value (PLV) and phase lag index (PLI), which depend on precise estimation of instantaneous phase differences. Phase estimates are sensitive to the temporal resolution of the signal: downsampling from higher rates, if performed without appropriate anti-aliasing filtering, can introduce phase distortions that alter connectivity estimates in ways that are difficult to detect without access to the original data [23]. Similarly, nonlinear descriptors such as sample entropy are sensitive to the temporal granularity of the signal. The reproducibility concern is therefore not that a single sampling frequency is universally required, but that undocumented and inconsistent resampling—especially when combined with connectivity-based features—can lead to non-comparable estimates across studies that nominally report the same biomarker [17,24].

3.4. Quality Control

Quality control (QC) should be treated as an important prerequisite for valid neurophysiological inference, yet it remains insufficiently standardized in EEG-based AD research. This issue is particularly relevant in clinical populations, where cognitive impairment, reduced compliance, movement, drowsiness, or medication effects may introduce group-dependent differences in signal quality relative to healthy controls. If QC procedures are not standardized and applied consistently across groups, differences in signal quality may be misinterpreted as neurophysiological group differences—a confound that is particularly relevant to clinical EEG research and remains insufficiently discussed in many methodological reports.

At the device level, EEG signal quality varies substantially across systems in ways that are not always visible to the analyst. Radüntz [25] systematically compared emerging EEG devices and showed that differences in noise floor and electrode characteristics could produce systematic signal differences that may be mistaken for neurophysiological effects if uncontrolled. Griesmaier et al. [26] documented that, even in routine bedside EEG, practical acquisition problems substantially compromised clinical interpretation. In multi-site AD studies—where device harmonization is not always verified or reported—device-specific noise profiles represent an additional source of between-site variability that can confound biomarker estimates [27].

The development of consensus QC benchmarks for EEG, analogous to MRI quality control (MRIQC) in neuroimaging [28], is an important unmet need. Establishing minimum QC

criteria applicable uniformly across sites and devices represents a critical next step toward improving cross-site comparability. Automated quality indices and machine learning-based QC frameworks offer a promising pathway toward scalable, operator-independent assessment [29,30]. The minimum reporting requirements for acquisition QC are summarized in Table 1.

Table 1. Acquisition parameters in EEG-based AD studies: methodological roles, reproducibility implications, and practical recommendations.

Typical Range	Common Failure	Implication for Reproducibility	Recommendation
Paradigm: shapes observable signal content			
REST 5–20 min; ERPs [11,17]	Recording duration, eyes condition, and vigilance state inconsistently reported	Spectral estimates may differ across studies due to uncontrolled conditions, even under identical paradigm labels	Report paradigm, eyes condition, and vigilance monitoring; justify pooling across paradigms
Channel density: shapes spatial resolution of network estimation			
19–25 ch (clinical); 32–64 ch or ≥ 128 ch (research) [11,18]	Low-density recordings used for connectivity analysis without justification	Sparse layouts may introduce spatial aliasing and volume conduction bias; construct validity of connectivity biomarkers may be weakened [22]	Justify channel density relative to intended analysis; 19 ch adequate for spectral features [4]; ≥ 32 ch preferable for connectivity [5]
Sampling rate: shapes frequency range and phase precision			
128–2048 Hz [11,17]	Downsampling procedure unreported	Inconsistent rates reduce comparability, especially for phase-sensitive measures	Report acquisition and analysis rates; apply anti-aliasing filter before downsampling
Recording duration: shapes the amount of artifact-free data available			
2–30 min (REST) [11,14]	Retained artifact-free duration omitted	Limited artifact-free data may reduce the stability of spectral and connectivity estimates	Report retained artifact-free duration; justify adequacy for the intended analysis
Quality control: important prerequisite for valid inference			
Rarely formalized; impedance criteria vary [31]	QC criteria absent; rejection rate unreported	Device noise and poor signal quality may mimic or obscure disease-related differences	Report device, impedance criteria, QC procedures, and proportion of data excluded

3.5. Summary

Across these parameters, a consistent pattern emerges: acquisition choices that may appear incidental can introduce feature-specific variability into downstream analyses, particularly when they are incompletely reported or insufficiently justified. Table 1 summarizes these acquisition parameters and their reproducibility implications.

4. EEG Preprocessing in AD Studies

EEG alterations in AD are typically characterized by spectral slowing, reduced signal complexity, and altered functional connectivity—effects that are often subtle and easily confounded by non-neural artifacts [32,33]. Preprocessing choices can substantially influence downstream analytical results, and Kessler et al. [34] provided systematic evidence that preprocessing decisions can shift decoding performance by margins comparable to or exceeding the effect of model selection—underscoring that preprocessing is not a neutral preparatory step but an integral part of the methodological chain.

4.1. Referencing

EEG signals are typically re-referenced to linked mastoids (A1–A2) or average reference [35,36]. Reference choice is not analytically neutral: its consequences are feature-specific and, crucially, asymmetric in ways that are particularly relevant to AD research. Spectral power features are generally less sensitive to reference scheme, whereas phase-based connectivity measures (PLV, PLI, coherence) are strongly affected. Mastoid reference introduces volume conduction that inflates apparent coherence between electrode pairs sharing a common reference pathway; Ruiz-Gómez et al. [22] demonstrated through computational modeling that this effect can substantially alter functional connectivity estimates. This issue may be particularly relevant for temporal and frontotemporal connectivity analyses [19,37]. Average reference assumes a relatively uniform spatial distribution of scalp activity, an assumption that may be less appropriate in the presence of asymmetric or focal cortical dysfunction.

4.2. Filtering

Band-pass filtering is commonly applied in the range of 0.5–45 Hz [5,38]. Filter settings are particularly relevant for AD studies because EEG-based AD markers are often derived from spectral slowing patterns, including increased delta/theta activity, reduced alpha/beta activity, and altered low-to-high-frequency band ratios [11,19]. High-pass cutoffs above 1 Hz may attenuate low-delta activity and reduce comparability across studies using different filter settings, especially when low-frequency power or ratio-based features are analyzed.

For connectivity analyses using PLV or PLI, filter type and order introduce an additional source of variability: phase distortion near filter edges can generate spurious synchrony or suppress genuine coupling [23]. Filtering choices optimized for spectral analyses may thus be suboptimal for connectivity analyses—a feature-dependent sensitivity that is rarely discussed in study methods sections.

4.3. Artifact Attenuation

Artifact attenuation is commonly addressed using independent component analysis (ICA) with automated component rejection (ICLabel [39] or ADJUST [40]), Artifact Subspace Reconstruction (ASR) [38,41], or wavelet-based denoising [42]. The central tension in AD research is that the biomarkers most directly implicated in the disease—delta and theta slowing—occupy the same frequency ranges most susceptible to over-attenuation. ICA-based automatic classifiers assign components to artifact categories based on spatial and spectral criteria. In principle, disease-related low-frequency components could be affected if automated criteria misclassify atypical neural components as artifacts, particularly when patient data have altered spatial or spectral characteristics. Delorme [43] demonstrated that excessive preprocessing reduces statistical sensitivity in EEG analyses, a finding with direct implications for biomarkers that depend on preserving low-frequency signal content. Regression-based ocular artifact correction methods such as o-CLEAN [44], which apply correction exclusively during detected blink intervals, have been shown to better preserve theta and alpha band content compared to whole-signal correction approaches—a property of particular relevance to AD research where low-frequency spectral features constitute important diagnostic biomarkers.

4.4. Segmentation and Epoch Length

Epoch length involves a tradeoff between frequency resolution and stationarity. Longer epochs provide finer spectral resolution—critical for the delta band, where $1/T_{\text{epoch}}$ must be sufficiently small to resolve sub-band structure—but increase the risk of non-stationarity

within segments. For connectivity features, short epochs yield noisy phase estimates and unreliable connectivity matrices [24]. Epoch count should not be reported as a proxy for sample size; the statistical dependence of epochs from the same subject is a validation concern addressed in Section 6.

4.5. Pipeline Integrity

Data-dependent preprocessing parameters—normalization statistics, artifact rejection thresholds, ICA component criteria—must be determined within training folds at each cross-validation iteration. Global preprocessing applied before data splitting allows test set information to influence training, constituting a form of data leakage whose effect on performance estimates is often underappreciated. Many researchers recognize that post-split normalization introduces leakage, but preprocessing thresholds, component-selection rules, connectivity thresholds, or feature-selection criteria derived from the full dataset can create similar risks. A systematic treatment of preprocessing-related leakage types is provided in Section 6.3. Pellegrini et al. [45] provide a useful framework for evaluating how preprocessing decisions propagate to downstream performance in clinical EEG classification.

4.6. Summary

Table 2 summarizes key preprocessing decisions and their reproducibility implications. Across all steps, a consistent pattern emerges: choices that appear technical and incidental can introduce systematic, feature-specific bias into downstream analyses—and because these choices are rarely reported in sufficient detail, their effects are difficult to detect or correct after the fact. Sharing analysis code and making preprocessing pipelines fully open would substantially reduce the ambiguity that currently limits cross-study comparisons.

Table 2. Preprocessing decisions in EEG-based AD studies: reproducibility implications and practical recommendations.

Common Practice	Implication for Reproducibility	Recommendation
Re-referencing: choice of reference scheme affects connectivity estimates more than spectral measures		
Linked mastoid or average reference [35,36]	PLV, PLI, and coherence are strongly sensitive to reference choice; mastoid reference introduces volume conduction that can alter inter-electrode coupling [22]; spectral power is generally less affected	Use a consistent reference scheme; report as a primary preprocessing parameter
High-pass filtering: cutoff choice affects δ -band biomarkers and phase-sensitive measures		
Variable high-pass cutoffs, typically 0.5–2 Hz	Cutoffs above 1 Hz progressively attenuate low- δ activity [11,19]; filter type and order introduce phase distortion that can affect PLV and PLI estimates [23]	Use $HP \leq 0.5$ Hz when δ -band is a primary outcome; report filter type, order, and exact cutoffs
Artifact attenuation: rejection criteria influence the balance between biomarker preservation and artifact removal		
ICA + ICLabel [39]/ADJUST [40]; ASR [41]; o-CLEAN [44]	Overly aggressive rejection may misclassify disease-related δ/θ components as artifacts, reducing statistical sensitivity [43]; pipelines with different rejection thresholds are difficult to compare	Apply predefined, conservative rejection criteria; report components or segments removed per subject

Table 2. Cont.

Common Practice	Implication for Reproducibility	Recommendation
Epoch length and segmentation: determines frequency resolution and statistical independence of samples		
Fixed-length epochs, commonly 1–30 s	Frequency resolution scales as $1/T_{\text{epoch}}$; short epochs reduce δ -band resolution and yield unstable connectivity matrices [24]; epochs from the same subject are statistically dependent and should not be treated as independent samples	Report epoch length, overlap, and count; avoid treating epoch count as subject-level sample size
Global parameter selection: fitting parameters before splitting allows test set information to influence training		
Normalization, scaling, thresholds, or feature selection applied before data splitting	Data-dependent parameters estimated from the full dataset incorporate test set information into preprocessing or feature selection, inflating performance estimates [7,46]	Fit all data-dependent parameters within training folds; document which steps were global or within-fold

5. Feature Representation and Modeling: An Interdependent Design Space

Feature representation constitutes the central interface between preprocessed EEG signals and downstream modeling. Rather than being a neutral transformation, it shapes the statistical structure of the data and conditions which modeling strategies are appropriate. Apparent differences in model performance across studies may therefore reflect differences in representation, upstream preprocessing, sample size, and validation design as much as differences in model architecture.

5.1. Low-Dimensional Representations and Conventional Machine Learning

Spectral, statistical, and complexity-based features quantify summary properties of EEG signals and remain among the most established descriptors in AD research. Common measures include band power, band ratios, peak alpha frequency (PAF), entropy-based metrics, and Hjorth parameters [11,19,33]. These representations are closely aligned with widely reported EEG alterations in AD, including increased delta and theta activity, reduced alpha and beta power, and slowing of the dominant alpha rhythm [18,19]. Thus, low-dimensional features should not be viewed merely as a pragmatic choice for small-sample settings; they can also be a principled choice when the biological question concerns global spectral reorganization.

From a methodological chain perspective, compact feature spaces offer an additional advantage: they are generally less sensitive than connectivity or high-dimensional tensor representations to moderate upstream preprocessing variation. This combination of neurophysiological grounding, relative preprocessing robustness, and compact dimensionality makes low-dimensional features practical for small-cohort AD studies. As an illustrative example, Siuly et al. [4] reported high MCI detection accuracy using AR and permutation entropy features with subject-wise cross-validation on 27 subjects, although the small sample size means that the reported accuracy should be interpreted cautiously. The primary limitation is that such features capture mainly marginal signal properties and provide only a partial description of distributed network dysfunction or dynamic reorganization [10,12].

5.2. High-Dimensional Structured Representations and Deep Learning

Time-frequency representations, wavelet decompositions, and multichannel tensor structures capture nonstationary and multi-scale dynamics that low-dimensional sum-

maries cannot represent [33,47,48]. Their hierarchical structure makes them conceptually well-suited to convolutional neural networks (CNNs) [49,50]. However, these representations also increase feature dimensionality and make model performance more sensitive to sample size, regularization, and validation design.

A frequently misunderstood point is that transforming EEG segments into large numbers of input samples does not increase subject diversity. Generating many epochs per subject increases apparent sample size, but the statistical dependence among epochs from the same individual means that the effective number of independent training examples remains constrained by the number of subjects [7,8]. A model trained on thousands of epochs from a small number of subjects has not seen thousands of independent examples of AD-related EEG; it has seen repeated samples from the same individuals. As a result, CNN-based models applied to small EEG-AD datasets may be prone to overfitting and especially sensitive to validation design. Developing explainability methods tailored to these high-capacity models would help clinicians assess whether model decisions reflect neurophysiologically valid features rather than dataset-specific or subject-level patterns.

5.3. Connectivity-Based Representations

Connectivity-based representations—including coherence, PLV, PLI, wPLI, and amplitude envelope correlation (AEC)—are conceptually motivated by the view of AD as a disorder of disrupted functional integration [10,37]. The disconnection syndrome hypothesis suggests that AD pathology is associated with disrupted long-range cortico-cortical communication, providing a neurophysiological rationale for connectivity-based biomarkers [37]. This conceptual alignment is one reason connectivity features have attracted substantial attention in EEG-based AD research.

In practice, however, connectivity estimates are among the more preprocessing-sensitive features in the EEG analysis pipeline. Reference scheme, filter type and order, epoch length, and sampling rate can all substantially affect connectivity estimates [11,22,24,51]. The consequence for reproducibility is that two studies using the same connectivity measure on similar populations may produce non-comparable results if upstream preprocessing choices differ or are incompletely reported. Trinh et al. [5] provided a concrete illustration of cross-site generalization challenges: using PLI-based connectivity across 180 participants from six hospitals, MCI vs. HC classification reached 90% leave-one-participant-out cross-validation (LOPO-CV) accuracy internally but dropped to 45% on the independent test set. This gap illustrates the difficulty of cross-site generalization and the limits of relying solely on internal validation.

5.4. Graph-Based Representations

Graph-based representations impose an additional abstraction layer by transforming connectivity matrices into topological network metrics, including clustering coefficient, global and local efficiency, modularity, and small-worldness [37,52]. They enable graph neural network modeling [53,54] and are theoretically motivated by evidence of altered network organization in AD [37].

However, graph metrics inherit the uncertainty of connectivity estimation and introduce additional sensitivity to graph construction choices, particularly thresholding strategy, binarization, and density threshold [51]. Small-world metrics can be particularly sensitive because small-worldness depends partly on graph density as well as the underlying connectivity structure. Different thresholding choices may therefore lead to different conclusions about network organization. Reported findings of abnormal small-world properties should be interpreted in light of the thresholding strategy used.

5.5. End-to-End Representation Learning Under Data Constraints

End-to-end approaches learn feature representations directly from minimally processed EEG, without predefined feature engineering [48,49]. In principle, this offers flexibility to discover disease-relevant patterns. In practice, the effectiveness of end-to-end learning in EEG-based AD research is strongly constrained by the size and heterogeneity of available datasets. When subject counts are small, high-capacity models may capture demographic confounds, site-specific noise profiles, or subject-identity features rather than disease-related signal [7,8,46].

This problem is amplified when end-to-end models are evaluated using segment-level splitting. Raw or minimally processed EEG segments from the same subject are highly similar, allowing models to exploit subject-specific patterns when segments from the same individual appear in both training and test sets. Such performance may decrease substantially under subject-level evaluation, underscoring the importance of validation design for high-capacity representation learning.

5.6. Synthesis: Representation, Validation, and the Methodological Chain

Across these representation types, the choice of representation influences what information is available to the model and how vulnerable the analysis may be to preprocessing variability, sample-size limitations, and validation errors. Low-dimensional subject-level features are generally more robust to upstream preprocessing variation and less vulnerable to segment-level leakage. High-dimensional, segment-based, connectivity-based, and graph-based representations can provide richer information, but they also amplify the consequences of improper cross-validation, segment-level splitting, global preprocessing, and insufficient sample size.

The combination of high-dimensional representation and segment-level validation—which has been reported across a substantial portion of the EEG-AD literature [7]—can create a compounding risk: the representation increases the opportunity to exploit subject-specific structure, while the validation design may fail to prevent such leakage. Table 3 summarizes these representation types and their associated methodological constraints.

Table 3. EEG representation types in AD studies: preprocessing sensitivity, sample-size sensitivity, and primary leakage risk. Sample-size figures are indicative only and depend on class balance, validation design, and feature dimensionality.

Preprocessing Sensitivity	Sample-Size Sensitivity Indicative	Typical Models	Primary Leakage Risk
Low-dimensional spectral/complexity [11,19,33]			
EEG slowing ($\uparrow \delta/\theta$, $\downarrow \alpha/\beta$); band power, PAF, entropy (ApEn, SampEn), Hjorth			
Lower: robust to moderate reference and filter variation	Lower (~ 30 – 50); compact feature space relatively stable in small cohorts	SVM, RF, kNN, LR	Global normalization before split
High-dimensional structured [33,50]			
Multi-scale abnormalities; spectrograms, wavelet coefficients, multichannel tensors			
Moderate: filter cutoff shapes spectrogram structure	High; segment inflation does not substitute for subject diversity	CNN, CNN-LSTM	Segment-level split; global standardization
Connectivity matrices [10,12,55]			
Disrupted long-range cortico-cortical communication; coherence, PLV, PLI, wPLI, AEC			
High: reference, filter phase, and epoch length alter estimates [22]	Moderate–high (~ 50 – 100); depends on epoch number and metric	SVM, RF, CNN (matrix)	Reference/filter inconsistency; global FC thresholding

Table 3. *Cont.*

Preprocessing Sensitivity	Sample-Size Sensitivity Indicative	Typical Models	Primary Leakage Risk
Graph-based topology [37,52,54]			
Altered network integration; clustering, efficiency, modularity, small-worldness			
High: inherits connectivity uncertainty; sensitive to graph construction choices [51]	High (~100 or more); depends on thresholding strategy	Graph kernels, GNN	Post hoc threshold selection; variable graph density across studies
End-to-end learned [49,53,56]			
Implicit hierarchical features; raw or minimally processed EEG			
Moderate–high: upstream filtering conditions learnable content	Very high; rarely satisfied by current EEG–AD datasets	CNN, RNN, transformers	Segment-level split; subject-identity leakage via oversampling

6. Validation and Reliability in EEG-Based AD Studies

Within the methodological chain, validation represents the final stage at which the reliability and generalizability of the entire analytical pipeline are assessed. Validation is not merely a neutral evaluation step; it strongly affects the extent to which reported performance can be interpreted as evidence of generalizable neurophysiological patterns rather than dataset-specific or methodological artifacts. Crucially, the consequences of improper validation are not uniform across representation types: high-dimensional and segment-based representations can amplify the effects of validation errors more strongly than low-dimensional subject-level features—a dependency that makes validation design inseparable from representation choice.

6.1. Subject-Level Versus Segment-Level Evaluation

A fundamental issue concerns the unit of analysis used for model evaluation. When segments from the same subject are included in both training and testing sets, the evaluation may no longer be independent. This segment-level splitting can introduce a form of data leakage because intra-subject similarity may allow models to exploit subject-specific patterns rather than disease-related features [7,8,46]. EEG data often allow the generation of large numbers of epochs from a small number of subjects, which can lead researchers to treat epoch count as sample size—a practice that makes segment-level splitting particularly prevalent in the literature. Young et al. [7] identified this as the single most common source of inflated performance in deep learning AD studies. By contrast, subject-level splitting keeps all segments from a given subject within the same fold. For example, Siuly et al. [4] implemented subject-wise 10-fold CV, representing a more appropriate evaluation design for small-sample studies.

6.2. Cross-Validation and Nested Design

Improper use of cross-validation can introduce optimistic bias at multiple levels. Most visibly, performing preprocessing steps such as normalization or feature selection on the full dataset prior to splitting allows test set information to influence training [7,46]. Less visibly, when feature selection is performed in the outer loop of cross-validation rather than nested within each fold, selected features have already been exposed to test set label information. The resulting bias grows with feature dimensionality: in high-dimensional connectivity or time-frequency representations, where hundreds or thousands of candidate features are available, unnested feature selection can produce substantially inflated accuracy estimates even when data splitting is otherwise correctly implemented.

When hyperparameter tuning or feature selection is involved, nested cross-validation is preferable for obtaining a less biased estimate of model performance [7,57].

6.3. Data Leakage and Pipeline Integrity

Data leakage can occur at multiple stages of the methodological chain, not only during model training but also during preprocessing and feature extraction [7,46]. Table 4 provides leakage types organized by stage, mechanism, and remediation strategy. Understanding the distinct origin of each leakage type is important for reducing leakage risk in analytical pipelines, as different sources require different preventive measures.

Table 4. Data leakage sources in EEG-based AD studies: mechanism and remediation.

Leakage Type (Stage)	Mechanism	Remediation
Label-informed preprocessing (Preproc)	Artifact rejection or cleaning criteria differ by diagnostic group; preprocessing implicitly encodes class labels	Apply identical, predefined preprocessing blinded to diagnosis
Shared preprocessing parameters (Preproc)	Artifact thresholds, cleaning rules, or preprocessing parameters tuned using the full dataset before splitting	Tune data-dependent parameters within training folds; apply fixed rules to test data
Global normalization (Preproc)	Mean/SD computed from full dataset, including test subjects; test set distribution informs training normalization	Fit normalization parameters within each training fold
Pre-split oversampling (Preproc)	SMOTE applied to full dataset; synthetic samples constructed using test-subject feature distributions	Apply oversampling exclusively within training folds
Global feature selection (Feature)	Feature ranking based on full dataset statistics; test set labels implicitly inform feature selection	Nest feature selection inside CV folds
Overlapping/adjacent epochs (Segmentation)	Temporally adjacent epochs from the same session split across folds; temporal autocorrelation exploited	Use non-overlapping epochs or temporal blocking in CV
Segment-level splitting (Validation)	Epochs from same subject distributed across train and test sets; intra-subject similarity exploited instead of disease signal	Split at subject level; report N subjects as effective sample size

6.4. External Validation and Generalizability

External validation using an independent cohort remains the most informative approach for assessing clinical applicability, yet it is still relatively uncommon in EEG-based AD studies [7,11,46]. The need for external validation is particularly acute in EEG-based AD research because EEG signals are highly sensitive to device characteristics, acquisition site, and operator practice. Internal cross-validation, even when correctly implemented at the subject level, may not fully capture how much performance depends on disease-related neurophysiology versus site-specific or device-specific factors. These factors may widen the gap between internal and external performance, particularly when acquisition protocols are not standardized across sites.

Trinh et al. [5] provide an instructive example: although their PLI-based MCI vs. HC model reached 90% LOPO-CV accuracy on the training cohort, performance dropped to 45% on an independent multi-site test set. This gap should not be attributed to a single

methodological factor; rather, it illustrates the difficulty of cross-site generalization and the need for external validation before making strong claims of clinical robustness.

The use of shared public datasets, such as OpenNeuro ds004504 [58], can support more transparent comparisons across methods, particularly when subject-level splits, preprocessing pipelines, and evaluation protocols are clearly specified. However, analyses based on a single public dataset should not be regarded as a substitute for external validation, because model development may still be influenced by dataset-specific characteristics. The absence of external validation should therefore be explicitly acknowledged in all studies. Looking forward, privacy-preserving multi-site model training represents a promising path toward addressing not only the sample-size constraints but also the acquisition heterogeneity—including device variability and protocol inconsistency—that characterize this field.

6.5. Dataset Size and Statistical Reliability

Many EEG-based AD studies involve fewer than 100 subjects [11,14], and this limitation has structural origins. EEG data collection in cognitively impaired populations is resource-intensive: patient compliance is variable, recording sessions require clinical supervision, and data loss from artifact rejection is often substantial. Longitudinal data—which would enable within-subject tracking of disease progression—is particularly scarce. These constraints appear to reflect broader field-level challenges rather than limitations of individual studies alone, and are difficult to address without coordinated multi-site data collection efforts.

Within this context, high classification accuracies—particularly when combined with small sample sizes, high-dimensional representations, and non-rigorous validation—should be interpreted with substantial caution [7,46]. Reporting uncertainty measures, such as confidence intervals or fold-wise variability, is important for assessing statistical reliability and for enabling meaningful comparison across studies. Benchmark datasets with pre-specified subject-level train-test splits would enable fairer cross-study comparison and facilitate more rigorous external validation.

6.6. Methodological Quality Map of Representative Studies

Table 5 provides a structured overview of methodological features across representative EEG-based AD classification studies at each stage of the analytical pipeline, illustrating the range of approaches and their associated design choices.

Table 5. Methodological features of representative EEG-based AD classification studies, illustrating a range of methodological approaches.

Study	Dataset/N	Acquisition	Preprocessing	Segmentation	Feature	Model	Validation	External Val.
Trinh et al. 2023 [5]	Proprietary; 6-site; 150 (50 AD/50 MCI/50 HC)	32-ch; 6 sites; 500 Hz; eyes-open REST	Ref: right mastoid; filtering NR; ICA + ADJUST; within-fold norm NR	Non-overlapping 3-s epochs	PLI (5 bands)	LDA + SFS; Acc = 82.50% (train)/75.00% (test)	Subject LOPO-CV; independent test N = 30	Multi-site test set
Siuly et al. 2020 [4]	Proprietary; 27 (11 MCI/16 HC)	19-ch; 256 Hz; eyes-closed REST; >30 min	SWT denoising (0.5–32 Hz); ref NR; within-fold norm NR	Non-overlapping 2-s epochs	AR (4th order) + PE + histogram (PAA)	ELM/SVM/KNN; best: ELM Acc = 98.78%, AUC = 0.98	Subject-wise 10-fold CV + LOPO-CV	No external set
Nour et al. 2024 [59]	Public; Florida State Univ. datasets; 140 (104 AD/36 HC)	19-ch; 128 Hz; 8-s segments; eyes condition partly inherited/not fully harmonized	Dataset-level artifact cleaning; Butterworth 0.5–45 Hz; band filtering into delta–gamma; ref and within-fold norm NR	1-s epochs; 128 × 19 arrays; total 1120 epochs	Raw /filtered 2D EEG arrays; no handcrafted feature extraction; band-specific inputs	DEL ensemble of five 2D-CNNs; weighted averaging; Acc = 97.9%	epoch-level 5-fold CV; subject-wise split NR	No external set
Huggins et al. 2021 [49]	Proprietary; 141 age-matched (52 AD/37 MCI/52 HA)	21-ch; 200 Hz; eyes-closed REST; ~10 min (Braintech 3.0)	1–60 Hz band-pass FIR; notch 21 & 42 Hz; ICA + MARA; within-fold norm NR	Non-overlapping 5-s epochs	CWT scalograms (Morse wavelet) → tiled topographic RGB images	AlexNet-based CNN; Acc = 98.9%	epoch-level 10-fold CV; subject-wise split NR	No external set
The following two studies share the same acquisition (OpenNeuro ds004504 [58]): 19-ch, 500 Hz, eyes-closed REST, ~12–14 min.								
Shamsi 2025 [46]	59 post-QC (31 AD/28 HC)	—	Resample 128 Hz; 0.5–45 Hz band-pass; notch 50 Hz; avg ref; within-fold z-norm	8-s epochs, 4-s overlap	WST + ROI pooling	Regularized logistic ensemble; AUC = 0.930	Subject-wise 5-fold GroupKFold	No external set
Zheng et al. 2023 [33]	65 (36 AD/29 HC)	—	A1–A2 ref; 0.5–45 Hz Butterworth; ICA; ASR; within-fold norm NR	4-s epochs, 50% overlap	Time-domain stats + RBP (5 bands) + entropy + graph sync metrics	Decision Tree/RF/SVM; best: RF Acc = 95.86%	Subject LOPO-CV	No external set

NR = not reported in the original publication. external val. = external validation. Assessments are based on published information only. AUC = area under the curve.

7. Limitations and Future Directions

Two main limitations of the present review should be acknowledged. This is a narrative rather than a systematic review: representative studies were selected to illustrate key methodological patterns rather than to provide exhaustive coverage of the literature. The scope is further limited to the EEG signal-processing and machine-learning pipeline; clinical factors such as vascular biomarkers, genetic markers, and cognitive measures are not systematically covered. The following directions point toward important extensions and open problems identified through the methodological chain analysis.

Standardized acquisition and QC protocols. As discussed in Section 3, the absence of standardized acquisition and QC criteria is one of the primary sources of between-study variability in EEG-based AD research. Analogous to MRIQC in neuroimaging [28], the development of community-agreed QC benchmarks for EEG-based AD studies would substantially improve signal reliability and result comparability. Building on existing expert recommendations [18], future work should establish minimum QC criteria—including quantitative signal quality indices—applicable uniformly across sites and devices.

Reproducible preprocessing pipelines. As shown in Section 4 and Table 2, preprocessing decisions—including reference scheme, filter cutoff, and artifact rejection aggressiveness—can introduce systematic, feature-specific bias into downstream analyses, yet these choices are rarely reported in sufficient detail to enable replication. The public release of fully specified, code-level preprocessing pipelines alongside published studies would be an important step toward reproducibility. Sensitivity analyses examining the effect of key preprocessing choices on downstream results should become more common practice, particularly for connectivity-based biomarkers [34,45].

Rigorous validation infrastructure. The prevalence of segment-level splitting and the scarcity of external validation identified in Section 6 highlight the need for more rigorous validation infrastructure. The field would benefit from benchmark datasets with pre-specified subject-level train-test splits, enabling fairer cross-study comparison. Public datasets such as that released by Miltiadous et al. [58] represent a step in this direction, though pre-specified splits remain uncommon. Approaches that enable multi-site model training while preserving data privacy represent a promising path toward larger effective cohort sizes.

Interpretability and clinical translation. The validation concerns identified in Section 6—particularly the difficulty of distinguishing genuine disease signal from subject-specific or site-specific patterns—underscore the importance of model interpretability for clinical translation. Future work should develop explainability methods tailored to EEG-based AD models, enabling clinicians to understand which features contribute to model decisions and to assess whether those features have neurophysiological validity.

Longitudinal designs. The majority of EEG-based AD studies reviewed here adopt cross-sectional designs, which cannot capture how biomarker patterns evolve as the disease progresses [11,14]. Studies examining longitudinal EEG trajectories—tracking individuals from MCI through conversion to AD—represent a particularly valuable but underrepresented design type, as they provide naturalistic evidence of biomarker sensitivity to disease progression that cross-sectional studies cannot supply.

Multimodal integration and clinical biomarkers. Clinical factors—including vascular biomarkers, demographics, lifestyle variables, comorbidities, genetic markers such as apolipoprotein E genotype, and cognitive measures—are essential components of a complete AD biomarker framework but lie beyond the scope of this methodologically focused work [2,3]. The integration of EEG with these factors represents a particularly important direction for future clinical translation [6]. Recent multimodal work [6] demonstrated that EEG-derived features were among the most sensitive indicators of early cognitive decline,

suggesting that paradigm selection and feature design choices can strongly influence which aspects of disease progression EEG can capture. Extension of the methodological chain concept to encompass multimodal pipelines will require the same principles of stage-wise consistency and leakage prevention applied with even greater complexity.

8. Conclusions

The central argument of this review is that reliability in EEG-based AD detection is not shaped by model choice alone, but by the integrity of the entire methodological chain. Acquisition shapes signal observability, preprocessing influences feature stability in ways that are feature class-dependent, representation substantially conditions model behavior, and validation is central to judging whether observed performance is clinically meaningful. High classification accuracy, in isolation, is an insufficient indicator of biomarker validity; it should be interpreted in the context of the full analytical pipeline. More broadly, similar reproducibility concerns have been documented in computer-aided diagnosis research for other medical imaging domains, such as lung disease detection using chest X-ray radiographs [60], suggesting these challenges extend beyond EEG-based neurological diagnosis. Accordingly, future progress may benefit as much from improved alignment, transparency, and rigor across all stages of the pipeline as from advances in model architecture alone.

Funding: This research was funded by Japan Society for the Promotion of Science (JSPS KAKENHI), grant number 26K15682 and 26K21664.

Data Availability Statement: No new data were created or analyzed in this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Alzheimer's Association. 2025 Alzheimer's disease facts and figures. *Alzheimers Dement.* **2025**, *21*, e70235.
2. Blennow, K.; de Leon, M.J.; Zetterberg, H. Alzheimer's disease. *Lancet* **2006**, *368*, 387–403. [[CrossRef](#)]
3. Jack, C.R., Jr.; Andrews, J.S.; Beach, T.G.; Buracchio, T.; Dunn, B.; Graf, A.; Hansson, O.; Ho, C.; Jagust, W.; McDade, E.; et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimers Dement.* **2024**, *20*, 5143–5169. [[CrossRef](#)]
4. Siuly, S.; Alçin, Ö.F.; Kabir, E.; Şengür, A.; Wang, H.; Zhang, Y.; Whittaker, F. A new framework for automatic detection of patients with mild cognitive impairment using resting-state EEG signals. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 1966–1976. [[CrossRef](#)]
5. Trinh, T.T.; Liu, Y.H.; Wu, C.T.; Peng, W.H.; Hou, C.L.; Weng, C.H.; Lee, C.Y. PLI-based connectivity in resting-EEG is a robust and generalizable feature for detecting MCI and AD: A validation on a diverse multisite clinical dataset. In *Proceedings of the 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Sydney, Australia, 24–27 July 2023*; IEEE: New York, NY, USA, 2023; pp. 1–6.
6. Lin, Y.; Shi, X.; Mu, J.; Ren, H.; Jiang, X.; Zhu, L.; Cai, X.; Lian, C.; Pei, Z.; Zhang, Y.; et al. Uncovering stage-specific neural and molecular progression in Alzheimer's disease: Implications for early screening. *Alzheimers Dement.* **2025**, *21*, e70182. [[CrossRef](#)]
7. Young, V.M.; Gates, S.; Garcia, L.Y.; Salardini, A. Data Leakage in Deep Learning for Alzheimer's Disease Diagnosis: A Scoping Review of Methodological Rigor and Performance Inflation. *Diagnostics* **2025**, *15*, 2348. [[CrossRef](#)]
8. Brookshire, G.; Kasper, J.; Blauch, N.M.; Wu, Y.C.; Glatt, R.; Merrill, D.A.; Gerrol, S.; Yoder, K.J.; Quirk, C.; Lucero, C. Data leakage in deep learning studies of translational EEG. *Front. Neurosci.* **2024**, *18*, 1373515. [[CrossRef](#)] [[PubMed](#)]
9. Vecchio, F.; Babiloni, C.; Lizio, R.; Fallani, F.D.V.; Blinowska, K.; Verrienti, G.; Frisoni, G.; Rossini, P.M. Resting state cortical EEG rhythms in Alzheimer's disease: Toward EEG markers for clinical applications: A review. *Suppl. Clin. Neurophysiol.* **2013**, *62*, 223–236. [[PubMed](#)]
10. Adebisi, A.T.; Veluvolu, K.C. Brain network analysis for the discrimination of dementia disorders using electrophysiology signals: A systematic review. *Front. Aging Neurosci.* **2023**, *15*, 1039496. [[CrossRef](#)] [[PubMed](#)]
11. Cassani, R.; Estarellas, M.; San-Martin, R.; Fraga, F.J.; Falk, T.H. Systematic review on resting-state EEG for Alzheimer's disease diagnosis and progression assessment. *Dis. Markers* **2018**, *2018*, 5174815. [[CrossRef](#)]
12. Paitel, E.R.; Otteman, C.B.; Polking, M.C.; Licht, H.J.; Nielson, K.A. Functional and effective EEG connectivity patterns in Alzheimer's disease and mild cognitive impairment: A systematic review. *Front. Aging Neurosci.* **2025**, *17*, 1496235. [[CrossRef](#)]

13. Ouchani, M.; Gharibzadeh, S.; Jamshidi, M.; Amini, M. A review of methods of diagnosis and complexity analysis of Alzheimer's disease using EEG signals. *BioMed Res. Int.* **2021**, *2021*, 5425569. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Perez-Valero, E.; Lopez-Gordo, M.A.; Morillas, C.; Pelayo, F.; Vaquero-Blasco, M.A. A review of automated techniques for assisting the early detection of Alzheimer's disease with a focus on EEG. *J. Alzheimer's Dis.* **2021**, *80*, 1363–1376. [\[CrossRef\]](#)
15. Costanzo, M.; Cutrona, C.; Leodori, G.; Malimpensa, L.; D'antonio, F.; Conte, A.; Belvisi, D. Exploring easily accessible neurophysiological biomarkers for predicting Alzheimer's disease progression: A systematic review. *Alzheimers Res. Ther.* **2024**, *16*, 244. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Yuan, Y.; Zhao, Y. The role of quantitative EEG biomarkers in Alzheimer's disease and mild cognitive impairment: Applications and insights. *Front. Aging Neurosci.* **2025**, *17*, 1522552. [\[CrossRef\]](#)
17. Dauwels, J.; Vialatte, F.; Cichocki, A. Diagnosis of Alzheimer's disease from EEG signals: Where are we standing? *Curr. Alzheimer Res.* **2010**, *7*, 487–505. [\[CrossRef\]](#)
18. Babiloni, C.; Arakaki, X.; Azami, H.; Bennys, K.; Blinowska, K.; Bonanni, L.; Bujan, A.; Carrillo, M.C.; Cichocki, A.; de Frutos-Lucas, J.; et al. Measures of resting state EEG rhythms for clinical trials in Alzheimer's disease: Recommendations of an expert panel. *Alzheimers Dement.* **2021**, *17*, 1528–1553. [\[CrossRef\]](#)
19. Jeong, J. EEG dynamics in patients with Alzheimer's disease. *Clin. Neurophysiol.* **2004**, *115*, 1490–1505. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Yamasaki, T.; Horie, S.; Ohyagi, Y.; Tanaka, E.; Nakamura, N.; Goto, Y.; Kanba, S.; Kira, J.i.; Tobimatsu, S. A potential VEP biomarker for mild cognitive impairment: Evidence from selective visual deficit of higher-level dorsal pathway. *J. Alzheimer's Dis.* **2016**, *53*, 661–676. [\[CrossRef\]](#)
21. Wu, S.Z.; Masurkar, A.V.; Balcer, L.J. Afferent and efferent visual markers of Alzheimer's disease: A review and update in early stage disease. *Front. Aging Neurosci.* **2020**, *12*, 572337. [\[CrossRef\]](#)
22. Ruiz-Gómez, S.J.; Hornero, R.; Poza, J.; Maturana-Candelas, A.; Pinto, N.; Gomez, C. Computational modeling of the effects of EEG volume conduction on functional connectivity metrics. Application to Alzheimer's disease continuum. *J. Neural Eng.* **2019**, *16*, 066019. [\[CrossRef\]](#)
23. Mehra, C.; Beyh, A.; Laiou, P.; Garces, P.; Jones, E.J.; Mason, L.; Buitelaar, J.; Johnson, M.H.; Murphy, D.; Loth, E.; et al. Zero-phase-delay synchrony between interacting neural populations: implications for functional connectivity-derived biomarkers. *Imaging Neurosci.* **2025**, *3*, IMAG-a. [\[CrossRef\]](#)
24. Briels, C.T.; Schoonhoven, D.N.; Stam, C.J.; de Waal, H.; Scheltens, P.; Gouw, A.A. Reproducibility of EEG functional connectivity in Alzheimer's disease. *Alzheimers Res. Ther.* **2020**, *12*, 68. [\[CrossRef\]](#)
25. Radüntz, T. Signal quality evaluation of emerging EEG devices. *Front. Physiol.* **2018**, *9*, 98. [\[CrossRef\]](#)
26. Griesmaier, E.; Neubauer, V.; Ralser, E.; Trawöger, R.; Kiechl-Kohlendorfer, U.; Keller, M. Need for quality control for aEEG monitoring of the preterm infant: A 2-year experience. *Acta Paediatr.* **2011**, *100*, 1079–1083. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Webb, S.J.; Shic, F.; Murias, M.; Sugar, C.A.; Naples, A.J.; Barney, E.; Borland, H.; Hellemann, G.; Johnson, S.; Kim, M.; et al. Biomarker acquisition and quality control for multi-site studies: The autism biomarkers consortium for clinical trials. *Front. Integr. Neurosci.* **2020**, *13*, 71. [\[CrossRef\]](#)
28. Esteban, O.; Birman, D.; Schaer, M.; Koyejo, O.O.; Poldrack, R.A.; Gorgolewski, K.J. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE* **2017**, *12*, e0184661. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Hu, B.; Peng, H.; Zhao, Q.; Hu, B.; Majoe, D.; Zheng, F.; Moore, P. Signal quality assessment model for wearable EEG sensor on prediction of mental stress. *IEEE Trans. Nanobiosci.* **2015**, *14*, 553–561.
30. Fickling, S.D.; Liu, C.C.; D'Arcy, R.C.; Hajra, S.G.; Song, X. Good data? The EEG quality index for automated assessment of signal quality. In *Proceedings of the 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 17–19 October 2019*; IEEE: New York, NY, USA, 2019; pp. 219–229.
31. Ferree, T.C.; Luu, P.; Russell, G.S.; Tucker, D.M. Scalp electrode impedance, infection risk, and EEG data quality. *Clin. Neurophysiol.* **2001**, *112*, 536–544. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Al-Qazzaz, N.K.; Ali, S.H.B.M.; Ahmad, S.A.; Chellappan, K.; Islam, M.S.; Escudero, J. Role of EEG as biomarker in the early detection and classification of dementia. *Sci. World J.* **2014**, *2014*, 906038. [\[CrossRef\]](#)
33. Zheng, X.; Wang, B.; Liu, H.; Wu, W.; Sun, J.; Fang, W.; Jiang, R.; Hu, Y.; Jin, C.; Wei, X.; et al. Diagnosis of Alzheimer's disease via resting-state EEG: Integration of spectrum, complexity, and synchronization signal features. *Front. Aging Neurosci.* **2023**, *15*, 1288295. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Kessler, R.; Enge, A.; Skeide, M.A. How EEG preprocessing shapes decoding performance. *Commun. Biol.* **2025**, *8*, 1039. [\[CrossRef\]](#)
35. Fraga, F.J.; Falk, T.H.; Kanda, P.A.; Anghinah, R. Characterizing Alzheimer's disease severity via resting-awake EEG amplitude modulation analysis. *PLoS ONE* **2013**, *8*, e72240. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Yang, X.; Fan, Z.; Li, Z.; Zhou, J. Resting-state EEG microstate features for Alzheimer's disease classification. *PLoS ONE* **2024**, *19*, e0311958. [\[CrossRef\]](#)
37. De Haan, W.; Pijnenburg, Y.A.; Strijers, R.L.; van der Made, Y.; van der Flier, W.M.; Scheltens, P.; Stam, C.J. Functional neural network analysis in frontotemporal dementia and Alzheimer's disease using EEG and graph theory. *BMC Neurosci.* **2009**, *10*, 101. [\[CrossRef\]](#)

38. Rostamikia, M.; Sarbaz, Y.; Makouei, S. EEG-based classification of Alzheimer's disease and frontotemporal dementia: A comprehensive analysis of discriminative features. *Cogn. Neurodyn.* **2024**, *18*, 3447–3462. [[CrossRef](#)]
39. Pion-Tonachini, L.; Kreutz-Delgado, K.; Makeig, S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage* **2019**, *198*, 181–197. [[CrossRef](#)]
40. Mognon, A.; Jovicich, J.; Bruzzone, L.; Buiatti, M. ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology* **2011**, *48*, 229–240. [[CrossRef](#)]
41. Chang, C.Y.; Hsu, S.H.; Pion-Tonachini, L.; Jung, T.P. Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings. *IEEE Trans. Biomed. Eng.* **2019**, *67*, 1114–1121. [[CrossRef](#)]
42. Grobbelaar, M.; Phadikar, S.; Ghaderpour, E.; Struck, A.F.; Sinha, N.; Ghosh, R.; Ahmed, M.Z.I. A survey on denoising techniques of electroencephalogram signals using wavelet transform. *Signals* **2022**, *3*, 577–586. [[CrossRef](#)]
43. Delorme, A. EEG is better left alone. *Sci. Rep.* **2023**, *13*, 2372. [[CrossRef](#)] [[PubMed](#)]
44. Ronca, V.; Flumeri, G.D.; Giorgi, A.; Vozzi, A.; Capotorto, R.; Germano, D.; Sciaraffa, N.; Borghini, G.; Babiloni, F.; Aricò, P. o-CLEAN: A novel multi-stage algorithm for the ocular artifacts' correction from EEG data in out-of-the-lab applications. *J. Neural Eng.* **2024**, *21*, 056023. [[CrossRef](#)] [[PubMed](#)]
45. Pellegrini, E.; Anschuetz, M.; Sander, C.; Wirths, J.; Hegerl, U.; Couvy-Duchesne, B. How preprocessing shapes EEG-based clinical classification: A systematic evaluation framework. *J. Neural Eng.* **2024**, *21*, 056007.
46. Shamsi, H. Alzheimer's diagnosis from EEG with reliable probabilities: Subject-wise, leakage-free evaluation and isotonic calibration. *J. Eng. Appl. Sci.* **2025**, *72*, 226. [[CrossRef](#)]
47. Chen, Y.; Wang, H.; Zhang, D.; Zhang, L.; Tao, L. Multi-feature fusion learning for Alzheimer's disease prediction using EEG signals in resting state. *Front. Neurosci.* **2023**, *17*, 1272834. [[CrossRef](#)]
48. Lal, U.; Chikkankod, A.V.; Longo, L. A comparative study on feature extraction techniques for the discrimination of frontotemporal dementia and Alzheimer's disease with electroencephalography in resting-state adults. *Brain Sci.* **2024**, *14*, 335. [[CrossRef](#)]
49. Huggins, C.J.; Escudero, J.; Parra, M.A.; Scally, B.; Anghinah, R.; Vitória Lacerda De Araújo, A.; Basile, L.F.; Abasolo, D. Deep learning of resting-state electroencephalogram signals for three-class classification of Alzheimer's disease, mild cognitive impairment and healthy ageing. *J. Neural Eng.* **2021**, *18*, 046087. [[CrossRef](#)]
50. Jiang, R.; Zheng, X.; Sun, J.; Chen, L.; Xu, G.; Zhang, R. Classification for Alzheimer's disease and frontotemporal dementia via resting-state electroencephalography-based coherence and convolutional neural network. *Cogn. Neurodyn.* **2025**, *19*, 46. [[CrossRef](#)] [[PubMed](#)]
51. Hallquist, M.N.; Hillary, F.G. Graph theory approaches to functional network organization in brain disorders: A critique for a brave new small-world. *Netw. Neurosci.* **2018**, *3*, 1–26. [[CrossRef](#)]
52. Duan, F.; Huang, Z.; Sun, Z.; Zhang, Y.; Zhao, Q.; Cichocki, A.; Yang, Z.; Sole-Casals, J. Topological network analysis of early Alzheimer's disease based on resting-state EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 2164–2172. [[CrossRef](#)]
53. Klepl, D.; He, F.; Wu, M.; Blackburn, D.J.; Sarriani, P. EEG-based graph neural network classification of Alzheimer's disease: An empirical evaluation of functional connectivity methods. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2022**, *30*, 2651–2660. [[CrossRef](#)]
54. Sharma, R.; Meena, H.K. Graph based novel features for detection of Alzheimer's disease using EEG signals. *Biomed. Signal Process. Control* **2025**, *103*, 107380. [[CrossRef](#)]
55. Zheng, H.; Xiao, H.; Zhang, Y.; Jia, H.; Ma, X.; Gan, Y. Time-Frequency functional connectivity alterations in Alzheimer's disease and frontotemporal dementia: An EEG analysis using machine learning. *Clin. Neurophysiol.* **2025**, *170*, 110–119. [[CrossRef](#)] [[PubMed](#)]
56. Shan, X.; Cao, J.; Huo, S.; Chen, L.; Sarriani, P.G.; Zhao, Y. Spatial-temporal graph convolutional network for Alzheimer classification based on brain functional connectivity imaging of electroencephalogram. *Hum. Brain Mapp.* **2022**, *43*, 5194–5209. [[CrossRef](#)]
57. Parvande, S.; Yeh, H.W.; Paulus, M.P.; McKinney, B.A. Consensus features nested cross-validation. *Bioinformatics* **2020**, *36*, 3093–3098. [[CrossRef](#)]
58. Miltiadous, A.; Tzimourta, K.D.; Afrantou, T.; Ioannidis, P.; Grigoriadis, N.; Tsalikakis, D.G.; Angelidis, P.; Tsipouras, M.G.; Glavas, E.; Giannakeas, N.; et al. A dataset of scalp EEG recordings of Alzheimer's disease, frontotemporal dementia and healthy subjects from routine EEG. *Data* **2023**, *8*, 95. [[CrossRef](#)]
59. Nour, M.; Senturk, U.; Polat, K. A novel hybrid model in the diagnosis and classification of Alzheimer's disease using EEG signals: Deep ensemble learning (DEL) approach. *Biomed. Signal Process. Control* **2024**, *89*, 105751. [[CrossRef](#)]
60. Devnath, L.; Summons, P.; Luo, S.; Wang, D.; Shaukat, K.; Hameed, I.A.; Aljuaid, H. Computer-aided diagnosis of coal workers' pneumoconiosis in chest x-ray radiographs using machine learning: A systematic literature review. *Int. J. Environ. Res. Public Health* **2022**, *19*, 6439. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.