6 (2026) 100298

Contents lists available at ScienceDirect

Neuroscience Informatics

journal homepage: www.elsevier.com/locate/neuri

Review

EEG-centred computational biomarkers and complementary physiological signals for mild cognitive impairment and early Alzheimer's disease: A systematic review 2005–2025

Ha Truong *, Victoria Kvasniak, Gautam Anand, Anubha Kalra, Farhaan Mirza, Kerri Spooner, Sruthy Krishna

Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand

ARTICLE INFO

Keywords:
Alzheimer's disease
Mild cognitive impairment
Electroencephalography
Biomedical signal processing
Machine learning

ABSTRACT

Background: Electroencephalography (EEG) provides a portable approach to assessing mild cognitive impairment (MCI) and early Alzheimer's disease (AD), while sleep and autonomic signals may offer complementary physiological information. The validation quality of these computational biomarkers remains uncertain.
Objective: To systematically evaluate signal-processing approaches, candidate biomarkers, computational models, and risk of bias in an EEG-centred evidence base for MCI and early AD, with complementary physiological studies considered as secondary evidence.
Methods: Following PRISMA 2020 guidelines, we systematically reviewed 118 studies (2005–2025), comprising 113 direct EEG studies, two studies of polysomnography-derived sleep architecture, and three complementary non-EEG studies, and appraised risk of bias using a PROBAST-informed framework.
Results: Recurring EEG findings included spectral slowing, reduced signal complexity, altered event-related potentials, and disrupted functional connectivity. Across 120 extracted performance rows, the median reported accuracy was 93.41%, and values above 95% were common. The proportion classified as High overall risk ranged from 38.1% to 69.5% across operational sample-size thresholds and was 22.0% when sample size was excluded. Twenty-two studies were at high risk of data leakage, only one reported external validation, and longitudinal evidence remained limited (9 studies; none used time-to-event modelling).
Conclusions: The EEG-centred literature presents a performance paradox: high reported accuracy coexists with data-leakage concerns, limited external validation, and little longitudinal evidence. Future studies should use design-specific sample-size planning, strict subject-level separation, external validation, and longitudinal outcome modelling to establish clinically reliable computational biomarkers.

1. Introduction

Dementia is a progressive neurodegenerative syndrome affecting more than 57 million people worldwide, with prevalence projected to increase substantially in the coming decades [1]. Alzheimer's disease (AD) accounts for most cases and is characterised by a prolonged preclinical and prodromal phase during which pathological and neurophysiological changes may precede overt dementia by many years [2]. Mild cognitive impairment (MCI), particularly amnestic MCI, represents an important stage for identifying individuals at increased risk of progression to AD [3]. Accessible physiological markers that distinguish healthy ageing, MCI, and early AD could therefore support screening, risk stratification, and enrolment in disease-modifying trials.

Structural and molecular imaging methods, including magnetic resonance imaging and positron emission tomography, provide important AD biomarkers but remain costly, infrastructure-dependent, and less suited to repeated or population-level assessment [4]. Electroencephalography (EEG) offers a portable, relatively low-cost, and non-invasive measure of neural activity. Accordingly, this review centres on EEG while considering polysomnography-derived sleep architecture and selected autonomic, movement, and respiratory measures as complementary indicators of physiological dysfunction in MCI and early AD. Electrooculography and electromyography contribute to standard polysomnographic sleep staging but are not represented as standalone modalities. In this review, "early stage" refers to MCI and prodromal or early AD as defined by the included studies. This term primarily

* Corresponding author.
E-mail addresses: ha.truong@aut.ac.nz (H. Truong), victoria.kvasniak@aut.ac.nz (V. Kvasniak), gautam.anand@aut.ac.nz (G. Anand), anubha.kalra@aut.ac.nz (A. Kalra), farhaan.mirza@aut.ac.nz (F. Mirza), kerri.spooner@aut.ac.nz (K. Spooner), sruthy.krishna@autuni.ac.nz (S. Krishna).

https://doi.org/10.1016/j.neuri.2026.100298
Received 20 June 2026; Received in revised form 5 October 2026; Accepted 15 October 2026
Available online 16 October 2026
2772-5286/© 2026 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

describes cross-sectional disease status and should not be interpreted as prospective prediction of incident dementia.

Previous reviews have identified several recurring EEG alterations in MCI and AD, including spectral slowing, reduced signal complexity, altered event-related potentials, and disrupted functional connectivity [5–7]. Computational analysis has progressed from statistical methods and handcrafted spectral, entropy, nonlinear, and connectivity features to conventional machine-learning models, including support vector machines, random forests, and gradient boosting [8–12]. More recently, deep-learning approaches have enabled representation learning from raw EEG, spectrograms, and connectivity matrices, while graph-based and explainable models have sought to improve neurophysiological relevance and interpretability [13–15]. These developments have produced numerous reports of high diagnostic accuracy and have expanded interest in EEG as a computational biomarker for early cognitive impairment.

Despite this progress, the evidential strength of reported performance remains uncertain. Studies differ substantially in cohort composition, diagnostic criteria, signal acquisition, preprocessing, feature definitions, classification tasks, and validation strategies [7,16]. Small samples, inadequate subject-level separation, data leakage, and reliance on internal validation may produce optimistic performance estimates that do not generalise to independent clinical populations. Furthermore, most studies evaluate cross-sectional discrimination between diagnostic groups, whereas relatively few investigate progression from MCI to AD or predict future clinical outcomes [15,17,18]. Consequently, high reported accuracy cannot be interpreted independently of methodological validity and study purpose.

To address these gaps, this review synthesises EEG biomarkers, computational methods, and methodological validity across 118 studies published from 2005 to 2025. The evidence base comprises 113 direct EEG studies, two studies of polysomnography-derived sleep architecture, and three studies of autonomic, movement, or respiratory physiology. We examine (i) physiological biomarkers, distinguishing EEG-derived findings from complementary sleep and autonomic evidence, and (ii) statistical, classical machine-learning, and deep-learning approaches.

The primary contribution is a systematic, study-level PROBAST-informed appraisal of five domains: sample size adequacy, validation strategy, class imbalance handling, data leakage, and reference standard clarity. By relating these domains to reported performance, we identify recurring biomarkers, characterise the performance paradox between high accuracy and methodological weakness, distinguish cross-sectional classification from longitudinal progression modelling, and assess external validation. This integrated analysis evaluates the reproducibility and clinical translation readiness of EEG-centred computational biomarkers.

2. Research methods

2.1. Systematic review protocol

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [19]. The review protocol was not prospectively registered. A structured search was performed in PubMed, Scopus, and IEEE Xplore to identify EEG-centred studies applying physiological signals and computational methods to MCI, prodromal AD, or early-stage AD classification or prediction. The search covered publications from January 2005 to December 2025. The final search was conducted on 10 February 2026.

The search strategy combined three conceptual domains: (1) early-stage cognitive impairment, (2) physiological signal modalities (EEG, ECG, EOG, EMG, sleep-related measures), and (3) computational modelling approaches (machine learning, deep learning, and advanced signal processing). Database-specific syntax and field tags were used, and complete search strings for each database, including filters and limits, are provided in Appendix A of the Supplementary Material. filters were applied to restrict results to human studies published in English. Some exclusion terms were applied at the search stage to reduce clearly irrelevant retrievals, while final eligibility decisions were made during title/abstract and full-text screening using the criteria in Table 1.

Table 1
Study exclusion criteria used during screening.

Category	Exclusion Criteria
Outcome Not Relevant	No assessment of Alzheimer's disease (AD), MCI, prodromal AD, or early cognitive decline; severity-only studies without early-stage populations.
Signal Not Relevant	No EEG, polysomnography-derived sleep measure, or computationally analysed complementary autonomic, movement, or respiratory signal; non-physiological or behavioural modalities (e.g., gesture recognition, gait, handwriting, VR, speech).
Method Not Relevant	No predictive modelling or classification analysis; descriptive-only or signal-processing studies without diagnostic evaluation.
Technology Focus	Brain-machine interface, human-computer interaction, or hardware validation studies without clinical dementia modelling.
Clinical Domain Mismatch	Clinical populations or disease-specific subgroups outside the Alzheimer's disease continuum when the physiological-signal model was not focused on MCI, prodromal AD, or early AD.
Population	Healthy volunteers only; non-Alzheimer's neurological conditions; moderate-severe dementia without early-stage focus.
Study Type	Reviews, conference papers, study protocols, inaccessible articles.

Operational eligibility prioritised studies that directly analysed EEG. Polysomnography studies were retained when computationally derived sleep architecture was evaluated in MCI or early AD, and complementary non-EEG studies were retained only when they computationally analysed autonomic, movement, or respiratory signals in the same target populations. These complementary studies were classified as secondary contextual evidence and were not used to support EEG-specific biomarker claims. The effect of retaining the three non-EEG-only studies was examined in a separate scope sensitivity analysis.

Specialised domains such as functional connectivity, graph theory, event-related potential (ERP) paradigms, quantitative EEG (qEEG), heart rate variability (HRV), wearable recordings, microstates, and sleep spindles were not individually specified as search terms; instead, they were captured through broader EEG, sleep, physiological-signal, feature-extraction, and signal-processing terms. This strategy reduced search complexity and vocabulary mismatches across databases but may have missed studies that used highly specialised terminology without explicit EEG, sleep, physiological-signal, or computational-modelling terms. Similarly, search-stage exclusions for terms such as MRI, PET, Parkinsonism, depression, and transcranial magnetic stimulation (TMS) may have reduced retrieval of multimodal validation or comorbidity-focused studies. These limitations should be considered when interpreting the comprehensiveness of the review.

2.2. Study selection process and data extraction

Study selection followed a three-stage process: (1) duplicate and incomplete-record removal, (2) title and abstract screening, and (3) full-text assessment. The database search yielded 366 records (139 PubMed, 181 Scopus, 46 IEEE Xplore). After removal of 104 duplicates and four records with missing data, 258 records were screened. Of these, 169 reports were assessed for eligibility and 118 studies were included in the review. Two reviewers independently assessed the 169 records for eligibility against the inclusion criteria. Inter-rater agreement was substantial (Cohen's $\kappa = 0.69$, 95% CI: 0.58–0.81). The 24 initial disagreements were resolved through consensus discussion, yielding 118 included and 51 excluded studies. The included reports were then reviewed by the same reviewers to extract key methodological and quantitative information, including signal modalities, feature categories, computational models, validation methods, and reported performance metrics (e.g., accuracy, sensitivity, AUC). Extracted variables were entered into structured data tables and then used to generate standardised summaries and

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

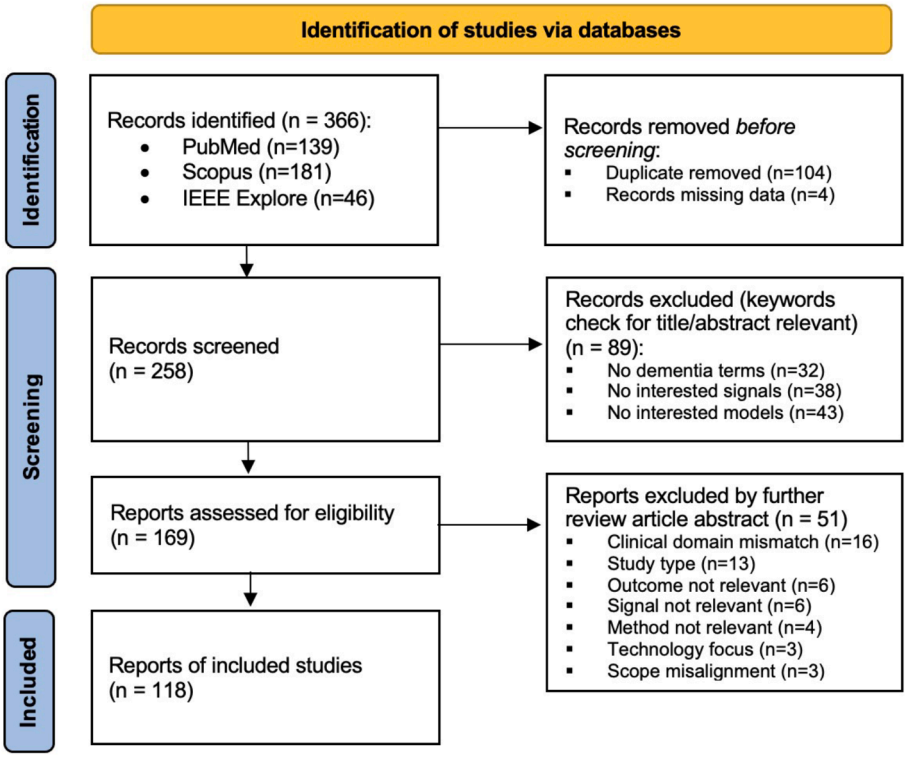


Fig. 1. PRISMA flow diagram of study selection.

plots for cross-study comparison. The complete selection process, including reasons for exclusion at each stage, is presented in the PRISMA flow diagram (Fig. 1).

2.3. Data synthesis and meta-analysis feasibility

Findings were synthesised narratively and descriptively. A formal meta-analysis was precluded by substantial heterogeneity across diagnostic contrasts, signal-acquisition protocols, preprocessing pipelines, feature definitions, classifier families, validation schemes, and reporting units. Most studies reported task-level accuracy without sufficient information to derive comparable subject-level effect sizes; some used epoch- or segment-level performance that may not be independent at the participant level; and sensitivity, specificity, and AUC were incompletely reported. Accuracy values were therefore not treated as exchangeable effect estimates or pooled measures of clinical performance. Instead, task-level summaries describe the distribution of published results and were interpreted in relation to validation rigour, sample size, class imbalance, and leakage risk.

Model-category summaries were used to characterise patterns in the published literature, including the central tendency, dispersion, and range of reported performance. These summaries provide an overview of the evidence landscape and support the identification of research trends and hypotheses for future benchmark studies. Because the included studies differed in cohorts, diagnostic tasks, preprocessing pipelines, feature sets, and validation designs, the summaries were interpreted in their study context rather than as controlled head-to-head comparisons.

Performance was extracted separately for each diagnostic contrast, with one value retained per study-task combination. Training-set results were excluded, and participant-level estimates were preferred to epoch- or segment-level estimates. Validation evidence was prioritised as external validation, independent subject-level testing, and subject-wise cross-validation. For the principal cohort at the highest available validation level, we retained the model or feature configuration with the highest reported accuracy; ties were resolved using the approach identified by the authors as primary or proposed. The resulting estimates represent the highest eligible accuracy under the most rigorous validation design reported for each task.

2.4. Risk of bias assessment

We assessed the methodological quality and risk of bias of the 118 included studies using a tailored framework based on the Prediction model Risk Of Bias ASsessment Tool (PROBAST) [20], focusing on five key domains relevant to computational biomarker modelling: sample size adequacy, validation strategy, handling of class imbalance, data leakage, and clarity of the reference standard (Table 2). Each study was rated as Low, High, or Unclear risk of bias in each domain. Because no single appraisal tool was designed specifically for the heterogeneous EEG, sleep, ECG, and computational-modelling studies included here, the framework should be interpreted as a structured review-specific screen informed by prediction-model appraisal principles, not as a full PROBAST + AI or Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis plus AI (TRIPOD + AI) assessment.

Published methodological guidance does not identify a universal total-participant minimum applicable across the heterogeneous statistical, machine-learning, and deep-learning studies included in this review. Prediction-model guidance recommends study-specific calculations based on outcome prevalence, candidate parameters, anticipated model performance, and acceptable overfitting [21,22]. EEG-specific methodological work likewise recommends design-specific power analysis or simulation rather than a fixed participant count [23,24]. The information required to reconstruct such calculations was rarely reported in the included studies. We therefore evaluated sample size using operational thresholds of 30, 50, 75, and 100 participants and repeated the overall assessment after excluding the sample-size domain. These thresholds were used only to examine the sensitivity of the review-level classifications and should not be interpreted as universal minimum sample sizes or evidence of model-specific adequacy.

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

Table 2
Definitions of risk-of-bias domains used in the adapted PROBAST framework for machine-learning and AI studies.

Domain	Low risk	High risk	Unclear
Validation strategy	External validation OR strong subject-level internal validation (e.g., leave-one-subject-out [LOSO], nested cross-validation [CV]).	Inappropriate validation (e.g., no validation, resubstitution testing, clear leakage).	Validation method not described.
Handling of class imbalance	Balanced groups OR sensitivity/specificity reported alongside accuracy.	Imbalanced groups AND only accuracy reported.	Class distribution not reported.
Data leakage	Explicit participant-level data splitting (e.g., subject-wise hold-out, LOSO).	Epoch/window-level split OR same participant data used in train and test.	Splitting strategy not described.
Reference standard clarity	Clear diagnostic criteria stated (e.g., NINCDS-ADRDA, DSM-5, Petersen, MMSE/MoCA thresholds).	Clearly inappropriate or non-clinical outcome definition.	Diagnostic criteria missing or vaguely described.
Sample-size operational sensitivity analysis	Under the corresponding sensitivity specification, total sample size ≥ 30, ≥ 50, ≥ 75, or ≥ 100 participants.	Under the corresponding sensitivity specification, total sample size < 30, < 50, < 75, or < 100 participants.	Total sample size not reported.

Note: The participant thresholds are operational sensitivity specifications rather than validated minimum sample sizes.

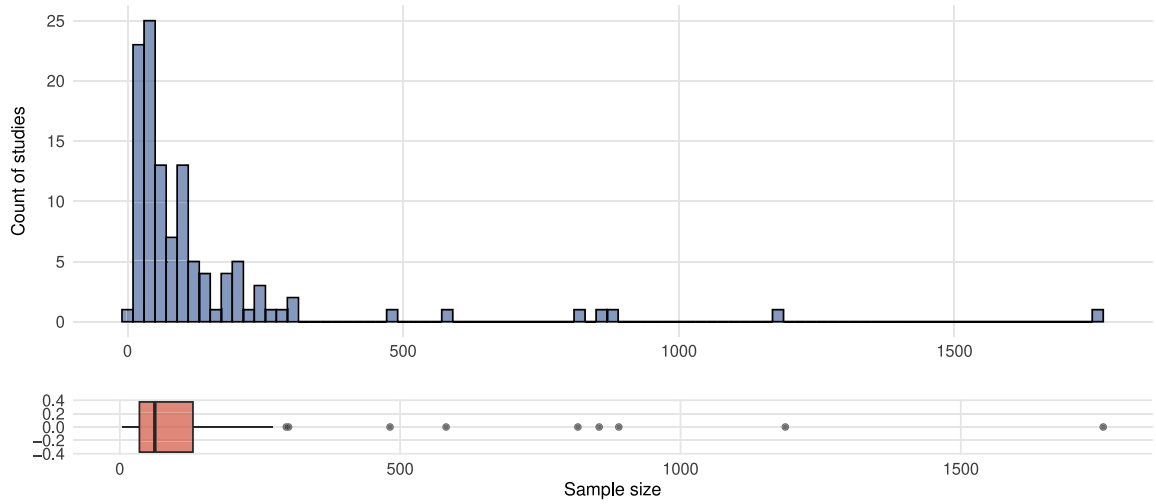


Fig. 2. Distribution of sample sizes across included studies. For visual clarity, the plot excludes one cohort of 8044 participants.

For each sensitivity specification, overall risk was classified as High when at least one included domain was High, Unclear when no included domain was High but at least one was Unclear, and Low only when all included domains were Low. Consequently, changing or excluding the sample-size specification could move a study from High to either Low or Unclear, depending on its remaining domain ratings.

3. Results

3.1. Study characteristics

Across the included studies, sample sizes were highly heterogeneous and strongly right-skewed (Fig. 2), with a clear concentration of studies in small to moderate sample sizes. Approximately 60% of studies enrolled fewer than 100 participants, with 18.6% including fewer than 30, while only a small proportion (7%) exceeded 500 participants. A limited number of large-scale datasets were reported, including cohorts of 890 [25], 1187 [26], 1754 [27], and 8044 participants [28]. Overall, the literature remains dominated by small to moderate sample sizes, which has important implications for model generalisability and risk of overfitting. Study designs varied in diagnostic composition, with some including all three groups (healthy controls, mild cognitive impairment, and Alzheimer's disease), while others focused on pairwise comparisons. As a result, reported sample sizes reflect total cohort size rather than standardised group structures.

3.1.1. Data acquisition and preprocessing

Physiological signals provide a non-invasive window into functional brain and autonomic changes associated with cognitive decline. The mutually exclusive modality audit in Fig. 3 confirms that direct EEG studies constituted 113 of 118 records (95.8%). Two studies (1.7%) evaluated sleep architecture derived from polysomnography, and three (2.5%) analysed complementary non-EEG signals. The complementary evidence comprised ECG and heart-rate-variability analyses in Geng et al. and Xavier et al. and sleep movement and respiration analysis in Khosroazad et al. [29–31]. These studies provide contextual evidence on autonomic and sleep-related physiology rather than EEG-specific biomarkers. Illustrative EEG studies used acquisition protocols based on the international 10–20 or 10–10 systems, including 19-channel clinical montages and higher-density configurations for network analyses [15,32,33]. Selected studies reported sampling rates from 128 to 1024 Hz and recording durations ranging from under one minute to approximately 30 min [14,34,35]. Resting-state recordings, particularly eyes-closed conditions, were commonly represented, while task-based paradigms such as P300 were also used [36]. Because montage, sampling rate, duration, recording setting, device portability, and home-versus-clinic acquisition context were not systematically extracted across all included studies, these examples should not be interpreted as a quantitative survey of acquisition practice.

Across the complete evidence set, the median reported accuracy was 93.41%, 22 studies were at high risk of data leakage, and one study re-

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

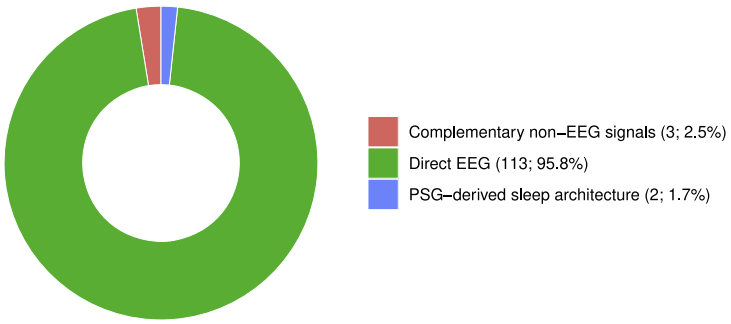


Fig. 3. Mutually exclusive evidence categories across the 118 included studies.

Table 3
Major feature domains and representative extraction methods.

Feature Domain	Representative Methods
Time-Frequency and Statistical Features	Mean, variance, higher-order moments; band power; peak alpha frequency; power spectral density; wavelet transforms; continuous wavelet transform; synchrosqueezed transforms
Nonlinear and Complexity Measures	Sample, fuzzy, and multiscale entropy; fractal dimension; transfer entropy; stochastic event synchrony
Functional Connectivity Metrics	Coherence; phase-lag index (PLI); weighted phase-lag index (wPLI); synchronization likelihood
Graph-Theoretic Network Features	Clustering coefficient; global efficiency; modularity; small-worldness; Minimum Spanning Tree metrics; visibility graphs
Event-Related Potential Features	P200; N200; P300; N400/P600 latency and amplitude measures
Sleep-Derived Features	N1/N2/N3/REM proportions; spindle density; slow oscillations; REM microstructure

ported external validation. A scope sensitivity analysis excluding the three complementary non-EEG studies yielded a median accuracy of 93.44%, retained the same counts for high leakage risk and external validation, and produced High overall-risk estimates ranging from 39.1% to 70.4% across the sample-size specifications. These closely comparable results indicate that the principal performance and methodological patterns were robust to the inclusion of complementary physiological evidence.

Representative preprocessing approaches included bandpass and notch filtering, followed by artefact removal using methods such as independent component analysis or artifact subspace reconstruction [37,38]. Some studies segmented EEG and sleep signals into short epochs, while ECG-derived measures were computed over longer windows [35]. Preprocessing choices can substantially influence feature stability and model performance, highlighting the need for transparent reporting and further standardisation [39,40].

3.1.2. Feature extraction and selection

Across EEG studies, feature extraction commonly clustered into three main domains: spectral and time–frequency features (e.g., band power and peak alpha frequency), nonlinear measures (including entropy and fractal metrics), and functional connectivity with graph-theoretic measures. Spectral features were most widely used [41], while nonlinear approaches captured signal complexity [42,43]. Functional connectivity measures were increasingly applied to characterise network-level alterations [44]. A comprehensive summary of extracted features across studies is provided in Table 3. Deep learning approaches were also used to learn representations directly from raw or transformed signals [45].

Feature selection was widely applied to address high dimensionality and limited sample sizes, using approaches such as statistical filtering, recursive feature elimination, and principal component analysis [46,47]. Across studies, reducing feature and channel dimensionality often maintained or improved reported performance by removing redundant inputs. Notably, several studies suggest that compact electrode configurations can achieve competitive accuracy, supporting further investigation of wearable and scalable diagnostic systems [11,48].

Table 4
Key EEG spectral alterations associated with early cognitive decline, with supporting studies.

EEG Feature	HC	MCI	AD	Refs
Delta Power (0.5-4 Hz)	Low	Increase	Strong increase	[25,49]
Theta Power (4–7 Hz)	Moderate	Increase	Strong increase	[25,53]
Alpha Power (8–13 Hz)	High	Decrease (posterior)	Strong decrease	[9,80]
Beta Power (13–30 Hz)	Moderate	Slight decrease	Decrease	[41,52]
PAF	~ 10 Hz	$\downarrow (\sim 9-9.5\ Hz)$	$\downarrow (\sim 8-9\ Hz)$	[9]

3.2. Physiological biomarkers

Studies reported a broad range of physiological biomarkers across EEG, sleep physiology, ECG and related modalities. These biomarkers reflect reported changes in oscillatory activity, signal complexity, network organisation and cognitive processing in early cognitive decline. The following sections summarise key biomarker families frequently observed across studies.

3.2.1. Linear signal features: Spectral and temporal descriptors

Spectral features are the most widely used EEG biomarkers for cognitive impairment, capturing oscillatory dynamics through power distributions, peak frequency shifts, and band ratios. A frequent finding is slowing of neural activity in MCI and AD, with increased delta and theta power and reduced posterior alpha and beta power, consistent with disrupted thalamocortical and cortico-cortical interactions [9,25,49]. Peak alpha frequency (PAF) has been reported as a useful marker, decreasing from around 10 Hz in healthy adults to lower values in MCI and AD [9]. Spectral ratios such as theta/alpha and theta/beta further summarise these changes and may improve discrimination, although sensitivity is generally higher for AD than early MCI [50]. Temporal statistical features provide complementary information on amplitude variability and waveform structure. Measures such as variance, skewness, kurtosis, zero-crossing rate, and Hjorth parameters have shown utility, particularly when combined with spectral features [41,51]. In addition, time–frequency decomposition methods, including wavelet and synchrosqueezing transforms, improve the resolution of transient oscillatory events and enhance sensitivity for early MCI detection [52].

A summary of these spectral and temporal alterations is provided in Table 4. Together, these linear descriptors form a core set of biomarkers but primarily capture quasi-linear signal properties, motivating the use of nonlinear and complexity-based measures discussed in the next section.

3.2.2. Nonlinear and complexity measures

Nonlinear and complexity-based features capture aspects of EEG dynamics not reflected in linear spectral measures, particularly signal irregularity, unpredictability, and multi-scale structure, which are

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

Table 5
Hierarchical organisation of EEG connectivity and graph-theoretic biomarkers in early cognitive impairment. The table summarises levels of network analysis, representative measures, inferred pathophysiological implications and typical change in MCI/AD.

Level of Analysis	Representative Measures	Pathophysiological Interpretation	Change in MCI/AD
Local spectral activity	Theta/alpha power distribution	Local neuronal dysfunction	Theta ↑; Alpha ↓; regional slowing
Pairwise functional connectivity	PLI, wPLI, coherence, wavelet coherence	Disrupted inter-regional communication	↓ Long-range frontal-posterior synchronisation; altered theta/alpha coupling
Global graph topology	Global efficiency, clustering coefficient, path length	Reduced integration and segregation balance	↓ Efficiency; ↓ clustering; ↑ path length
Topological network structure	Minimum Spanning Tree metrics; visibility graphs	Early large-scale network reconfiguration	Altered hub organisation; increased network fragmentation

known to decline in early cognitive impairment. Entropy-based measures, including approximate, sample, fuzzy, and multiscale entropy, often show reduced signal complexity in MCI and further reductions in AD, sometimes correlating with cognitive performance [43,54]. Multivariate extensions capture inter-channel interactions and may provide additional sensitivity to early impairment. Other nonlinear descriptors, such as fractal dimension and Lempel–Ziv complexity, similarly reflect reduced flexibility of neural activity [42]. Wavelet-based nonlinear features may further enhance discrimination, particularly when combined with channel optimisation strategies [55]. Information-theoretic approaches provide complementary insights into disrupted neural communication. Measures such as transfer entropy and nonlinear synchronisation capture directional interactions and network-level dynamics, and some studies report improved sensitivity compared with classical coherence measures in detecting early dysfunction [56,57].

Overall, nonlinear and complexity measures offer a complementary perspective on early cognitive decline by characterising changes in neural variability and network dynamics. These features may be useful for short-duration recordings and portable screening applications, but require further validation.

3.2.3. Network-level connectivity and graph topology

Early Alzheimer's disease is increasingly conceptualised as involving large-scale network disruption rather than solely focal neuronal loss. From this perspective, EEG biomarkers extend beyond local spectral abnormalities to encompass altered interactions between distributed cortical regions and changes in global network topology (Table 5). This framework aligns with the disconnection hypothesis of neurodegeneration and provides a biologically plausible basis for diagnostic discrimination and progression modelling.

At the pairwise level, functional connectivity measures capture synchronisation between cortical regions. A recurring finding across studies is reduced long-range coupling, particularly between frontal and posterior areas, together with altered theta and alpha band interactions in MCI and AD [18,58]. These disturbances are consistent with impairment of integrative processes that support memory, executive function, and attentional control. Metrics such as phase-lag index (PLI), weighted phase-lag index (wPLI), synchronization likelihood, and coherence have been reported to distinguish healthy controls from MCI and AD across independent cohorts [59,60]. Graph-theoretic modelling extends this analysis by characterising global network organisation. Several studies report reduced global efficiency, decreased clustering coefficients, and disrupted hub architecture in MCI and AD, suggesting a shift toward a more fragmented and less integrated network structure [44,61,62]. Minimum Spanning Tree representations derived from wPLI connectivity further suggest disease-related topological alterations across MCI, AD, and frontotemporal dementia, and in some cases have been reported to be more stable than conventional connectivity estimators [63]. Longitudinal analyses suggest that progressive reductions in network density and increasing divergence from normative topology are associated with conversion from MCI to AD [64,65], highlighting the potential of network metrics for prognostic assessment.

Recent computational advances reinforce the relevance of network information. Graph-based representations, including visibility graph transformations of ERP dynamics, have shown promising reported performance in identifying prodromal AD and MCI converters [66,67]. Graph convolutional networks that explicitly model connectivity structure have also been reported to perform well in early impairment classification [68], supporting the view that preserving network topology may enhance sensitivity to subtle disease-related alterations. Collectively, connectivity and graph-theoretic biomarkers suggest functional disruption of cortical systems in early cognitive impairment. Network-level disturbances may complement local spectral abnormalities as candidate indicators of MCI and possible predictors of Alzheimer's disease progression.

3.2.4. Event-related potentials

Event-related potentials (ERPs) offer high temporal resolution to index neural processes that may deteriorate before overt behavioural symptoms emerge. Early sensory components, such as the P200 and N200, exhibit altered amplitudes and prolonged latencies in individuals with MCI/AD, suggesting disruptions in attentional allocation and conflict-monitoring [27,69]. Later cognitive components show recurrent abnormalities. The P300, particularly its P3b sub-component, shows reduced amplitude and delayed latency across oddball and working-memory paradigms, making it a commonly studied marker for differential diagnosis [12,69]. Similarly, language-related components such as the N400 and P600 are attenuated and delayed in early impairment, consistent with altered semantic integration and syntactic processing [70].

Some ERP measures have also been explored for prognostic modelling beyond cross-sectional discrimination. For instance, source-level features of the resting-state P3b have been reported to differentiate stable MCI from cases that progress to AD [36], and task-evoked responses during working-memory paradigms may yield stronger predictive accuracy than resting-state measures alone [71]. The practical appeal of these markers is reinforced by their feasibility in low-resource settings; even single-channel recordings paired with lightweight classifiers have shown promising MCI classification performance [11]. A summary of key ERP components, their characteristic changes, and typical experimental paradigms is presented in Table 6.

3.2.5. Sleep-derived biomarkers

Sleep physiology may provide a window into early neurodegeneration, as both non-rapid eye movement (NREM; N1–N3) and rapid eye movement (REM) sleep support memory consolidation, synaptic homeostasis, and glymphatic clearance, all of which are disrupted in Alzheimer's disease. In healthy adults, N1 typically accounts for approximately 5% of total sleep time, N2 for 45–55%, N3 for 15–25%, and REM sleep for 20–25%, although these proportions vary with age. Standard staging follows criteria from the American Academy of Sleep Medicine (AASM) and integrates EEG, EOG, and EMG signals to characterise sleep architecture [72]. A summary is provided in Table 7.

Empirical evidence links disrupted sleep architecture with early cognitive decline. Reduced slow-wave sleep (N3), lower spindle density, and shortened REM duration are associated with amyloid burden and impaired memory consolidation [73,74]. Large-scale polysomnographic analyses report that both AD and amnestic MCI show re-

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

Table 6
Summary of major ERP components, directionality of change, and typical paradigms in early cognitive impairment.

ERP Component	Cognitive Function	Typical Paradigm	Change in MCI/AD
P200	Early attention and perceptual processing	Oddball; visual discrimination	Amplitude ↓ ; Latency ↑(inconsistent)
N200	Conflict detection and cognitive control	Oddball; Go/No-Go	Amplitude ↓ (AD); latency alterations in MCI
P300 (P3b)	Attention allocation and working memory updating	Oddball; working memory	Latency ↑ (delayed); Amplitude ↓ (reduced)
N400 / P600	Semantic processing and language integration	Semantic priming; language tasks	Latency ↑ ; Amplitude ↓

Table 7
Typical EEG, EOG, EMG patterns and change of proportion in MCI/AD across sleep stages.

Stage	EEG	EOG	EMG	Change in MCI/AD
N1 - Transition from wake to sleep	Low-voltage mixed-frequency, theta (4–8 Hz)	Slow rolling eye movements	Slight decrease from wake	↑ (increased proportion; fragmented sleep)
N2 - Light sleep; largest portion	Sleep spindles (12–15 Hz), K-complexes	Minimal	Reduced	↑ N2 proportion and/or ↓ spindle density
N3 - Deep slow-wave sleep; memory consolidation	High-amplitude delta (0.5-4 Hz)	Minimal	Low	↓ (reduced slow-wave sleep)
REM - Dreaming; memory consolidation	Low-voltage mixed-frequency, similar to wake	Rapid eye movements	Very low (atonia)	↓ (reduced REM duration)

duced total sleep time, decreased N3 and REM sleep, and increased lighter stages (N1/N2), consistent with fragmented and less restorative sleep [75]. Lower N3 and REM proportions have also been associated with poorer learning and episodic memory performance, supporting a possible mechanistic link between sleep disruption and cognitive decline.

Sleep EEG has also shown discriminative performance for distinguishing MCI from controls [76]. Semi-supervised approaches using spatio-temporal sleep features report diagnostic accuracy around 90% for AD, and large-scale spectral and coherence analyses achieve area under the receiver operating characteristic curve (AUROC) values between 0.73 and 0.78 [28,77]. Sleep measures have also been incorporated into risk prediction models; a polysomnography-based nomogram in 817 participants reported AUC values of 0.968 in training and 0.953 in validation cohorts for aMCI detection [78]. Together, these findings suggest that sleep-derived measurements may complement waking EEG in computational pipelines for MCI and prodromal AD, although the evidence base remains smaller than for waking EEG.

3.3. Computational modelling

The development of computational models for EEG-centred MCI and AD classification has evolved from statistical approaches to machine learning and, more recently, deep learning (Fig. 4). Early studies (pre-2015) were based on signal-processing and statistical modelling, relying on handcrafted features such as spectral power, coherence, and entropy, analysed using linear classifiers and conventional statistical methods [79,80]. These approaches established foundational feature sets but were limited in their ability to capture complex patterns. From approximately 2015 onwards, classical machine-learning methods became dominant, driven by increased data availability and advances in feature engineering. Models such as support vector machines, random forests, and k-nearest neighbours enabled more flexible classification, particularly when combined with connectivity, nonlinear, and time–frequency features [56,81]. Since 2018, deep learning has expanded rapidly, with convolutional, recurrent, and graph-based architectures applied to raw signals and derived representations such as spectrograms and connectivity matrices [14,38]. These models enable representation learning directly from data and have reported high performance in some datasets, particularly larger or more structured datasets.

Rather than replacing earlier methods, these paradigms now coexist as complementary approaches. Classical models may remain useful in small-sample settings and offer greater interpretability, while deep learning provides scalable representation learning when adequate data and validation are available. Hybrid strategies, combining deep feature extraction with conventional classifiers, may help balance performance and clinical interpretability.

3.3.1. Modelling architectures

Computational models can be broadly grouped into three strategies, differing in how physiological signals are represented and where feature learning occurs.

Traditional machine learning relies on handcrafted features derived during preprocessing, including spectral power, entropy, Hjorth parameters, and connectivity measures (Fig. 5). These representations are typically combined with classifiers such as support vector machines, random forests, and k-nearest neighbours. Such models may remain useful, particularly in small datasets or when interpretability is required, and have reported high performance using well-designed feature sets [47,82]. **Deep learning** integrates feature learning within the model, enabling automatic extraction of spatial, temporal, and spectral representations from raw or minimally processed signals. Architectures such as convolutional, recurrent, and attention-based networks have been applied to time series, spectrograms, and connectivity matrices, with high reported performance across some classification tasks [14,35,45]. These approaches may be advantageous in larger datasets but typically involve reduced interpretability.

Hybrid approaches combine deep networks as feature extractors with classical machine-learning classifiers (Fig. 6). This strategy is especially useful in small-sample settings, where learned representations can improve separability while simpler classifiers provide stable decision boundaries. Multimodal extensions, integrating signals such as EEG and heart rate variability, further enhance performance compared to unimodal models [83,84].

Overall, these modelling paradigms are complementary rather than competing. Traditional methods offer interpretability and robustness in limited data, while deep learning enables scalable representation learning, and hybrid approaches provide a practical balance between the two.

3.3.2. Model performance patterns

Reported performance was most consistently available as accuracy (Table 8): all 120 task rows included accuracy, whereas validation type was available for 57 rows (47.5%), sensitivity for 40 rows (33.3%), specificity for 41 rows (34.2%), and AUC for 19 rows (15.8%). Consequently, the quantitative summaries below use accuracy for descriptive comparability, but they should be interpreted alongside the limited availability of threshold-independent and class-imbalance-aware metrics.

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

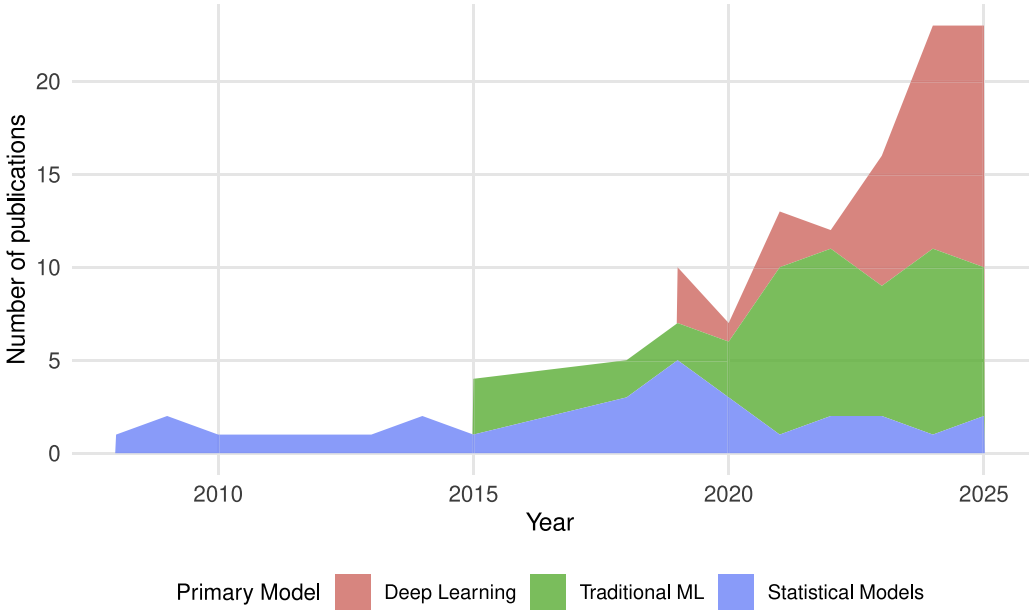


Fig. 4. Evolution of computational modelling approaches in EEG-centred MCI and AD studies (2005–2025).

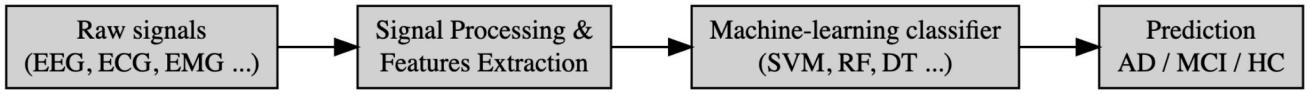


Fig. 5. Traditional machine learning using handcrafted EEG and physiological-signal features.



Fig. 6. Deep learning and hybrid pipelines for EEG-centred MCI and AD modelling.

Table 8
Availability of reported performance metrics across extracted task-level rows.

Metric	Available task rows
Accuracy	120 (100%)
Validation type	57 (47.5%)
Sensitivity	40 (33.3%)
Specificity	41 (34.2%)
F1-score	31 (25.8%)
AUC	19 (15.8%)
Precision	23 (19.2%)
Recall	17 (14.2%)

Table 9
Descriptive summary of reported accuracy patterns by model category.

Model Category	N	Mean Accuracy	SD	Median	IQR	Min	Max
Classical ML	66	90.52	7.77	92.00	10.20	71.25	100.00
Deep Learning	43	93.48	6.28	95.04	7.73	72.67	100.00
Statistical	11	90.03	6.97	93.06	12.03	79.06	98.04

Cross-study summaries of model categories provide a useful overview of how reported performance is distributed across the literature. However, these patterns require contextual interpretation because the included studies differed in cohort composition, diagnostic contrast, sample size, signal acquisition, preprocessing, feature selection, class balance, reporting unit, and validation strategy. The summaries are therefore intended to describe published performance patterns and their variability, rather than to isolate the independent effect of model architecture.

Table 9 summarises the descriptive model-category results. Within the published evidence, deep-learning approaches (N = 43 task-level rows) had the highest median reported accuracy (95.04%), followed by statistical approaches (N = 11; median 93.06%) and classical machine-learning approaches (N = 66; median 92%). Classical machine-learning results showed substantial dispersion (IQR 10.2), while statistical approaches had the widest IQR (12.03). Deep-learning results had an IQR of 7.73. These descriptive patterns show how reported performance and variability are distributed across the current literature. However, the observed differences may reflect variation in study populations, diagnostic tasks, dataset complexity, preprocessing, validation rigour, and reporting practices in addition to modelling approach. Higher median reported accuracy in one category should therefore not, by itself, be interpreted as evidence of inherent model superiority.

Within the HC-versus-MCI subset, the exploratory fitted line in Fig. 7 had a positive slope ($\beta = 0.651$ percentage points per year; $R^2 = 0.111$). Each of the 55 observations represents a different study. This descriptive pattern does not establish a temporal improvement in model performance because the studies differ in cohorts, acquisition and preprocessing, modelling approaches, validation designs, and reporting practices.

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

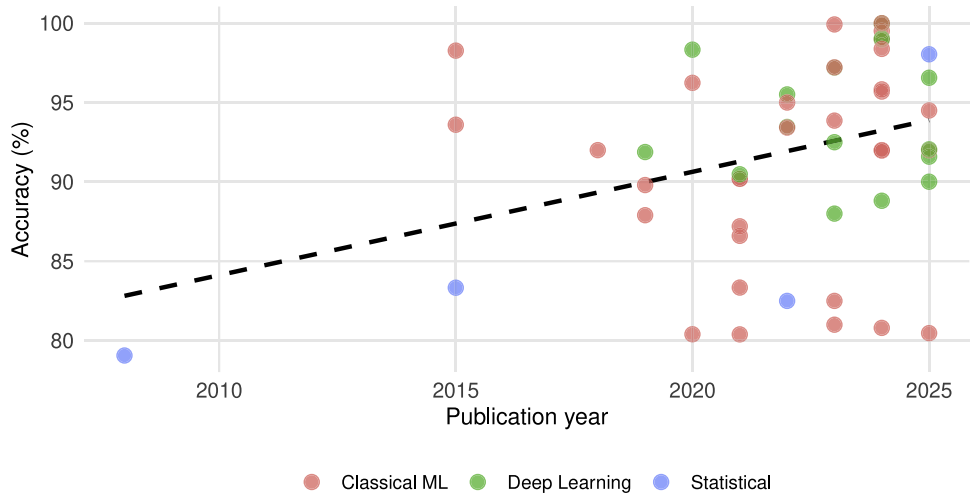


Fig. 7. Exploratory trend in reported accuracy for HC-versus-MCI classification over time. Each point represents one study, and the dashed line is a descriptive linear fit without an inferential confidence band.

Table 10
Summary statistics of accuracy across main classification tasks.

Task	N	Mean Accuracy	SD	Median	IQR	Min	Max
HCvsMCI	55	91.97	6.23	92.05	8.81	79.06	100.0
HCvsMCIvsAD	31	90.75	7.89	93.40	9.56	72.63	98.9
MCIvsAD	17	87.90	8.92	90.00	12.61	71.25	100.0

Near-perfect or perfect values should therefore be interpreted in relation to sample size, validation type, and leakage risk rather than treated as evidence of clinical readiness.

3.3.3. Task-level performance

Performance was summarised at the **task level**, where each diagnostic contrast reported in a study (e.g., HC vs MCI, MCI vs AD, HC vs MCI vs AD) was treated as one observation. A study could contribute different contrasts, so the number of task rows could exceed the number of studies, but no study contributed more than one selected value within the same diagnostic task.

Performance varied by task type, as shown in Table 10. HC–MCI classification was the most frequently evaluated task (N = 55) and had high reported accuracy (median 92.05%, IQR 8.81). Three-class HC–MCI–AD classification (N = 31) showed comparable central performance (median 93.4%) with slightly greater dispersion, reflecting increased multiclass decision complexity. In contrast, the MCI–AD task (N = 17) had the lowest median reported accuracy (90%) and the widest variability (IQR 12.61), indicating substantial heterogeneity in performance across studies. This reduced separability likely reflects the physiological and cognitive proximity between adjacent clinical stages relative to contrasts involving healthy controls.

As a descriptive heterogeneity summary in the absence of a formal meta-analysis, the HC-vs-MCI accuracy subset (N = 55) spans 79.06% to 100% (range 20.9 percentage points; IQR 8.81; SD 6.23; coefficient of variation 6.8%). A spread of this magnitude within a single diagnostic contrast, in a literature with widely varying acquisition protocols, cohort definitions, validation schemes, and class-balance handling, is consistent with substantial between-study heterogeneity and supports the decision to synthesise narratively rather than pool quantitatively.

3.3.4. Exploratory association with sample size

Within the HC-versus-MCI subset, the exploratory fit between reported accuracy and log-transformed sample size had a negative slope (β = -3.592 percentage points per ten-fold increase in sample size; R^2 = 0.042; Fig. 8). Each of the 55 observations represents a different study. The weak descriptive fit should not be interpreted as an independent effect of sample size because cohort composition, model type, preprocessing, validation rigour, and reporting practices also varied across studies. Smaller studies may be more vulnerable to overfitting and optimistic internal validation, whereas larger cohorts may contain greater clinical and acquisition heterogeneity. The plot is therefore hypothesis-generating rather than inferential.

3.3.5. Interpretability and clinical translation

Across the reviewed studies, both traditional machine learning and deep learning were frequently associated with high reported performance, with many recent works reporting accuracies above 95 percent and several reporting perfect or near-perfect accuracy under internal validation settings [14,35,84]. These values should be treated as reported estimates rather than as direct evidence of deployment-ready performance, particularly when cohorts were small, validation was internal, or subject-level separation was unclear. Traditional machine-learning models often performed well when paired with carefully engineered spectral, entropy, nonlinear, and connectivity features [12,82]. Deep-learning models were increasingly used with spectrograms, connectivity matrices, and raw multichannel EEG, but claims of superiority remain difficult to establish without standardised datasets, external validation, and comparable task definitions [45,84,85].

Within the deep-learning subset, architectural coverage was uneven: convolutional networks dominated reported applications, while transformer, graph neural network, and recurrent architectures such as long short-term memory (LSTM), gated recurrent unit (GRU), and bidirectional long short-term memory (BiLSTM) models were each represented by only a small number of task-level entries. Direct comparison of these specific architectures on the same diagnostic contrast was therefore not statistically meaningful in the present dataset, and reported performance differences between architectural families should not be interpreted as evidence of intrinsic superiority. A more systematic architecture-by-architecture comparison would require benchmark datasets, fixed validation protocols, and matched preprocessing pipelines, which the current literature does not consistently provide.

The major trade-off lies in interpretability. Deep learning can learn richer signal representations, but the internal representations are difficult to interpret and may be unstable when training data are limited. Emerging explainable AI methods such as SHapley Additive exPlanations (SHAP) and attention visualisation help address this limitation [86], yet interpretability remains lower than in many classical models. Traditional machine learning retains an advantage in transparency because feature weights, channel contributions and decision rules can

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

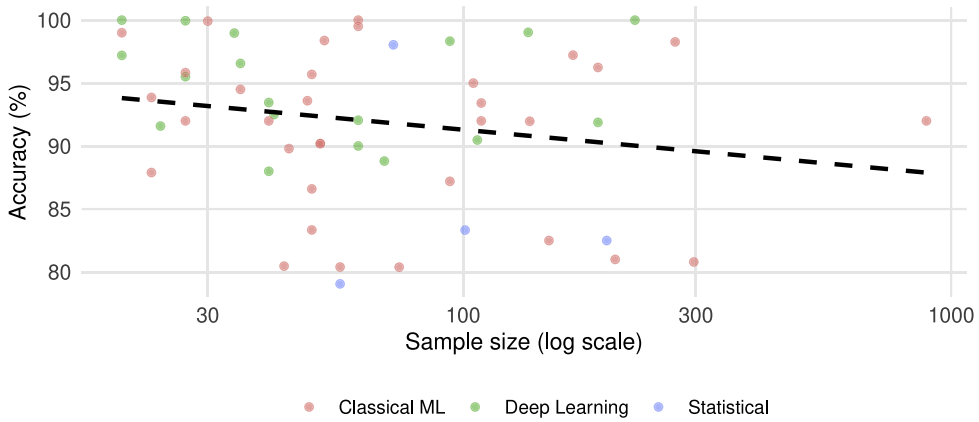


Fig. 8. Exploratory relationship between reported accuracy and sample size for HC-versus-MCI classification. Each point represents one study, and the dashed line is a descriptive linear fit without an inferential confidence band.

often be inspected directly. In summary, traditional machine learning remains well suited to smaller datasets with strong handcrafted features, while deep learning may provide scalability when larger, well-validated datasets are available. Hybrid pipelines bridge these strengths by using deep networks for representation learning and classical models for more interpretable decision-making.

3.3.6. *Classification vs prognostic prediction*

Most studies focus on cross-sectional classification tasks (e.g., HC vs MCI or MCI vs AD), and many report high performance across both machine-learning and deep-learning approaches. This indicates substantial activity in EEG-centred diagnostic discrimination, but it does not establish prospective early detection. A focused audit identified 9 of 118 studies with progression-related analyses (Table 11). These studies are heterogeneous: some used baseline EEG to predict later outcomes, some compared stable and converted MCI groups after follow-up [36], and others explored within-person EEG changes over time. Early work suggested that EEG features may capture changes associated with future conversion from MCI to AD [87]. More recent studies have examined network-based and nonlinear features, including functional connectivity, graph metrics, and signal complexity, as candidate indicators of progression risk [49,64,65]. Temporal dynamics and time-varying connectivity have also been explored for predicting decline before clinical diagnosis [17,18,88].

However, several methodological limitations constrain interpretation. First, many studies rely on retrospective discrimination using baseline data rather than explicit probability-based risk models [64,65]. Second, performance is often reported at the epoch or intra-subject level, which can overestimate clinically relevant subject-level accuracy [15,17]. Third, most studies rely on internal validation within single cohorts, with no independent external replication, limiting generalisability [18,88]. Fourth, incident-dementia and conversion studies rarely report time-to-event modelling, censoring, or competing-risk handling, making it difficult to determine whether models estimate future risk or simply discriminate observed diagnostic groups. Overall, cross-sectional classification is much more developed than prognostic modelling.

When the 9 progression-related studies were examined using the customised PROBAST-informed framework, the overall classification varied with the operational sample-size specification. Four studies were rated High at thresholds of 30 and 50 participants, and five were rated High at thresholds of 75 and 100 participants; two studies remained Unclear under each threshold. When sample size was excluded, none was rated High, seven were rated Low, and two were rated Unclear. Among the remaining domains, validation, class-imbalance handling, and reference-standard clarity were each rated Low in 8 studies and Unclear in 1, while data-leakage risk was Low in 7 and Unclear in 2. These domain ratings should not be interpreted as evidence of clinical readiness: none of the nine studies reported external validation, and none used time-to-event modelling or explicitly addressed censoring and competing risks. Small longitudinal cohorts, few participants who progressed to AD, variation in follow-up duration, heterogeneous progression definitions, and different assessment schedules further limit comparison across studies. Collectively, the available evidence supports exploratory associations between EEG features and progression, but it does not yet establish generalisable individual-level prognostic models.

3.4. *Risk of bias*

The validation and reporting profile in Table 12 highlights a major translational bottleneck for the included computational models. External validation was reported in only 1 of 118 studies, whereas 74 studies relied on internal validation only. Class imbalance was present in 48 studies, and 22 studies were rated high risk for data leakage. These findings indicate that an important limitation is not simply model choice, but the reliability of validation and reporting practices.

Table 13 shows the number and percentage of studies rated Low, High, or Unclear in each methodological domain, with sample size displayed separately under each operational threshold. Class-imbalance handling was rated Low risk in 99% of studies, validation strategy in 86%, and reference-standard clarity in 87%. Most studies with Low validation risk used approaches such as leave-one-subject-out cross-validation or subject-wise hold-out. Studies with clear reference standards reported diagnostic criteria including NINCDS-ADRDA, DSM-5, Petersen criteria for MCI, or defined cognitive-test thresholds such as MMSE or MoCA cut-offs.

Data leakage remained an important concern: 68% of studies were rated Low risk, 19% High risk, and 13% Unclear. Identified concerns included possible epoch- or window-level overlap between training and test data, unclear subject-level separation, and feature-selection procedures that may have used information from the full dataset. Such practices can lead to overfitting, inflate model-performance estimates, and limit interpretability.

As shown in Fig. 9, the proportion classified as High overall risk ranged from 38.1% using the 30-participant threshold to 69.5% using the 100-participant threshold. When the sample-size domain was excluded, 22.0% remained at High overall risk, 56.8% were Low, and 21.2% were Unclear. The overall-risk distribution therefore depended substantially on the sample-size specification. Lowering or excluding the threshold moved some studies from High to Unclear because incomplete reporting remained in other domains.

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

Table 11
Progression-related EEG evidence. The studies differ in cohort composition, follow-up design, endpoint definition, and validation strategy; reported performance values should therefore not be compared directly.

Study	Cohort and follow-up	Progression-related endpoint and EEG representation	Main reported result and interpretive limitation
[87]	29 participants; MCI outcomes observed over 1–3 years	MCI-to-AD progression; multivariate statistical pattern recognition	Four of five MCI participants with AD-like patterns later progressed; exploratory pilot cohort.
[64]	25 MCI participants; two evaluations over 3 months	MCI-to-AD conversion; nonlinear connectivity and network-density change	Four converters showed reduced network density; descriptive analysis in a small cohort.
[65]	15 participants; repeated EEG over 3 months	MCI-to-AD evolution; network coefficients and motifs	Proposed conversion coefficient to support expert review; no validated prospective risk model.
[36]	175 participants, including 135 MCI; follow-up duration not clearly reported	Stable versus converted MCI; spatial-filtered resting-state P3b dynamics	Stable-versus-converter accuracy was 74.1%; subgroup discrimination after follow-up.
[15]	Four MCI converters; two evaluations over 3 months	Within-person MCI-to-AD change; HD-EEG channel-frequency maps with explainable CNN	Reported intra-subject accuracy up to 98.97%; only four converters and no subject-level prognostic validation.
[49]	205 participants; longitudinal MCI subset (n = 72) followed for up to 3 years	Stable versus converted MCI; resting-state holospectral features	Longitudinal analysis reported sensitivity of 0.90 and specificity of 0.80; internal cohort.
[88]	137 participants; progression-trend analysis over 9–18 months in an availability-limited subset	Cognitive-decline trend; time-varying resting-state connectivity score	Progression-trend accuracy was 84.61%; subset size was not stated in the extracted report.
[17]	107 aMCI participants; 18-month follow-up	Stable versus progressive aMCI; spectral, nonlinear, and connectivity features	Subject-based accuracy was 78.37% (AUC 83.89%); epoch-based estimates were much higher.
[18]	65 aMCI participants; four EEG assessments over 12 months	Stable versus progressive aMCI; longitudinal spectral, nonlinear, and connectivity features	SVM accuracy was 94.92%; requires independent external replication.

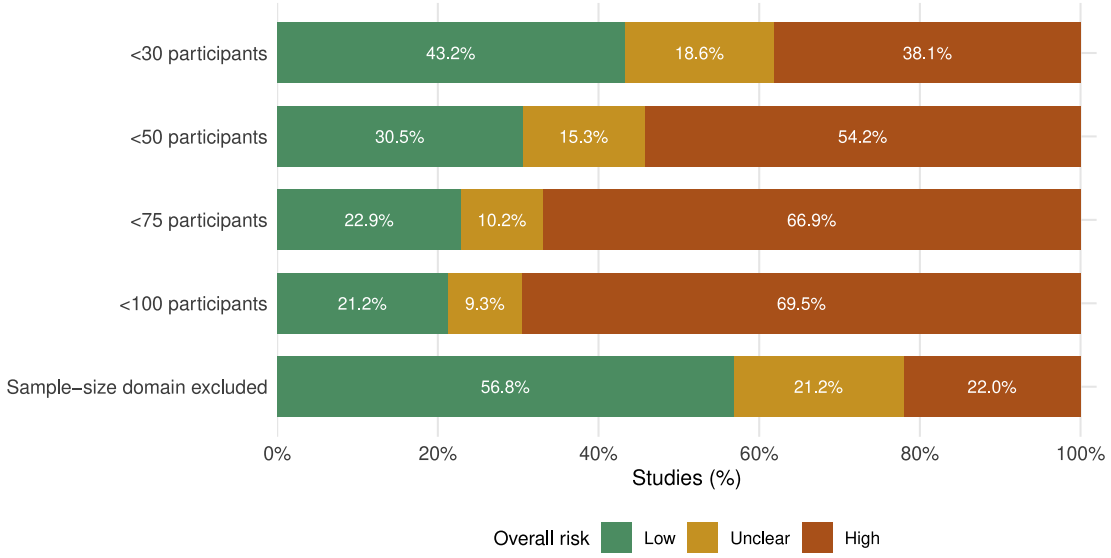


Fig. 9. Sensitivity of overall risk-of-bias classifications to the operational sample-size specification. Bars show the percentage of 118 studies classified as High, Unclear, or Low overall risk using participant thresholds of 30, 50, 75, and 100, and after excluding the sample-size domain. Participant thresholds are operational sensitivity specifications and do not represent validated minimum sample sizes.

Table 12
Validation and reporting profile across included studies.

Indicator	Studies n (%)
External validation reported	1 (0.8%)
Internal validation only	74 (62.7%)
No/unclear predictive validation	43 (36.4%)
Class imbalance reported	48 (40.7%)
Sensitivity or specificity reported	90 (76.3%)
High data-leakage risk	22 (18.6%)

Table 13
Risk-of-bias judgements across methodological domains, with sample size shown under four operational specifications.

Domain	Low risk n (%)	High risk n (%)	Unclear n (%)
Sample size: <30 participants	95 (80.5%)	22 (18.6%)	1 (0.8%)
Sample size: <50 participants	68 (57.6%)	49 (41.5%)	1 (0.8%)
Sample size: <75 participants	52 (44.1%)	65 (55.1%)	1 (0.8%)
Sample size: <100 participants	46 (39%)	71 (60.2%)	1 (0.8%)
Validation strategy	101 (85.6%)	12 (10.2%)	5 (4.2%)
Class imbalance handling	117 (99.2%)	0 (0%)	1 (0.8%)
Data leakage	80 (67.8%)	22 (18.6%)	16 (13.6%)
Reference standard clarity	103 (87.3%)	0 (0%)	15 (12.7%)

The sensitivity of the composite classification does not remove the independently observed methodological concerns. Twenty-two studies were at high risk of data leakage, only one reported external validation, and many relied exclusively on internal validation. Conclusions concerning reproducibility and generalisability should therefore be based on the complete domain profile rather than a single overall-risk percentage.

4. Discussion and conclusions

This review advances the evaluation of EEG-centred computational biomarkers by integrating three elements that have commonly been

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

considered separately: neurophysiological biomarker findings, computational model performance, and methodological validity. The recurring observations of spectral slowing, reduced signal complexity, altered event-related potentials, and disrupted functional connectivity reinforce the biological consistency of candidate EEG markers reported in previous reviews [5–7]. The present synthesis extends this evidence by determining whether the computational studies supporting these markers used methods capable of producing reliable and generalisable performance estimates.

The PROBAST-informed assessment identifies a performance paradox, although the estimated prevalence of High overall risk depends on how sample-size concerns are operationalised. Across thresholds of 30, 50, 75, and 100 participants, the proportion classified as High ranged from 38.1% to 69.5% and decreased to 22.0% when sample size was excluded. No single overall percentage should therefore be interpreted as threshold-independent. Nevertheless, frequently reported accuracy above 95% coexisted with 22 studies at high risk of data leakage, extensive reliance on internal validation, and only one external validation. In addition, only nine studies examined longitudinal progression, and none used time-to-event modelling. These findings provide a quantitative explanation for the gap between high reported accuracy and limited clinical adoption. The contribution of this review therefore extends beyond summarising which biomarkers or models performed well. It establishes how validation design, sample adequacy, leakage control, and study purpose determine the credibility and translational relevance of that performance.

Our findings caution against overconfident interpretation of high accuracy claims in the absence of rigorous validation. The following priorities emerge for future research:

- Participant-level validation should become standard. Epoch-level splitting should be explicitly avoided; leave-one-subject-out or subject-wise hold-out validation is important for unbiased performance estimation.
- Sample size adequacy requires justification. Studies should report power calculations or demonstrate that sample sizes are sufficient for the modelling approach used, particularly for deep learning.
- External validation is urgently needed. Models validated only on single-centre, internal datasets cannot be assumed to generalise; multi-site replication studies are important.
- Model-class selection should reflect data scale and validation constraints. Traditional machine learning paired with well-engineered features may remain effective in small datasets where interpretability and stability matter, while deep learning may provide scalability for larger and multimodal datasets; both face validation and interpretability challenges that motivate hybrid and explainable approaches.
- Prognostic modelling should be prioritised. Cross-sectional classification is more developed than longitudinal modelling; the field should move toward prediction of conversion risk and time-to-event outcomes.
- Subgroup and recording-context reporting should improve. The extracted datasets did not support sex- or age-stratified performance analysis, and recording context was not consistently available in a structured form. Future studies should report subgroup performance, acquisition setting, device type, and protocol details to support fairness, reproducibility, and real-world deployment assessment.

This review has several limitations. The protocol was not prospectively registered, and the customised PROBAST-informed framework represents a review-specific methodological screen rather than a complete validated PROBAST + AI assessment. The search was limited to three databases and English-language journal articles; broad search terms, search-stage exclusions, and the exclusion of conference and grey literature may have reduced coverage of specialised, multimodal, or engineering studies. Study heterogeneity precluded formal meta-analysis, and selecting the highest eligible accuracy for each study-task combination emphasised the upper range of reported performance. The findings should therefore be interpreted as a descriptive synthesis rather than pooled estimates of expected clinical performance.

In conclusion, EEG-centred computational biomarker approaches show promise for distinguishing MCI and early AD from healthy ageing, particularly when they integrate interpretable spectral, nonlinear, event-related, and connectivity-derived features. However, the current literature is characterised by a performance paradox: impressive reported accuracies coexist with methodological weaknesses that threaten validity. Without rigorous, independent validation and attention to fundamental sources of bias, the risk of overfitting and non-generalisable findings remains high. Future progress depends less on developing ever more complex models and more on design-specific sample-size planning, strict subject-level separation, external validation, and longitudinal outcome modelling within reproducible analytical pipelines.

Human and animal rights

The authors declare that the work described has not involved experimentation on humans or animals.

Informed consent and patient details

The authors declare that this report does not contain any personal information that could lead to the identification of the patient(s) and/or volunteers.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used OpenAI Codex to assist with language editing and improving readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This work has been supported by:Auckland University of Technology (AUT).

Declaration of competing interest

The authors declare that they have no known competing financial or personal relationships that could be viewed as influencing the work reported in this paper.

Acknowledgements

The authors thank Arunkumar S for reviewing and editing earlier drafts of this manuscript.

Data availability

The extracted review datasets and analysis scripts supporting this study will be made publicly available upon acceptance. Until then, access is available on reasonable request to the corresponding author.

Supplementary material

Supplementary material associated with this article can be found in the online version at 10.1016/j.neuri.2026.100298.

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

References

[1] World Health Organization, Dementia: key facts, 2023, https://www.who.int/news-room/fact-sheets/detail/dementia. Accessed: 2023-10-01.

[2] C.R. Jack, D.A. Bennett, K. Blennow, M.C. Carrillo, B. Dunn, S.B. Haeberlein, D.M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, E. Liu, J.L. Molinuevo, T. Montine, C. Phelps, K.P. Rankin, C.C. Rowe, P. Scheltens, E. Siemers, H.M. Snyder, R. Sperling, NIA-AA research framework: toward a biological definition of Alzheimer's disease, In:Alzheimer's Dement. 14(4) (2018) 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

[3] R.C. Petersen, Mild cognitive impairment, In: Continuum (Minneapolis, Minn.) 22(2) (2016) 404–418. https://doi.org/10.1212/CON.0000000000000313

[4] G.B. Frisoni, N.C. Fox, C.R. Jack, P. Scheltens, P.M. Thompson, The clinical use of structural MRI in alzheimer disease, In: Nat. Rev. Neurol. 6(2) (2010) 67–77. https://doi.org/10.1038/nrneurol.2009.215

[5] S. Yang, J.M. Sanchez Bornot, K. Wong-Lin, G. Prasad, M/EEG-based bio-markers to predict the MCI and Alzheimer's disease: a review from the ML perspective, In: IEEE Trans. Biomed. Eng. 66(10) (2019) 2924–2935. https://doi.org/10.1109/TBME.2019.2898871

[6] S. Ehteshamzad, Assessing the potential of EEG in early detection of Alzheimer's disease: a systematic comprehensive review (2000–2023), In: J. Alzheimer's Dis. Rep. 8(1) (2024) 1153–1169. https://doi.org/10.3233/ADR-230159

[7] V. Papaliagkas, et al., The role of quantitative EEG in the diagnosis of Alzheimer's disease: a critical review, In: Diagnostics 15 (2025) 1965. https://doi.org/10.3390/diagnostics15151965

[8] J. Dauwels, F. Vialatte, T. Musha, A. Cichocki, A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG, In: NeuroImage 49(1) (2010) 668–693. https://doi.org/10.1016/j.neuroimage.2009.06.056

[9] R. Lizio, C. Babiloni, C. Del Percio, A. Losurdo, L. Vernò, M. De Tommaso, A. Montemurno, G. Dalfino, P. Cirillo, A. Soricelli, R. Ferri, G. Noce, M.T. Pascarelli, V. Catania, F. Nobili, F. Famá, F. Orzi, F. Giubilei, C. Buttinelli, A.I. Triggiani, G.B. Frisoni, A.M. Scisci, N. Mastrofilippo, D.A. Procaccini, L. Gesualdo, Different abnormalities of cortical neural synchronization mechanisms in patients with mild cognitive impairment due to Alzheimer's and chronic kidney diseases: an EEG study, In: J. Alzheimers Dis. (2018). https://doi.org/10.3233/JAD-180245

[10] Y.-T. Hsiao, C.-F. Tsai, C.-T. Wu, T.-T. Trinh, C.-Y. Lee, Y.H. Liu, MCI detection using kernel eigen-relative-power features of EEG signals, In: Actuators 10(7) (2021). https://doi.org/10.3390/act10070152

[11] S. Khatun, B.I. Morshed, G.M. Bidelman, A single-channel EEG-based approach to detect mild cognitive impairment via speech-evoked brain responses, In: IEEE Trans. Neural Syst. Rehabil. Eng. (2019). https://doi.org/10.1109/TNSRE.2019.2911970

[12] Y.H. Liu, T.T. Trinh, C.F. Tsai, J.K. Yang, C.Y. Lee, C.T. Wu, A novel working memory task-induced EEG response (WM-TIER) feature extraction framework for detecting Alzheimer's disease and mild cognitive impairment, In: Biosensors (2025). https://doi.org/10.3390/bios15050289

[13] M. Buscema, F. Vernieri, G. Massini, F. Scrascia, M. Breda, P.M. Maria Rossini, E. Grossi, An improved I-FAST system for the diagnosis of Alzheimer's disease from unprocessed electroencephalograms by using robust invariant features, In: Artif. Intell. Med. 64(1) (2015) 59. https://doi.org/10.1016/j.artmed.2015.03.003

[14] C. Ieracitano, N. Mammone, A. Hussain, F.C. Morabito, A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia, In: Neural Netw. (2020). https://doi.org/10.1016/j.neunet.2019.12.006

[15] F.C. Morabito, C. Ieracitano, N. Mammone, An explainable artificial intelligence approach to study MCI to AD conversion via HD-EEG processing, In: Clin. EEG Neurosci. (2023). https://doi.org/10.1177/15500594211063662

[16] C.T. Briels, D.N. Schoonhoven, C.J. Stam, H. de Waal, P. Scheltens, A.A. Gouw, Reproducibility of EEG functional connectivity in Alzheimer's disease, In: Alzheimer's Res. Ther. 12 (2020) 68. https://doi.org/10.1186/s13195-020-00632-3

[17] Y. Ge, J. Yin, C. Chen, S. Yang, Y. Han, C. Ding, J. Zheng, Y. Zheng, J. Zhang, An EEG-based framework for automated discrimination of conversion to Alzheimer's disease in patients with amnestic mild cognitive impairment: an 18-month longitudinal study, In: Front. Aging Neurosci. (2024). https://doi.org/10.3389/fnagi.2024.1470836

[18] Y. Ge, Y. Fei, C. Ding, S. Yang, Y. Fang, Y. Zheng, J. Yin, Q. Pan, N. Zhang, X. Zhang, X. Lu, J. Zhang, Predicting progression from amnestic mild cognitive impairment to Alzheimer's disease using longitudinal EEG data: a 12-month cohort study, In: Front. Aging Neurosci. (2025). https://doi.org/10.3389/fnagi.2025.1719981

[19] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J.M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, L.A. McGuinness, L.A. Stewart, J. Thomas, A.C. Tricco, V.A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, In: BMJ 372 (2021) n71. https://doi.org/10.1136/bmj.n71

[20] K.G.M. Moons, J.A.A. Damen, T. Kaul, L. Hooft, C. Andaur Navarro, P. Dhiman, A.L. Beam, B. Van Calster, L.A. Celi, S. Denaxas, et al., PROBAST + AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods, In: BMJ 388 (2025) e082505. https://doi.org/10.1136/bmj-2024-082505

[21] R.D. Riley, J. Ensor, K.I.E. Snell, F.E. Harrell, G.P. Martin, J.B. Reitsma, K.G.M. Moons, G. Collins, M. van Smeden, Calculating the sample size required for developing a clinical prediction model, In: BMJ 368 (2020) m441. https://doi.org/10.1136/bmj.m441

[22] M. van Smeden, K.G.M. Moons, J.A.H. de Groot, G.S. Collins, D.G. Altman, M.J.C. Eijkemans, J.B. Reitsma, Sample size for binary logistic prediction models: beyond events per variable criteria, In: Stat. Methods Med. Res. 28 (2019) 2455–2474. https://doi.org/10.1177/0962280218784726

[23] M.J. Larson, K.A. Carbine, Sample size calculations in human electrophysiology (EEG and ERP) studies: a systematic review and recommendations for increased rigor, In: Int. J. Psychophysiol. 111 (2017) 33–41. https://doi.org/10.1016/j.ijpsycho.2016.06.015

[24] E. Guttmann-Flury, X. Sheng, D. Zhang, X. Zhu, A priori sample size determination for the number of subjects in an EEG experiment, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 5180–5183. https://doi.org/10.1109/EMBC.2019.8857482

[25] C. Simfukwe, S.S.A. An, Y.C. Youn, Contribution of scalp regions to machine learning-based classification of dementia utilizing resting-state qEEG signals, In: Neuropsychiatr. Dis. Treat (2024). https://doi.org/10.2147/NDT.S486452

[26] M.J. Kim, Y.C. Youn, J. Paik, Deep learning-based EEG analysis to classify normal, mild cognitive impairment, and dementia: algorithms and dataset, In: Neuroimage (2023). https://doi.org/10.1016/j.neuroimage.2023.120054

[27] J.H. Bae, M. Choi, J.J. Lee, K.H. Lee, J.U. Kim, Connectivity changes in two-channel prefrontal ERP associated with early cognitive decline in the elderly population: beta band responses to the auditory oddball stimuli, In: Front. Aging Neurosci. (2024). https://doi.org/10.3389/fnagi.2024.1456169

[28] E.M. Ye, H. Sun, P.V. Krishnamurthy, N. Adra, W. Ganglberger, R.J. Thomas, A.D. Lam, M.B. Westover, Dementia detection from brain activity during sleep, In: Sleep (2023). https://doi.org/10.1093/sleep/zsac286

[29] D. Geng, Q. Wang, W. Zheng, Y. Yin, P. Xu, G. Xu, ECG- and HRV-based hybrid architecture-early detection of Alzheimer's disease and mild cognitive impairment, In: Appl. Sci. 15(23) (2025). https://doi.org/10.3390/app152312555

[30] A. Xavier, S. Noble, J. Joseph, A. Ghosh, T.G. Issac, Heart rate and its variability from short-term ECG recordings as potential biomarkers for detecting mild cognitive impairment, In: Am. J. Alzheimers Dis. Other Demen. (2024). https://doi.org/10.1177/15333175241309527

[31] S. Khosroazad, A. Abedi, M.J. Hayes, Sleep signal analysis for early detection of alzheimer's disease and related dementia (ADRD), In: IEEE J. Biomed. Health Inform. 27(5) (2023) 2264. https://doi.org/10.1109/JBHI.2023.3235391

[32] G.L. Read, I.J. Innis, Electroencephalography (EEG), in: The International Encyclopedia of Communication Research Methods, Wiley, 2017. https://doi.org/10.1002/9781118901731.iecrm0080

[33] E. Sibilano, A. Brunetti, D. Buongiorno, M. Lassi, A. Grippo, V. Bessi, S. Micera, A. Mazzoni, V. Bevilacqua, An attention-based deep learning approach for the classification of subjective cognitive decline and mild cognitive impairment using resting-state EEG, In: J. Neural Eng. (2023). https://doi.org/10.1088/1741-2552/acb96e

[34] X. Li, H. Zhou, R. Su, J. Kang, Y. Sun, Y. Yuan, Y. Han, X. Chen, P. Xie, Y. Wang, Q. Liu, A mild cognitive impairment diagnostic model based on IAAFT and BiLSTM, In: Biomed. Signal Process. Control 80 (2023). https://doi.org/10.1016/j.bspc.2022.104349

[35] H. Zhou, L. Yin, R. Su, Y. Zhang, Y. Yuan, P. Xie, X. Li, STCGRU: a hybrid model based on CNN and BiGRU for mild cognitive impairment diagnosis, In: Comput. Methods Programs Biomed. (2024). https://doi.org/10.1016/j.cmpb.2024.108123

[36] C. Porcaro, F. Vecchio, F. Miraglia, G. Zito, P.M. Rossini, Dynamics of the "Cognitive" brain wave P3b at rest for Alzheimer dementia prediction in mild cognitive impairment, In: Int. J. Neural Syst. (2022). https://doi.org/10.1142/S0129065722500228

[37] X. Li, C. Yang, P. Xie, Y. Han, R. Su, Z. Li, Y. Liu, The diagnosis of amnestic mild cognitive impairment by combining the characteristics of brain functional network and support vector machine classifier, In: J. Neurosci. Methods (2021). https://doi.org/10.1016/j.jneumeth.2021.109334

[38] S. Chen, C. Zhang, H. Yang, L. Peng, H. Xie, Z. Lv, Z.G. Hou, A multi-modal classification method for early diagnosis of mild cognitive impairment and Alzheimer's disease using three paradigms with various task difficulties, In: IEEE Trans. Neural Syst. Rehabil. Eng. 32 (2024) 1456. https://doi.org/10.1109/TNSRE.2024.3379891

[39] D.V. Puri, J.P. Gawande, J.L. Rajput, S.L. Nalbalwar, A novel optimal wavelet filter banks for automated diagnosis of Alzheimer's disease and mild cognitive impairment using electroencephalogram signals, In: Decis. Anal. J. 9 (2023). https://doi.org/10.1016/j.dajour.2023.100336

[40] T.M. Rutkowski, T. Komendziński, M. Otake-Matsuura, Mild cognitive impairment prediction and cognitive score regression in the elderly using EEG topological data analysis and machine learning with awareness assessed in affective reminiscent paradigm, In: Front. Aging Neurosci. (2023). https://doi.org/10.3389/fnagi.2023.1294139

[41] D. Pirrone, E. Weitschek, P. Di Paolo, S. DE SALVO, M.C. De Cola, EEG signal processing and supervised machine learning to early diagnose Alzheimer's disease, In: Appl. Sci. 12(11) (2022). https://doi.org/10.3390/app12115413

[42] F.C. Morabito, D. Labate, A. Bramanti, F.L. Foresta, G. Morabito, I. Palamara, H.H. Szu, Enhanced compressibility of EEG signal in Alzheimer's disease patients, In: IEEE Sens. J. 13(9) (2013) 3255. https://doi.org/10.1109/JSEN.2013.2263794

[43] R. Su, X. Li, Z. Li, Y. Han, W. Cui, P. Xie, Y. Liu, Constructing biomarker for early diagnosis of aMCI based on combination of multiscale fuzzy entropy and functional brain connectivity, In: Biomed. Signal Process. Control 70 (2021). https://doi.org/10.1016/j.bspc.2021.103000

[44] N. Mammone, S. DE SALVO, L. Bonanno, C. Ieracitano, S. Marino, A. Marra, A. Bramanti, F.C. Morabito, Brain network analysis of compressive sensed high-density EEG signals in AD and MCI subjects, In: IEEE Trans. Ind. Inform. 15(1) (2019) 527. https://doi.org/10.1109/TII.2018.2868431

[45] M. Azeker, M.S. Ozerdem, Deep insights into MCI diagnosis: a comparative deep learning analysis of EEG time series, In: J. Neurosci. Methods (2024). https://doi.org/10.1016/j.jneumeth.2024.110057

[46] Y.-T. Hsiao, C.-T. Wu, C.-F. Tsai, Y.-H. Liu, T.-T. Trinh, C.Y. Lee, EEG-based classification between individuals with mild cognitive impairment and healthy controls using conformal kernel-based fuzzy support vector machine, In: Int. J. Fuzzy Syst. 23(8) (2021) 2432. https://doi.org/10.1007/s40815-021-01186-8

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.

[47] E. Perez-Valero, C. Morillas, M.A. Lopez-Gordo, I. Carrera-Muñoz, S. López-Alcalde, R.M. Vílchez-Carrillo, An automated approach for the detection of Alzheimer's disease from resting state electroencephalography, In: Front. Neuroinform. (2022). https://doi.org/10.3389/fninf.2022.924547
[48] K. Lee, K.-M. Choi, S. Park, S.-H. Lee, C.H. Im, Selection of the optimal channel configuration for implementing wearable EEG devices for the diagnosis of mild cognitive impairment, In: Alzheimer's Res. Ther. 14(1) (2022). https://doi.org/10.1186/s13195-022-01115-3
[49] K.T. Chu, W.C. Lei, M.H. Wu, J.L. Fuh, S.J. Wang, I.T. French, W.S. Chang, C.F. Chang, N.E. Huang, W.K. Liang, C.H. Juan, A holo-spectral EEG analysis provides an early detection of cognitive decline and predicts the progression to Alzheimer's disease, In: Front. Aging Neurosci. (2023). https://doi.org/10.3389/fnagi.2023.1195424
[50] Y. Ding, Y. Chu, M. Liu, Z. Ling, S. Wang, X. Li, Y. Li, Fully automated discrimination of Alzheimer's disease using resting-state electroencephalography signals, In: Quant. Imaging Med. Surg. (2022). https://doi.org/10.21037/qims-21-430
[51] S. Saroja, N.J. Nalini, Hybrid optimized techniques for EEG based mild cognitive impairment detection using time domain feature extraction, In: Ing. Syst. Inf. 30(1) (2025) 231. https://doi.org/10.18280/isi.300120
[52] T. Nimmy John, P. Subha Dharmapalan, N. Ramshekhar Menon, Exploration of time-frequency reassignment and homologous inter-hemispheric asymmetry analysis of MCI-AD brain activity, In: BMC Neurosci. 20(1) (2019). https://doi.org/10.1186/s12868-019-0519-3
[53] S.J. Ruiz-Gómez, R. Hornero, J. Poza, A. Maturana-Candelas, N. Pinto, C. Gómez, Computational modeling of the effects of EEG volume conduction on functional connectivity metrics. application to Alzheimer's disease continuum, In: J. Neural Eng. (2019). https://doi.org/10.1088/1741-2552/ab4024
[54] J. Liu, H. Lu, X. Zhang, X. Li, L. Wang, S. Yin, D. Cui, Which multivariate multi-scale entropy algorithm is more suitable for analyzing the EEG characteristics of mild cognitive impairment?, In: Entropy (2023). https://doi.org/10.3390/e25030396
[55] M. Aljalal, S.A. Aldosari, M. Molinas, F.A. Alturki, Selecting EEG channels and features using multi-objective optimization for accurate MCI detection: validation using leave-one-subject-out strategy, In: Sci. Rep. (2024). https://doi.org/10.1038/s41598-024-63180-y
[56] J. McBride, X. Zhao, N. Munro, G. Jicha, C. Smith, Y. Jiang, Discrimination of mild cognitive impairment and Alzheimer's disease using transfer entropy measures of scalp EEG, In: J. Healthc. Eng. (2015). https://doi.org/10.1260/2040-2295.6.1.55
[57] J. Dauwels, F. Vialatte, T. Weber, T. Musha, A. Cichocki, Quantifying statistical interdependence by message passing on graphs-part II: multidimensional point processes, In: Neural Comput. (2009). https://doi.org/10.1162/neco.2009.11-08-899
[58] E. Gallego-Jutgl, J. Sol-Casals, F.B. Vialatte, J. Dauwels, A. Cichocki, A theta-band EEG based index for early diagnosis of Alzheimer's disease, In: J. Alzheimers Dis. (2015). https://doi.org/10.3233/JAD-140468
[59] B. Oltu, M.F. Akşahin, S. Kibaroglu, A novel electroencephalography based approach for Alzheimer's disease and mild cognitive impairment detection, In: Biomed. Signal Process. Control 63 (2021). https://doi.org/10.1016/j.bspc.2020.102223
[60] J. Lin, W. Huang, Automated diagnosis of mild cognitive impairment through connectivity analysis of EEG signals and a DL scheme, In: J. Eng. Appl. Sci. 72(1) (2025). https://doi.org/10.1186/s44147-025-00674-0
[61] A. Josefsson, A. Ibáñez, M. Parra, J. Escudero, Network analysis through the use of joint-distribution entropy on EEG recordings of MCI patients during a visual short-term memory binding task, In: Healthc Technol. Lett. (2019). https://doi.org/10.1049/htl.2018.5060
[62] C.A. Frantzidis, A.B. Vivas, A. Tsolaki, M.A. Klados, M. Tsolaki, P.D. Bamidis, Functional disorganization of small-world brain networks in mild Alzheimer's disease and amnestic mild cognitive impairment: an EEG study using relative wavelet entropy (RWE), In: Front. Aging Neurosci. (2014). https://doi.org/10.3389/fnagi.2014.00224
[63] H. Biglari, R. Boostani, M. Yazdi, S. Sanei, A new minimum spanning tree-based analysis of EEG and MEG with application to classification of Alzheimer's disease, In: IEEE Access 13 (2025) 119023. https://doi.org/10.1109/ACCESS.2025.3585196
[64] N. Mammone, C. Ieracitano, H. Adeli, A. Bramanti, F.C. Morabito, Permutation jaccard distance-based hierarchical clustering to estimate EEG network density modifications in MCI subjects, In: IEEE Trans. Neural Netw. Learn. Syst. (2018). https://doi.org/10.1109/TNNLS.2018.2791644
[65] P. Lo Giudice, N. Mammone, F.C. Morabito, R.G. Pizzimenti, D. Ursino, L. Virgili, Leveraging network analysis to support experts in their analyses of subjects with MCI and AD, In: Med. Biol. Eng. Comput. (2019). https://doi.org/10.1007/s11517-019-02004-y
[66] J. Zhang, J. Xia, X. Liu, J. Olichney, Machine learning on visibility graph features discriminates the cognitive event-related potentials of patients with early Alzheimer's disease from healthy aging, In: Brain Sci. (2023). https://doi.org/10.3390/brainsci13050770
[67] Y. Liu, J. Xia, Z. Kan, J. Zhang, S. Toprani, J.B. Brewer, M. Kutas, X. Liu, J. Olichney, Unveiling early signs of preclinical Alzheimer's disease through ERP analysis with weighted visibility graphs and ensemble learning, In: Bioengineering 12(8) (2025). https://doi.org/10.3390/bioengineering12080814
[68] C. Wang, Z. Wang, M.T. Herrero, T. Xu, F. Chu, H. Zeng, M. Tao, Early diagnosis of mild cognitive impairment and Alzheimer's disease using multimodal feature-based deep learning models in a chinese elderly population, In: Asian J. Psychiatr. (2025). https://doi.org/10.1016/j.ajp.2025.104632
[69] M.B. Babić Leko, M.K. Skoric, N. Klepac, F. Borovečki, L.L. Horvat, Z. Sonicki, P.R. Hof, G. Šimić, Event-related potentials improve the efficiency of cerebrospinal fluid biomarkers for differential diagnosis of Alzheimer's disease, In:Curr. Alzheimer Res. 15(13) (2018) 1244. https://doi.org/10.2174/1567205015666180911151116
[70] J.M. Olichney, J. Taylor, J. Gatherwright, D.P. Salmon, S.L. Bressler, M. Kutas, Patients with MCI and N400 or P600 abnormalities are at very high risk for conversion to dementia, In: Neurology 70(19) (2008) 1763–1770. https://doi.org/10.1212/01.wnl.0000281689.28759.ab
[71] S. Kongwudhikunakorn, S. Kiatthaveephong, K. Thanontip, P. Leelaarporn, M. Piriyajitakonkij, T. Charoenpattarawut, P. Autthasan, R. Chaisaen, P. Dujada, T. Sudhawiyangkul, V. Senanarong, T. Wilaiprasitporn, A pilot study on visually stimulated cognitive tasks for EEG-based dementia recognition, In: IEEE Trans. Instrum. Meas. 70 (2021). https://doi.org/10.1109/TIM.2021.3120131
[72] R.B. Berry, R. Brooks, C.E. Gamaldo, S.M. Harding, C.L. Marcus, B.V. Vaughn, The AASM manual for the scoring of sleep and associated events: updates for 2017 (version 2.4), In: J. Clin. Sleep Med. 13(5) (2017) 665–666. https://doi.org/10.5664/jcsm.6576
[73] M.K. Scullin, Sleep, memory, and aging: the link between slow-wave sleep and episodic memory changes from younger to older adults, In: Psychol. Aging 28(1) (2013) 105–114. https://doi.org/10.1037/a0028830
[74] B.A. Mander, S.M. Marks, J.W. Vogel, V. Rao, B. Lu, J.M. Saletin, S. Ancoli-Israel, W.J. Jagust, M.P. Walker, Beta-amyloid disrupts human NREM slow waves and related hippocampus-dependent memory consolidation, In: Nat. Neurosci. 18 (2015) 1051-1057. https://doi.org/10.1038/nn.4035
[75] L. Cui, Y. Wan, Y. Lu, M. Meng, N. Zhang, Sleep disturbances and their correlation with memory impairment in patients with Alzheimer's disease and amnestic mild cognitive impairment, In: J. Alzheimers Dis. Rep. (2025). https://doi.org/10.1177/25424823241309063
[76] D. Geng, C. Wang, Z. Fu, Y. Zhang, K. Yang, H. An, Sleep EEG-based approach to detect mild cognitive impairment, In: Front. Aging Neurosci. (2022). https://doi.org/10.3389/fnagi.2022.865558
[77] L. Gallego-Vinaras, J.M. Mira-Tomas, A.M. Gaeta, G. Pinol-Ripoll, F. Barbe, P.M. Olmos, A. Munoz-Barrutia, Alzheimer's disease detection in EEG sleep signals, In: IEEE J. Biomed. Health Inf. (2025). https://doi.org/10.1109/JBHI.2024.3478380
[78] S. Liu, C. Ao, Z. Li, L.H. Chio, Y. Gu, H. Zhan, H. Li, H. Li, Z. Li, Q. Wang, Development and validation of a polysomnography-based nomogram for predicting amnestic mild cognitive impairment in middle-aged and elderly people, In: Sleep Med. (2025). https://doi.org/10.1016/j.sleep.2025.106564
[79] K. Akrofi, M.C. Baker, M.W. O'Boyle, R.B. Schiffer, Clustering and modeling of EEG coherence features of Alzheimer's and mild cognitive impairment patients, In: Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (2008). https://doi.org/10.1109/IEMBS.2008.4649350
[80] J.L. Cantero, M. Atienza, A. Cruz-Vadell, A. Suarez-Gonzalez, E. Gil-Neciga, Increased synchronization and decreased neural complexity underlie thalamocortical oscillatory dynamics in mild cognitive impairment, In: Neuroimage (2009). https://doi.org/10.1016/j.neuroimage.2009.03.018
[81] G. Fiscon, E. Weitschek, A. Cialini, G. Felici, P. Bertolazzi, S. De Salvo, A. Bramanti, P. Bramanti, M.C. De Cola, Combining EEG signal processing with supervised methods for Alzheimer's patients classification, In: BMC Med. Inf. Decis. Mak. (2018). https://doi.org/10.1186/s12911-018-0613-y
[82] H.S. Ohal, S. Mantri, Exploring EEG-based biomarkers for improved early Alzheimer's disease detection: a feature-based approach utilizing machine learning, In: Meas. Sens. 36 (2024). https://doi.org/10.1016/j.measen.2024.101403
[83] B. Jalloul, S. Chaabene, B. Bouaziz, W. Mahdi, Empirical mode decomposition and attention-driven deep fusion of electroencephalographic and heart rate variability data for cognitive impairment detection, In: SN Comput. Sci. 6(7) (2025). https://doi.org/10.1007/s42979-025-04418-y
[84] M. Khosravi, H. Parsaei, K. Rezaee, M.S. Helfroush, Fusing convolutional learning and attention-based Bi-LSTM networks for early Alzheimer's diagnosis from EEG signals towards IoMT, In: Sci. Rep. (2024). https://doi.org/10.1038/s41598-024-77876-8
[85] B. Cansiz, H.O. Ilhan, N. Aydin, G. Serbes, Olfactory EEG based Alzheimer disease classification through transformer based feature fusion with tunable Q-factor wavelet coefficient mapping, In: Front. Neurosci. (2025). https://doi.org/10.3389/fnins.2025.1638922
[86] G. Grammenos, A.G. Vrahatis, P. Vlamos, D. Palejev, T. Exarchos, Alzheimer's Disease Neuroimaging Initiative, Predicting the conversion from mild cognitive impairment to Alzheimer's disease using an explainable AI approach, In: Information 15(5) (2024) 249. https://doi.org/10.3390/info15050249
[87] J. Snaedal, G.H. Johannesson, T.E. Gudmundsson, S. Gudmundsson, T.H. Pajdak, K. Johnsen, The use of EEG in Alzheimer's disease, with and without scopolamine a pilot study, In: Clin. Neurophysiol. (2010). https://doi.org/10.1016/j.clinph.2010.01.008
[88] J. Deng, B. Sun, V. Kavčič, M. Liu, B. Giordani, T. Li, Novel methodology for detection and prediction of mild cognitive impairment using resting-state EEG, In: Alzheimer's Dement. 20(1) (2024) 145. https://doi.org/10.1002/alz.13411

Downloaded for Anonymous User (n/a) at Auckland University of Technology from ClinicalKey.com.au by Elsevier on October 26, 2026. For personal use only. No other uses without permission. Copyright ©2026. Elsevier Inc. All rights reserved.