Check for updates

Biomarker Insights
Volume 21: 1–24
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11772719261496180
journals.sagepub.com/home/bmi
Sage

Multimodal AI is Advancing Biomarker Discovery in Parkinson's and Alzheimer's Disease: Evidence, Limitations, and the Translational Gap

Albert Śledzianowski[1], Artur Chudzik[1], and Andrzej W. Przybyszewski[1]

Abstract

Artificial intelligence is rapidly expanding the biomarker landscape for Parkinson's and Alzheimer's diseases. Multimodal systems combining neuroimaging, speech, gait, physiological signals, cognitive measures, and molecular markers show substantial research-stage performance. In studies with extractable accuracy and denominator data, reported Parkinson's disease classification accuracies ranged from 94.25% (95% Wilson CI: 93.78–94.68%) to 98.6% (95% Wilson CI: 95.71–99.67%), while Alzheimer's disease diagnostic or conversion-prediction accuracies ranged from 65.8% (95% Wilson CI: 62.54–68.87%) to 99.26% (95% Wilson CI: 97.89–99.75%). Therefore, the field is no longer limited by the absence of candidate AI biomarkers. However, the central bottleneck is translational validation. External validation remains uncommon across medical AI, with rates of 3.7% in high-impact medical informatics studies, 6% in diagnostic medical-imaging AI studies, and 14.7% in ICU AI scoring-system studies. In neurodegenerative disease, many models still rely on single cohorts, internal splits, incomplete metric reporting, and weak linkage between AI-derived digital phenotypes and established molecular or clinical biomarkers. Consequently, performance degradation in real-world populations, reduced specificity, calibration drift, and cohort bias help explain why extensive machine learning activity has not yet translated into routine clinical deployment. In this narrative review, we therefore synthesize quantitative evidence across motor, cognitive, neuropsychiatric, remote-monitoring, and neuroimaging domains. We argue that progress in AI for Parkinson's and Alzheimer's disease now depends less on additional model novelty alone than on standardized validation pipelines, external cohort testing, calibration assessment, interpretable multimodal fusion, and clinically meaningful endpoints. Nonetheless, with these safeguards, multimodal AI can move from promising biomarker discovery toward reproducible, clinically deployable precision diagnostics in neurodegenerative disease.

Plain Language Summary

This review explains why artificial intelligence may be useful for Parkinson's disease and Alzheimer's disease. Both diseases are complex and can affect the brain, movement, speech, thinking, mood, and everyday behavior. Because no single test captures this full picture, AI tools that combine many types of information may be especially helpful. The review shows that many AI tools already exist and often perform well in specific research settings. They can use information such as brain imaging, speech, gait, eye movements, cognitive tests, wearable sensors, smartphone and watch measurements, clinical data, and biological markers. These tools may help detect disease earlier, monitor changes over time, support risk prediction, and improve clinical decision-making. However, the main challenge is proving that these tools work reliably beyond the datasets in which they were first developed. Many studies are still based on small or single-center datasets, internal validation only, incomplete reporting, or limited testing in real-world clinical settings. Before multimodal AI can be used

[1]Faculty of Computer Science, Polish-Japanese Academy of Information Technology, Warsaw, Poland

Corresponding author:
Andrzej W. Przybyszewski, Faculty of Computer Science, Polish-Japanese Academy of Information Technology, 86 Koszykowa Street, Warsaw 02-008, Poland.
Email: przy@pja.edu.pl

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

widely in clinical care, it needs stronger testing on independent and larger datasets, better calibration, clearer interpretation, and well-defined clinical use cases.

Keywords
Narrative Review, Multimodal AI, Biomarkers, Parkinson's disease, Alzheimer's disease

Received: 29 May 2026; Revised: 18 August 2026; Accepted: 22 September 2026

Introduction

Parkinson's disease (PD) and Alzheimer's disease (AD) are progressive neurodegenerative disorders that often begin long before conventional clinical diagnosis. Motor, cognitive, affective, autonomic, sleep, imaging, and molecular abnormalities may emerge gradually and unevenly, while compensatory mechanisms can obscure the relationship between pathology and symptoms.[1-3] This creates a central diagnostic problem: early disease is rarely expressed through a single dominant marker, but through weak, heterogeneous, and temporally evolving signals distributed across multiple biological and behavioral domains.

Artificial intelligence (AI) is well suited to this problem because it can model complex patterns across digital and biological biomarkers, including neuroimaging, speech, gait, eye movements, wearable-sensor data, cognitive measures, neuropsychiatric symptoms, physiological signals, and molecular markers. In controlled research settings, machine learning and deep learning models have reported strong performance for disease classification, conversion prediction, subtype discovery, and progression monitoring. The field is therefore no longer limited by the absence of candidate AI biomarkers; across several domains, proof-of-principle performance has already been demonstrated.

The remaining challenge is translational. High discrimination in curated datasets does not by itself establish clinical usefulness. Many models still rely on single cohorts, internal cross-validation, incomplete metric reporting, limited calibration assessment, and weak evaluation under real-world conditions. This limitation is consistent with the broader medical AI literature, where external validation remains uncommon across high-impact medical informatics, diagnostic imaging, and ICU scoring-system studies.[4-6] For AD and PD, clinical AI biomarker development therefore requires a shift from model novelty alone toward external validation, calibration, missing-modality handling, interpretability, prospective testing, and clinical workflow integration.

This narrative review synthesizes evidence on AI-based biomarker discovery and translational validation in PD and AD. As summarized in Figure 1, we organize the literature from prodromal disease emergence and affective mechanisms through oculomotor signals, temporal-structural modeling, multimodal integration, cognitive and neuroimaging-based prediction, remote monitoring, and deployment readiness. The aim is not only to summarize model performance, but to identify which methodological conditions are needed for multimodal AI biomarkers to move from research-stage signal detection toward reproducible and clinically useful decision support.

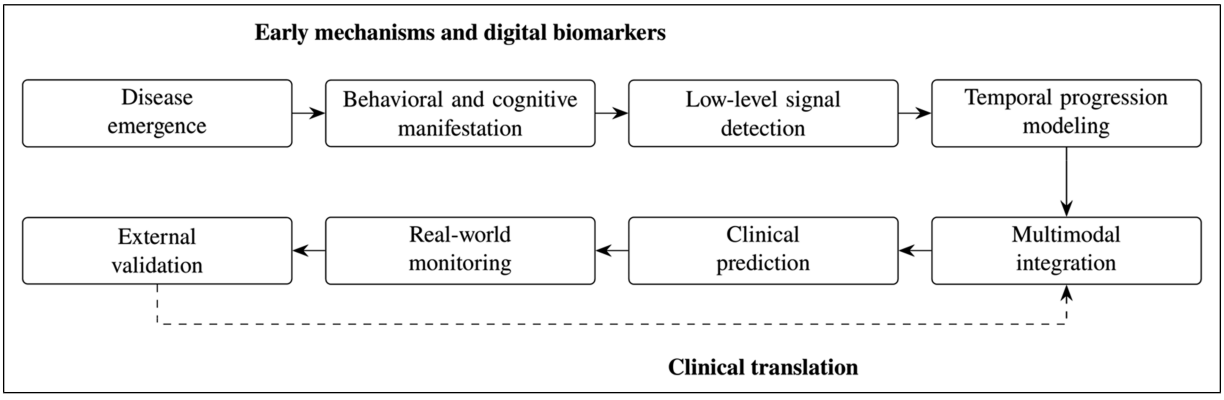


Figure 1. Narrative structure of the review. The paper progresses from disease emergence and behavioral/cognitive manifestation to low-level digital biomarker detection, temporal progression modeling, multimodal integration, clinical prediction, real-world monitoring, and external validation. The dashed feedback arrow indicates that validation should iteratively inform biomarker selection, model refinement, and clinical deployment

Methods

We conducted a narrative literature review focused on the use of artificial intelligence methods for biomarker discovery and translational validation in Parkinson's disease and Alzheimer's disease. The review was designed to synthesize evidence across heterogeneous biomarker domains, including neuroimaging, speech, gait, wearable and remote-monitoring signals, cognitive and neuropsychiatric measures, physiological signals, and molecular biomarkers. Because the literature spans multiple methodological traditions and outcome definitions, we prioritized conceptual synthesis and methodological appraisal rather than formal meta-analysis. The review was not registered as a systematic review protocol, and no pooled effect estimates were calculated.

The review was prepared with reference to the Scale for the Assessment of Narrative Review Articles (SANRA), which emphasizes the importance of a clear rationale, explicit aims, transparent literature search description, appropriate referencing, scientific reasoning, and presentation of relevant endpoint data in narrative reviews.[7] The present review therefore combines narrative synthesis with structured evidence extraction, while remaining distinct from a systematic review or meta-analysis. SANRA alignment was operationalized by defining the review rationale and aims in the Introduction, specifying the literature-discovery strategy and eligibility logic below, extracting study-level methodological and endpoint information into structured tables, and interpreting the evidence through a consistent translational-validation framework.

Search Strategy

Relevant publications were identified through iterative searches of PubMed, IEEE Xplore, and Google Scholar. Searches were not restricted by publication year, and the retained literature included publications from 1991 to 2026. Search terms combined disease-related keywords, such as "Parkinson's disease", "Alzheimer's disease", "neurodegenerative disease", "mild cognitive impairment", and "prodromal", with artificial intelligence and biomarker-related terms, including "machine learning", "deep learning", "artificial intelligence", "digital biomarker", "multimodal biomarker", "fusion", "external validation", "calibration", and "clinical translation". Modality-specific terms included "MRI", "fMRI", "PET", "EEG", "MEG", "speech", "voice", "gait", "wearable sensor", "smartphone", "remote monitoring", "eye tracking", "saccade", "cognitive assessment", "neuropsychiatric symptoms", "CSF", "genetics", and "molecular biomarker".

To improve coverage beyond database searches, we used complementary reference-management and citation-tracing strategies. Mendeley was used to manage references, remove duplicates, and organize studies by disease area, biomarker modality, and methodological focus. Litmaps was used to visualize citation relationships and identify clusters of related publications. In addition, backward citation mining was performed by screening the reference lists of key papers, and forward citation mining was used to identify more recent studies citing highly relevant publications.

Search results and citation-tracing outputs were screened for relevance to four evidence roles: conceptual and pathophysiological background; primary AI or machine learning biomarker studies; review-level syntheses; and methodological or translational-validation evidence. Records were retained for detailed discussion when they were directly relevant to AI-based biomarker discovery, multimodal integration, disease progression, remote monitoring, or translational validation in PD or AD. Studies and reviews with sufficient methodological detail were summarized in the supplementary evidence tables to support cross-domain comparison of datasets, modalities, AI methods, validation strategies, performance metrics, clinical applications, and limitations.

Eligibility

Studies were considered eligible for detailed discussion if they applied artificial intelligence, machine learning, deep learning, or related computational methods to biomarker discovery, disease classification, disease-stage prediction, conversion prediction, monitoring, or clinical outcome prediction in Parkinson's disease, Alzheimer's disease, or related prodromal conditions. We included studies that examined digital, behavioral, neuroimaging, physiological, cognitive, neuropsychiatric, or molecular biomarkers and reported quantitative performance metrics such as accuracy, sensitivity, specificity, area under the receiver operating characteristic curve, F1-score, correlation with clinical measures, or prediction error.

Studies were prioritized when they reported sufficient methodological detail regarding the data source, population, biomarker modality, model type, validation strategy, and clinical prediction task. We excluded publications that did not involve AI-based analysis, did not focus on Parkinson's disease or Alzheimer's disease, lacked extractable performance or validation information, or were purely technical reports without clear clinical or biomarker relevance.

Data Extraction

For studies summarized in the main text and supplementary tables, we extracted information on disease area, biomarker modality, data source, sample characteristics, AI methodology, prediction task, validation approach, reported performance metrics, and whether external validation was performed. The extracted evidence is summarized by domain in Supplemental Tables 1–4. Particular attention was given to methodological factors relevant to clinical translation, including the use of independent test cohorts, cross-validation design, risk of data leakage, cohort size, class imbalance, reporting completeness, calibration assessment, interpretability, and linkage between AI-derived features and established clinical or biological markers.

For accuracy values highlighted in the abstract, 95% Wilson confidence intervals were calculated descriptively from the reported accuracy and participant denominator. Correct classifications were estimated as the nearest integer to reported accuracy multiplied by the denominator. These intervals were not pooled and should be interpreted as study-level uncertainty estimates rather than meta-analytic confidence intervals.

Synthesis

The evidence was synthesized qualitatively, with emphasis on biomarker modality, clinical task, and methodological rigor. Rather than ranking studies solely by reported model performance, we interpreted findings in relation to validation strength, reproducibility, external generalizability, and translational relevance. This approach allowed us to compare evidence across heterogeneous domains while highlighting recurring methodological limitations that may restrict clinical deployment. Because the review was narrative, the supplementary tables should be interpreted as structured evidence summaries rather than exhaustive systematic-review inventories. The synthesis therefore begins with the biological reason this field matters clinically: AD and PD may be detectable before conventional diagnostic thresholds are reached.

Prodromal Disease Emergence

Alzheimer's disease (AD) and Parkinson's disease (PD) are neurodegenerative diseases (NDs), a group of disorders in which selected populations of neurons progressively lose function and die. Because NDs become more common with age, they are an increasing source of disability in aging societies. Their clinical consequences include memory loss, disorientation, impaired daily functioning, tremor, slowness of movement, gait disturbance, and affective or cognitive changes. Yet these visible symptoms usually appear late in the biological course of disease.

This delay creates one of the central problems in AD and PD care. By the time a patient reaches the clinic with clear dementia or parkinsonism, the underlying process may have been active for many years, and substantial brain or nervous-system changes may already have accumulated.[1-3] Early detection is therefore not only a diagnostic ambition; it is a prerequisite for prevention trials, disease-modifying treatment, and timely clinical planning. If reliable tools could detect AD or PD before classical symptoms become obvious, they could change the point at which clinicians enter the disease course.

The evidence increasingly suggests that such early signals do exist. In AD, biomarker frameworks distinguish the underlying pathophysiological process from the later clinical syndrome, showing that amyloid, tau, neurodegeneration, and subtle cognitive change can precede dementia by many years.[1,8] Longitudinal evidence supports the clinical relevance of this construct: in 311 cognitively normal older adults, CSF-defined preclinical AD stages were associated with increasing 5-year progression to symptomatic AD, from 2% in biomarker-normal individuals to 11%, 26%, and 56% across stages 1, 2, and 3, respectively.[9]

PD shows a similarly long but clinically broader prodromal interval. Before cardinal motor symptoms allow a standard clinical diagnosis, patients may show REM sleep behaviour disorder, olfactory loss, constipation and other autonomic features, subtle motor dysfunction, affective symptoms, and abnormal dopaminergic imaging.[2,3,10] Retrospective data from 1,467 patients further illustrate this temporal spread: loss of smell, sleep problems, fatigue, stress, and bladder or bowel problems were among the features most often reported more than five years before diagnosis, whereas tremor and bradykinesia clustered closer to the diagnostic transition.[11]

The difficulty is that prodromal disease is not simply a smaller version of diagnosed disease. Brain plasticity, cognitive reserve, compensatory learning, and threshold effects can obscure the relationship between pathology and symptoms. In AD, cognitive reserve theory proposes that education, occupation, leisure activity, neural reserve, and neural compensation can allow individuals with similar clinical status to carry different pathological burdens.[12] In PD, prodromal trajectories can differ in their motor, autonomic, cognitive, affective, imaging, genetic, and pathological profiles, motivating subtype and probability-based approaches rather than a single-marker diagnostic model.[10] Thus, the central challenge is not only to measure early abnormalities, but to interpret many weak, heterogeneous, and partly compensated signals together.

This compensation can be understood as a learning problem. Some adaptation is guided by explicit feedback, similar to supervised learning, while other adaptation depends on reinforcement learning: actions are selected and modified according to reward, prediction, and environmental feedback.[13] Because dopamine is central to reinforcement learning and is also depleted in PD, prodromal PD may disturb not only movement but also adaptation, motivation, reward-guided behaviour, and emotional stability. Related social reinforcement-learning disturbances have also been discussed in AD in association with medial temporal and parietal atrophy.[14] This learning perspective helps explain why no two patients show exactly the same clinical pattern during a long prodromal interval.

This is where multimodal and pattern-based AI becomes clinically relevant. The question is no longer whether prodromal signals can be measured, but whether they can be combined into reliable and clinically interpretable tools. In PD, similarity network fusion has shown that MRI morphometry, dopamine transporter binding, CSF assays, and clinical measures can be integrated into continuous phenotypic axes that capture patient heterogeneity more parsimoniously than discrete clinical categories.[15] In prodromal PD specifically, a recent scoping review found that multimodal imaging and AI-driven approaches reported sensitivities of 43–84% and specificity up to 96%, while emphasizing that clinical implementation remains constrained by limited validation and standardization.[16]

One way to formalize this problem is to treat prodromal disease as a similarity-recognition task. This view is close to the logic of pattern recognition in biological vision. Object recognition does not depend on a single feature, but on similarity between the current sensory pattern and stored internal representations in higher visual areas.[17] By analogy, a person who is still clinically classified as normal may already show a fuzzy pattern of attributes that resembles the pattern observed in more advanced disease. The working hypothesis is therefore straightforward: if rules extracted from disease progression are meaningful, they should show similarity to rules describing more advanced disease states. In this framework, the question is not whether one symptom crosses a diagnostic threshold, but whether the whole attribute pattern is moving toward a disease-like state.

Przybyszewski and Śledzianowski applied this idea to PD by comparing attributes of patients at different stages of progression with model patterns derived from more advanced disease.[18] This work emphasized that PD progression is not uniform across motor, emotional, and cognitive domains; individual patients may show different dominant problems even when they share the same diagnostic label.[19] Such heterogeneity is precisely why prodromal prediction requires pattern-based models rather than simple symptom counting.

A related approach was developed for AD using cognitive attributes rather than the PD-specific clinical features. Instead of monitoring only a single dominant marker, the method examined patterns across multiple cognitive test results and compared them with patterns observed in MCI and AD. Rough set theory and granular computing were used to represent these attribute patterns as granules and to generate rules describing similarity between clinically normal subjects, MCI, and AD patients.[20-22]

In this AD model, general granules were first derived from normal subjects, MCI, and more advanced AD patients. These granules produced rules that separated broad clinical stages, but they also allowed a second, more clinically interesting step: comparison between AD-related granules and granules from individuals still classified as healthy. Similarity to AD-like granules suggested possible early neurodegenerative change even when standard neurological and neuropsychological evaluation remained normal. The Clinical Dementia Rating sum of boxes (CDRSUM) was used as a clinically accepted decision attribute.[23] In this setting, some normal subjects were classified by AD rules as already resembling mild dementia, illustrating how a pattern-based approach may reveal prodromal structure before conventional clinical thresholds are crossed.[21,22] A further aim of this logic is to look for universal classifications, meaning similar clinical conclusions emerging from different but related sets of rules.

This logic belongs to a broader interpretable pattern-based literature rather than to a single research line. In AD and dementia, rough-set and fuzzy-rule approaches have been used to model diagnostic uncertainty in MRI, cognitive, and clinical data. Cheng and Liu combined hybrid MRI segmentation, wavelet-packet features, and a rough-set LEM2 classifier for degenerative brain disease classification, reporting 99.89% and 99.36% internal accuracy on T1-and T2-weighted image sets, respectively, but also showing how strongly performance depended on image quality, segmentation, and user-defined regions of interest.[24] Mazarbhuiya et al proposed a neighborhood rough-set classifier for NACC-derived AD records, using both medical-history and cognitive-function attributes and reporting approximately 96–98% internal accuracy compared with substantially lower SVM, KNN, and naive Bayes comparators.[25] Chen et al approached the same interpretability problem from a fuzzy-systems direction: dominant-set clustering was used to initialize compact fuzzy rules, which were then optimized in an adaptive neuro-fuzzy inference system (ANFIS) framework for OASIS-2 dementia diagnosis, yielding human-readable rules, mean cross-validated accuracy of 81.69%, precision of 82.7%, recall of 74.7%, and specificity of 87.3%.[26] At the review level, fuzzy deep learning has also been proposed as a way to combine neuroimaging, fuzzy segmentation, uncertainty handling, and explainable decision support in AD.[27]

Comparable ideas also appear in PD. Revett et al used rough sets for feature selection in voice-based PD classification, illustrating an early attempt to connect feature relevance, rule extraction, and dysphonia-based decision support.[28] Sarin et al later developed fuzzy classifiers for dynamic handwriting analysis across ParkinsonHW, PaHaW, and NewHandPD, explicitly framing the method around interpretable IF–THEN rules and the trade-off between accuracy, compactness, and clinical usability; the best implementation reached 96.67% accuracy on ParkinsonHW, whereas performance dropped to 69.33% on PaHaW, underscoring the importance of dataset, task, medication state, and generalizability.[29] These studies are heterogeneous in modality, scale, and validation quality, but together they support the same central point: rough, granular, and fuzzy models are useful because prodromal and early neurodegenerative disease often appears as a graded, uncertain, and multidomain pattern rather than as a crisp diagnostic threshold.

The strength of this broader approach is that it matches the biological problem: prodromal AD and PD are heterogeneous, partially compensated, and distributed across multiple domains. Its current limitation is also the broader limitation of the field. High prediction accuracy in small or internally resampled datasets does not automatically translate into broad clinical coverage or external validity. The next step is therefore not only to build more sensitive models, but to test whether rule-based and AI-derived prodromal patterns remain reliable in larger, prospectively followed cohorts, including online and remote-testing designs, across populations and clinical settings.

Affective Mechanisms

Affective symptoms are often described as psychiatric comorbidities, but in AD and PD this description is too narrow. Depression, anxiety, apathy, irritability, sleep disturbance, and hallucinations can also be part of the neurodegenerative phenotype itself. They may appear before diagnosis, accompany the earliest clinical phase, or track progression after diagnosis. Mood and motivation are brain functions, and the neural systems that support them can be affected by the same disease processes that later produce memory loss, tremor, and motor slowing.

This idea is especially clear in PD. Depression affects a large proportion of patients and is several times more common than in the general population.[30,31] Importantly, it can precede motor diagnosis rather than only follow it. In the Rotterdam Study, motor and non-motor features were examined up to 23 years before PD diagnosis, and anxiety and depression were detected about five years before clinical diagnosis.[32] In newly diagnosed PD, the ICICLE-PD cohort showed that depression, anxiety, apathy, and hallucinations were more frequent in PD than in controls across 36 months; higher baseline motor severity was associated with increasing neuropsychiatric burden, and higher neuropsychiatric burden was associated with poorer quality of life.[33] These findings support the view that affective symptoms are not merely reactive distress, but clinically meaningful components of early and progressing PD.

The biological plausibility of this relationship is strong. Early PD affects not only the posterior putamen and motor circuit, but also associative and limbic basal-ganglia circuits. In untreated early PD, depression, apathy, sleep disturbance, and anxiety are common, and depression and apathy have been linked to ventral striatal and mesolimbic dopaminergic denervation, with serotonergic and noradrenergic systems also implicated.[34] Reviews of PD depression similarly emphasize dopaminergic, serotonergic, and noradrenergic dysfunction; cortico-limbic, striato-thalamic-prefrontal, and mediotemporal-limbic network disruption; and possible contributions from neuroinflammation, neuroimmune dysregulation, stress hormones, and metabolic factors.[31,35] Thus, affective symptoms in PD are mechanistically connected to reward learning, motivation, autonomic regulation, and the broader non-motor network pathology of the disease.

In AD, the relationship is more difficult to interpret because depression and cognitive decline overlap clinically. Symptoms such as loss of interest, psychomotor slowing, fatigue, impaired concentration, sleep disturbance, and appetite change can arise from depression, dementia, aging, medical illness, or their combination.[36] This overlap is one reason diagnosis is difficult. However, it is also why affective symptoms may carry useful signal, especially in late-onset AD, where late-life depression has been discussed as a risk factor, prodromal marker, and clinically important coexisting pathway rather than a purely reactive state.[37] Engedal et al found that sadness, anxiety, suicidal thoughts, poor self-esteem, multiple physical complaints, and pessimism distinguished AD patients with depression from those without depression.[38] Lyketsos et al proposed a broader classification of Alzheimer's disease-associated neuropsychiatric disturbance, including affective symptoms such as depression, irritability, anxiety, and euphoria together with associated symptoms such as aggression, agitation, delusions, hallucinations, sleep disturbance, and appetite disturbance.[39]

Longitudinal AD and MCI studies support this broader disease-signal view. Migliorelli et al distinguished major and minor depression in AD, with major depression more likely to begin before dementia onset and minor depression more often emerging after cognitive decline.[40,41] In a population-based cohort of 2,220 older adults, depressive symptoms were associated with a higher probability of MCI, with increasing rates across negligible, low, and moderate/high depressive-symptom groups.[42] Neuropsychiatric symptoms are also common across AD populations, with depression, apathy, anxiety, irritability, sleep disturbance, and related symptoms contributing to clinical heterogeneity and downstream

functional status.[43] In MCI, Teng et al found that all participants who progressed to AD had neuropsychiatric symptoms at baseline, compared with 59% of those who remained stable or improved; depression and apathy were more common among progressors.[44] In a much larger NACC cohort of 4,932 participants with MCI, apathy alone increased risk of conversion to AD, and the combination of apathy and depression increased risk further, whereas depression alone was not a significant predictor after adjustment.[45] These findings suggest that apathy, depression, and anxiety should be interpreted as separable but interacting features, not collapsed into one nonspecific psychiatric label.

Biomarker and imaging studies sharpen this interpretation. In non-demented older adults, Johansson et al found that apathy and anxiety, but not depressive symptoms, were associated with cerebral amyloid-β accumulation and longitudinal cognitive decline; apathy was also associated with atrophy and white-matter lesions, while anxiety interacted with amyloid-β status to predict faster cognitive decline.[46] In a later Swedish BioFINDER analysis of 356 cognitively unimpaired older adults followed biennially for up to eight years, baseline amyloid-β pathology was associated with increasing apathy and anxiety, and the amyloid effect on apathy was only partly mediated by cognitive decline.[47] Belleville et al similarly concluded that neuropsychiatric symptoms, including depression, apathy, anxiety, and sleep disturbances, can contribute to early AD prediction when combined with cognition and neuroimaging, although their sensitivity is generally lower than cognitive measures.[48] The emerging message is not that affective symptoms are specific enough to diagnose AD alone, but that they may improve risk estimation when placed inside a multimodal pattern.

This is the point at which granular-computing logic becomes useful again. In the PD progression project, depression was treated not only as a comorbidity but as an attribute that might change the rule structure of disease prediction.[49] Patients were analyzed across medication-only, recent deep brain stimulation (DBS), and longer-term DBS groups, with UPDRS values followed across three visits separated by approximately six months. Rules derived from the first medication-only visit predicted later medication-only states more accurately when Beck Depression Inventory results were included: accuracy increased from 0.70 to 0.765 for the second visit and from 0.70 to 0.80 for the third visit. The same logic appeared in cross-group prediction, where prediction of pre-surgical DBS status improved from 0.50 to 0.64 with Beck scores. DBS-derived rules also predicted later DBS and long-term DBS states with variable accuracy, including DBSW3, POPW1, POPW2, and POPW3 at 0.625, 0.77, 0.50, and 0.33, respectively. Although the dataset was small and coverage was limited, the result is conceptually important: depression improved prediction because it captured clinically relevant variance that motor scores alone did not fully explain.

Recent machine learning evidence reaches a similar conclusion in AD. Shah et al tested whether neuropsychiatric symptoms improved ML prediction of cognitive decline when added to demographic and AD biomarker data. Across community-dwelling older adults, adding neuropsychiatric symptoms improved prediction in several global and domain-specific cognitive outcomes, particularly when combined with neuroimaging biomarkers, although effects were small and domain-specific.[50] This supports the broader argument of this review: affective variables should not be discarded as noise simply because they are clinically complex. Their complexity may be exactly why pattern-based AI can use them, provided that models are externally validated and interpreted cautiously.

Affective symptoms therefore form a natural bridge between prodromal disease and lower-level digital biomarkers. They are higher-level clinical expressions of disrupted neural systems, reflecting changes in reward, motivation, stress, sleep, autonomic regulation, and cognition. Oculomotor signals, discussed next, represent a lower-level and more continuous behavioral output of neural control systems. Both belong in a multimodal biomarker framework because neurodegeneration leaves traces at multiple levels: in mood and motivation, in cognition and movement, and in the fine temporal structure of sensorimotor behavior.

Oculomotor Signals

Oculomotor signals constitute an important source of low-level temporal information in biomarker analysis, as they directly reflect the dynamics of eye movement control, which is strongly linked to the functioning of neural systems. In contrast to high-level clinical indicators, they exhibit a continuous and high-resolution temporal nature, enabling precise modeling of subtle changes in visual behavior.

Incorporating the temporal structure of oculomotor signals provides information about the microdynamics of perceptual-motor processes, such as saccade latency, amplitude, velocity, acceleration, and fixation stability. This allows for the identification of patterns that are not visible in static analysis and the discrimination of cases with similar global parameters but different underlying dynamics, which is of significant prognostic relevance.

Temporal models learn not only the distribution of signal values but also their temporal structure, including fluctuations, oscillations, and both short- and long-term dependencies. As a result, they can capture oculomotor control instability and subtle deviations from normal behavior that often emerge in the early stages of disease.

This temporal approach enables a more precise representation of neurodegenerative processes manifested in eye movement control. In particular, in prodromal phases where changes are minor and irregular, analysis of low-level signals can increase detection sensitivity and prediction stability by exploiting information embedded in the temporal structure of the data.

The reviewed literature can be grouped into several methodological directions based on how eye-tracking signals are represented and modeled.

End-to-End Temporal Sequence Modeling (Raw Time-Series Signals)

End-to-end temporal models ask whether diagnostic information is lost when eye movements are reduced too early to summary descriptors. This is the strongest argument for preserving raw saccadic waveforms or fixation-interval dynamics: disease-related abnormalities may be carried not only by mean latency, amplitude, or velocity, but by the shape and ordering of the signal itself.

Two recent examples illustrate this direction. A hierarchical machine learning framework for Parkinsonian disorder classification used full saccadic waveforms to distinguish PD, progressive supranuclear palsy (PSP), and healthy controls.[51] A related deep time-series approach applied InceptionTime and ROCKET-based models to fixation-interval sequences and achieved up to 96% classification accuracy under an internal subject-exclusive split.[52] Read together, these studies suggest that raw temporal representations can capture information that handcrafted summaries may discard. Their shared limitation is interpretability and validation strength: learned representations are harder to translate into familiar clinical descriptors, and external generalizability remains unestablished.

Feature-Based Machine Learning and Ensemble Methods (Engineered Descriptors)

Feature-based approaches make the opposite trade-off. They deliberately compress eye-tracking recordings into interpretable descriptors, then use classical machine learning or ensemble methods to combine them. This can be advantageous in clinical datasets, where sample sizes are often modest and clinicians may want to know whether saccade, fixation, pupil, or blink features are driving a prediction.

This logic is visible in PD screening work that combines handcrafted visual-exploration features with a Mixture of Experts ensemble, achieving strong internal discrimination (AUC $\approx$ 0.95).[53] It is also reflected in video-based eye-tracking classification and staging, where engineered saccade, pupil, and blink features support classical machine learning models.[54] These studies show that deep learning is not inherently superior when feature engineering captures clinically meaningful dynamics. The cost is that aggregation may smooth away subtle waveform structure that raw-sequence models can exploit.

Unsupervised Learning and Multivariate Statistical Modeling

Unsupervised and multivariate approaches are useful when the clinical question is exploratory or labels are scarce. Rather than learning a conventional disease classifier, these methods estimate reference behavioral structure, detect deviations from that structure, or quantify joint abnormalities across several oculomotor variables.

One-class anomaly detection has been used to model Parkinsonian fixation behavior without requiring balanced labelled datasets; in the reported GANomaly configuration, samples that depart from the learned Parkinsonian fixation representation are treated as outliers.[55] Multivariate statistical modeling takes a more interpretable route by jointly analyzing latency, damping, amplitude, and related saccadic metrics.[56] The common strength of these methods is sensitivity to distributed abnormalities that may be weak in any single feature. Their common vulnerability is dependence on the reference distribution: if recording conditions, cohort composition, or task structure shift, the learned abnormality profile may not transport cleanly.

Biomarker-Driven and Correlational Clinical Analysis (Non-Predictive Modeling)

Not all useful oculomotor work is predictive. Biomarker-driven studies test whether specific eye-movement features track motor subtype, axial impairment, cognition, or disease severity. They therefore provide physiological plausibility for later AI modeling, even when they do not themselves build complex classifiers.

Video-oculographic analysis of PD motor subtypes shows that saccade amplitude, velocity, and square-wave jerks vary systematically across clinical phenotypes.[57] Memory-guided saccade latency has also been linked to freezing of gait and axial motor impairment,[58] while tablet-based eye-movement assessment connects oculomotor metrics with cognition and disease severity.[59] These correlational studies are not designed primarily as predictive classifiers, but they help anchor AI features in

clinically interpretable neurobiology. Their limitation is that association with phenotype does not by itself establish diagnostic prediction, external generalizability, or incremental value over standard clinical assessment.

Summary

Across the reviewed studies, both classical and deep learning approaches can benefit from modeling temporal structure in oculomotor signals. Methods operating on raw saccadic waveforms or fixation-interval time series can preserve fine-grained dynamics of eye movement control that are partly lost in static feature summaries, while engineered-feature and ensemble approaches remain competitive when descriptors capture clinically meaningful oculomotor properties.

The main comparative distinction is therefore not deep learning versus classical machine learning, but representation strategy. Raw-sequence models are strongest when high-resolution waveforms are available and the goal is to preserve temporal microstructure, but they sacrifice interpretability and require careful subject-level validation. Engineered-feature and ensemble models are more transparent and can perform well in smaller cohorts, but may discard subtle timing information. Unsupervised or anomaly-detection approaches are attractive when labels are limited, whereas biomarker-driven correlational studies provide physiological plausibility without necessarily establishing predictive utility.

Across studies, saccade latency, amplitude, fixation stability, and related eye-movement measures repeatedly emerge as biomarkers linked to motor and cognitive aspects of Parkinson's disease.

However, methodological limitations are recurrent. As shown in Supplemental Table 2, many studies rely on single-center datasets, lack external validation, or report incomplete evaluation metrics, which limits the generalizability of results. Despite strong predictive performance in controlled settings, the clinical transferability of these models remains constrained by dataset heterogeneity and evaluation design. This limitation leads directly to the broader temporal problem: if neurodegeneration unfolds through trajectories rather than isolated measurements, models must distinguish true disease dynamics from short-range signal structure or dataset-specific regularities.

Temporal-Structural Modeling

Neurodegenerative diseases are not static classification problems. AD and PD unfold through changing trajectories: symptoms fluctuate, biomarkers evolve, compensatory mechanisms fail gradually, and clinically similar patients may diverge over months or years. Temporal-structural modeling addresses this problem by treating disease as a dynamic process rather than a single cross-sectional snapshot.

The central idea is simple but important. A model can use not only the value of a biomarker at one visit, but also the direction, speed, and pattern of change across observations. These changes may be expressed directly as first-order increments (Δx), time-normalized changes ($\Delta x / \Delta t$), or higher-order dynamics ($\Delta^2 x$), which describe acceleration or deceleration of pathological processes. In modern deep learning models, however, these temporal relations are often learned implicitly through hidden states, recurrent memory, temporal convolutions, or ordered feature sequences.

This distinction matters most in early and prodromal disease, where abnormalities are often weak, nonlinear, and distributed across modalities. A patient may look nearly normal at one time point but already show an abnormal trajectory in gait, cognition, imaging, speech, or clinical history. Temporal models therefore promise to detect disease-relevant change before static thresholds are crossed. At the same time, not every study that uses a sequence model captures true disease progression. The reviewed literature spans a methodological spectrum, from genuine longitudinal modeling to local sliding-window analysis and artificial sequencing of static data.

Explicit Sequence Modeling (RNN/LSTM/GRU and Real Temporality)

The strongest temporal formulation uses genuinely sequential or longitudinal data. In recurrent neural networks, long short-term memory networks, and gated recurrent units, hidden states encode previous observations and allow predictions to depend on history rather than isolated measurements. This architecture matches the clinical problem: in AD and PD, trajectory may be more informative than status at a single visit.

The evidence spans several kinds of real temporality. In gait analysis, LSTM models applied to vertical ground-reaction-force sequences from PhysioNet achieved 98.6% accuracy for binary PD classification and 96.6% accuracy for multiclass severity rating, illustrating how recurrent models can exploit dependencies within gait cycles.[60] In routine healthcare data, LSTM screening models used longitudinal diagnostic histories to identify PD risk up to four years before diagnosis, reaching 94.25% accuracy with AUC values ranging from 0.839 to 0.923 in a single national claims cohort.[61] In AD, recurrent models have been used to predict disease progression and future biomarker values from longitudinal

neuroimaging and biomarker data, with reported AUC of approximately 0.90 and lower mean absolute error than baseline approaches.[62] Together, these studies move temporal modeling closer to the clinical question of how a patient's disease process is likely to evolve.

These examples are clinically persuasive because the data themselves contain temporal order. Their weakness is practical rather than conceptual. Sequence models are sensitive to sequence length, visit timing, missing observations, measurement noise, and cohort-specific care patterns. In the reviewed studies, strong internal discrimination was not matched by independent external validation or complete reporting of calibration and complementary metrics, limiting claims about transportability.

Hybrid Spatiotemporal Modeling (CNN–RNN)

Hybrid CNN–RNN architectures combine local pattern extraction with temporal or ordered dependency modeling. Convolutional components learn spatial, segment-level, or frame-level representations; recurrent components then model how those representations unfold. This design is most persuasive when both parts of the architecture correspond to real structure in the data.

In gait-based PD assessment, CNN–LSTM modeling of vertical ground-reaction-force signals allows local waveform structure and temporal evolution to be learned jointly, extends naturally from diagnosis toward severity estimation, and has been reported to improve performance over classical approaches.[63] In fMRI-based AD classification, CNN–LSTM models link spatial brain features with ordered functional representations; Farhan et al reported approximately 99% accuracy with near-perfect AUC in several configurations in an ADNI cohort of 413 subjects, but this was based on internal 10-fold cross-validation only and external generalizability was not established.[64] A more ambiguous case arises when structural MRI scans are converted into video-like sequences for AD, PD, and control classification.[65] This approach has reported very high internal performance and can improve representation learning across ordered slices, but it should not be interpreted as direct evidence that disease progression over time has been modeled.

Hybrid spatiotemporal models therefore sit between mechanistic plausibility and technical convenience. They can integrate local structure with temporal or ordered dependencies, but their interpretation depends on what the ordered dimension actually represents. When the order corresponds to gait cycles or functional time series, temporal claims are stronger; when it corresponds to constructed image slices, the model is better understood as representation learning.

Pseudo-Sequencing (Sequencing of Static Data)

Pseudo-sequencing refers to cases in which static data, such as MRI slices or extracted feature vectors, are arranged into an ordered sequence so that recurrent or temporal models can be applied. This can be valuable, but it answers a narrower question than longitudinal modeling: whether ordered representations improve classification, not whether patient-level disease trajectories have been learned.

MRI-based AD studies illustrate the distinction. Ebrahimi et al extracted MRI slice features with a ResNet-18 backbone and then modeled the resulting ordered representations with RNN, LSTM, GRU, and temporal convolutional networks, with the TCN reaching 91.78% accuracy, 91.56% sensitivity, and 92% specificity under internal evaluation.[66] Salehi et al similarly reported high binary-classification performance using LSTM networks for MRI-based AD detection, including AUC and accuracy values near the upper 0.90 range under internal resampling.[67] These results support the technical usefulness of ordered image representations. They do not, however, establish longitudinal validity, because the recurrent memory operates across constructed image order rather than repeated observations of the same patient over disease time.

The limitation of pseudo-sequencing is therefore conceptual as much as methodological. Such models may be strong classifiers, but they should be presented as representation-learning approaches rather than as direct models of neurodegenerative progression.

Local Memoryless Temporal Modeling (CNN + Sliding Window)

Local temporal modeling uses convolutional architectures, often with sliding windows, to capture short-range patterns in time-series data. Unlike recurrent models, these methods do not maintain a global state across the full sequence. They can be highly effective when local waveform shape is informative, but they are less suited to modeling patient-level history or long-range task structure.

This distinction matters for sensor-based neurodegenerative disease classification. Multi-branch CNNs applied to functional-mobility sensor signals can learn local movement dynamics for AD, PD, and control classification through sliding-window segmentation and 1D convolutions.[68] Related CNN approaches include accelerometer-based AD stage

identification, where reported accuracy reached 90.91%,[69] and gait-dynamics classification across neurodegenerative disorders using a twin-layered CNN.[70] In neuroimaging, dynamic functional brain networks from rs-fMRI have been modeled with spatiotemporal graph convolution and attention for AD-related cognitive dysfunction and PD subtyping.[71] Across these examples, the gain comes from local or graph-structured temporal patterning rather than full longitudinal memory.

Hybrid approaches try to bridge this gap. For example, EMD preprocessing combined with CNN–LSTM modeling has been used for PD gait analysis, with reported internal accuracy of 98.32%.[72] The broader tradeoff is consistent across domains: local CNNs can be efficient and discriminative, while recurrent or hybrid models are better aligned with long-range dependency when the data genuinely contain it.

Summary

Across these studies, temporal-structural modeling can improve the representational richness of AI biomarkers when the temporal or ordered structure is meaningful. The strongest case is made when the data are genuinely longitudinal or sequential, as in gait cycles, repeated clinical histories, or longitudinal AD biomarker trajectories. Hybrid CNN–RNN models add value when spatial or local signal structure must be integrated with temporal organization. Pseudo-sequencing and sliding-window CNNs can also improve classification, but they answer a narrower question: they often learn ordered structure within data rather than disease progression itself.

Comparatively, recurrent models are most clinically persuasive when the prediction target is progression, conversion, or future severity, because their architecture matches the longitudinal nature of the disease process. CNN-based sliding-window models may outperform simpler models for high-frequency sensor data because they capture local waveform motifs efficiently, but they do not necessarily model patient-level disease history. Pseudo-sequential MRI or fMRI models should therefore be interpreted primarily as representation-learning methods, not as direct evidence that temporal disease trajectories have been learned.

The translational message is therefore balanced. Temporal models are well matched to neurodegenerative disease because AD and PD are dynamic processes. Nonetheless, many studies summarized in Supplemental Table 1 still lack external validation, rely on single datasets, and emphasize accuracy while omitting F1-score, calibration, or full AUC reporting. Consequently, temporal modeling should be viewed as a promising route toward clinically meaningful progression biomarkers, but not as a substitute for rigorous validation across independent cohorts, devices, and clinical settings. It also leaves a further question open: because disease progression is rarely expressed in one signal alone, temporal structure must often be integrated with imaging, clinical, cognitive, molecular, and behavioral information.

Multimodal Integration

Multimodal integration is the computational step that turns heterogeneous biomarkers into a unified diagnostic or prognostic representation. In neurodegenerative disease, this is particularly important because no single modality captures the full clinical phenotype: imaging may reflect structural or functional brain change, speech and gait capture motor and temporal control, cognitive testing measures clinical impairment, and biological markers provide molecular context. A scoping review of clinically accessible multimodal learning studies in AD and PD identified 46 papers including 11,750 participants, illustrating both the rapid growth of this field and its emphasis on combining imaging, cognitive, speech/language, gait, hand-movement, eye-movement, genetic, and other clinical data.[73] Multimodal AI models therefore seek to reduce diagnostic ambiguity by combining complementary information from MRI, PET, EEG, speech, genetic markers, clinical records, and behavioral data into shared representations.

Deep multimodal models are especially attractive because they can learn hierarchical features directly from raw or weakly processed inputs and can model nonlinear relationships across modalities. CNNs are typically used for spatial feature extraction, transformers for cross-modal attention, graph neural networks for relational learning, and recurrent or temporal modules for longitudinal dependencies. Attention and graph-based mechanisms can also support interpretability by identifying which modalities, regions, or relationships contribute most strongly to prediction. In principle, such systems can capture subtle disease signatures that are missed by unimodal approaches, especially in early disease stages when single-modality signals are weak.

The design question is where integration occurs. Fusion can happen at the input level, in an intermediate latent representation, at the decision level, or within hybrid architectures that combine spatial, temporal, and cross-modal dependencies.

Early Fusion

Early fusion integrates heterogeneous modalities at the input or near-input stage, before a shared model performs representation learning. This can involve channel-wise stacking of images, concatenation of minimally processed features, or construction of a single input vector from multiple sources. The strength of early fusion is simplicity: all modalities are presented to the model simultaneously, allowing a unified representation to be learned from the beginning. Its weakness is that it assumes reasonable alignment, scale compatibility, and completeness across modalities.

Odusami et al proposed an explainable deep learning framework for AD diagnosis using multimodal PET and MRI fusion.[74] The model combined metabolic information from PET with structural information from MRI using input-level fusion and achieved approximately 0.74 accuracy on ADNI, outperforming unimodal baselines while providing feature-attribution-based explainability. This channel concatenation or image-level fusion is efficient, but inter-modal relationships are learned implicitly rather than modeled explicitly.

Venugopalan et al used ADNI MRI, SNP genetic information, and clinical or neuropsychological measures for classification of AD, MCI, and cognitively normal groups.[75] MRI was modeled with a 3D-CNN, while SNP and clinical data were encoded through stacked denoising autoencoders. Modality-specific embeddings were then concatenated before classification. The multimodal model outperformed MRI-only, clinical-only, and SNP-only models in several tasks and improved recall and F1-score for more difficult MCI classification, although the evaluation remained within ADNI-derived splits rather than an independent external cohort.

Leony et al compared early and late fusion strategies for AD classification, with emphasis on interpretability.[76] In the early fusion setting, features from multiple modalities were concatenated, including weighted and hybrid weighted variants that assigned different importance to modalities before classification with traditional machine learning models. Weighted early fusion achieved balanced accuracy of 0.920, slightly exceeding late fusion at 0.912 in internal evaluation, and interpretability analysis highlighted the dominant contribution of cognitive and neuropsychological measures. These findings suggest that early fusion can perform well when feature harmonization is strong, but it remains sensitive to missing data, scale differences, and modality imbalance.

Intermediate Fusion

Intermediate or feature-level fusion processes each modality through a dedicated branch and combines learned embeddings in a shared latent space. This offers a balance between modality-specific learning and cross-modal integration. It allows each modality to retain an appropriate encoder while still giving the model an opportunity to learn interactions before final prediction. The cost is greater architectural complexity and higher risk of overfitting, especially in small medical datasets.

Dentamaro et al used PPMI MRI and clinical data to classify prodromal PD using multimodal deep learning with an Excitation Network for joint co-learning.[77] The strongest model, DenseNet combined with the Excitation Network, achieved 96.6% accuracy, 98.9% AUC, and 96.5% F1-macro in stratified tenfold cross-validation (internal validation only), outperforming most unimodal alternatives and supporting the value of combining imaging with clinical features. Importantly, the study also used explainable AI to identify regions such as right temporal and left prefrontal areas, linking model outputs to plausible prodromal pathophysiology rather than treating multimodal performance as a black-box result.

Jha et al proposed an attention-based multimodal deep learning framework for early neurodegenerative disease prediction across healthy control, MCI, AD, and PD groups.[78] The paper reports integration of structural MRI, functional connectivity/fMRI, DTI, PET, clinical biomarkers, and digital/behavioral biomarkers through modality-specific feature pathways followed by attention-based intermediate fusion. It reports a primary cohort of 850 participants, including 180 participants with PD, and a separate retrospective external-validation cohort of 180 ADNI-3 participants. Reported performance was 96.8% overall accuracy and AUC 0.959 internally, with 95.1% accuracy on the ADNI-3 validation cohort. These findings illustrate the intended role of attention-based multimodal fusion, although clinical relevance still depends on prospective validation, transparent cohort description, and reproducibility in independent settings.

Singh et al proposed an end-to-end deep learning framework integrating MRI, PET, EHR, and genomic data for AD, MCI, and control classification.[79] The study compared fusion strategies and found that intermediate multiplicative fusion performed best on external validation, outperforming unimodal and partial-modality baselines. Performance was stronger in younger age groups and declined in older cohorts, underscoring an important translational issue: multimodal models may improve prediction while still remaining sensitive to cohort age, disease stage, and population structure.

Late Fusion

Late fusion combines predictions or high-level outputs from independently trained modality-specific models. Averaging, voting, gated attention, stacking, and meta-learning can all be used at this stage. The advantage is that each modality can be optimized separately, reducing inter-modality interference and preserving modality-specific information. The limitation is that cross-modal interactions are modeled weakly, because integration occurs only near the final decision.

Ahmed et al proposed a multimodal deep learning framework for AD prediction using MRI–PET fusion.[80] Structural MRI was processed with a 3D-ResNet, PET was processed with a dedicated 3D CNN, and gated attention combined modality-level embeddings at a late stage. The paper reports evaluation across ADNI, AIBL, and OASIS-3, with accuracy up to 98.7%, F1-score up to 0.894, and AUC values from 0.92 to 0.96. Compared with unimodal and early-fusion baselines, the model is reported to improve accuracy by approximately 5–6% and reduce false positives, especially in early MCI detection. These values should be interpreted as benchmark results rather than definitive evidence of clinical readiness.

Zhang et al developed a multimodal graph-neural-network framework for AD diagnosis using sMRI, PET, and phenotypic data.[81] Each modality was processed in a separate graph branch, with brain images represented within a population graph framework. Information was partially integrated through node representations and graph structures, and the final diagnostic decision was produced through late fusion of branch outputs. Combining sMRI, PET, and phenotypic data improved performance over unimodal and CNN-based baselines, and ablation studies showed that removing any modality degraded performance, especially for NC–MCI–AD staging.

Overall, late fusion often performs well because it reduces error propagation between modalities and allows modality-specific optimization. However, because interactions are introduced late, these models may miss subtle synergistic disease patterns that depend on relationships among modalities rather than independent modality-level evidence.

Hybrid Spatiotemporal Fusion

Hybrid spatiotemporal multimodal models combine multimodal fusion with explicit temporal or relational modeling. These architectures may use CNNs for spatial representation, transformers for attention, recurrent networks for temporal structure, and graph neural networks for inter-modality or population relationships. They are especially relevant when the clinical question involves not only diagnosis but progression, staging, or personalized prediction.

Kaliappan et al discussed a broad HybridNet-NDM-style architecture for predictive analytics and healthcare management in neurodegenerative diseases, combining CNNs, LSTMs, graph convolutional networks, and attention-based fusion.[82] The paper is better interpreted as a general hybrid neural-network framework for neurodegenerative-disease analytics than as direct evidence for AD-specific classification performance. Its relevance here is architectural: it illustrates how spatial, temporal, graph-based, and attention mechanisms may be combined in multimodal systems, while leaving disease-specific validation as an open requirement.

Anand et al developed a multimodal framework for neurodegenerative disease detection using EEG, fMRI, MEG, and genomic data.[83] EEG was processed with time–frequency and wavelet methods, fMRI with independent component and correlation analysis, MEG with beamforming and source localization, and genomic data with graph neural networks. Extracted representations were fused with a GRU-based mechanism and classified with a 1D CNN. The model improved performance over existing methods, with reported gains in precision, accuracy, recall, AUC, specificity, and detection delay, suggesting that temporal and cross-modal fusion can strengthen detection of complex disease patterns.

Summary

Across these studies, multimodal models generally outperform unimodal approaches by leveraging complementary biomedical information. Supplemental Table 3 compares the reviewed multimodal studies and reviews by data source, sample size, modality combination, AI method, clinical task, reported metrics, validation strategy, and main limitation. Early, intermediate, and late fusion differ mainly in the stage of integration: raw input concatenation, latent feature alignment, or final decision aggregation. Early fusion can be effective when modalities are complete, harmonized, and naturally aligned, but it is vulnerable to missing data and scale imbalance. Intermediate fusion is often better suited to heterogeneous medical data because each modality can have its own encoder before cross-modal learning, although this increases model complexity and overfitting risk. Late fusion is robust when modality-specific models are strong or data completeness varies, but it may miss synergistic interactions that only appear when modalities are modeled jointly. This broader evidence is consistent with Huang et al, who found promising performance across clinically accessible multimodal AD and PD studies but emphasized small samples, missing or noisy modalities, and insufficient validation as recurrent barriers.[73]

At the progression level, Jiang et al reviewed 59 studies using at least two modalities, including clinical, genetic, biomarker, and neuroimaging variables, to predict PD progression.[84] They found that integrating multiple modalities often improved prediction compared with fewer modalities under similar conditions, but also emphasized that gains depend on the modality combination, prediction target, and validation design. In prodromal PD, Serag et al similarly found encouraging performance across multimodal imaging and AI-driven diagnostic approaches, while concluding that screening use remains limited by lack of validation and standardization.[16] Thus, multimodal integration is not automatically a route to clinical readiness. Its value depends on whether models can handle missing modalities, preserve interpretability, remain calibrated, and generalize across cohorts, devices, and care settings. The most developed testing ground for these questions is cognitive and neuroimaging-based prediction, where high internal performance and external transportability can be compared directly.

Cognitive and Neuroimaging-Based Prediction

Cognitive and neuroimaging-based models occupy a central position between mechanistic biomarker discovery and clinical prediction. In AD, this literature is dominated by MCI-to-AD conversion, structural MRI, FDG-PET, amyloid and tau markers, and ADNI-derived multimodal datasets. In PD, comparable approaches increasingly use MRI, DAT-SPECT, clinical scales, cognitive measures, and PPMI-derived biomarkers to predict prodromal status, cognitive decline, motor progression, and subtype evolution.

The Dentamaro et al. PPMI study provides a useful PD example because it explicitly combines structural MRI with clinically accessible features for prodromal-stage classification rather than relying on imaging alone.[77] In AD, cross-cohort studies show why this domain is clinically attractive but methodologically fragile: models trained on ADNI or related cohorts can achieve high internal performance, yet external validation across cohorts or acquisition environments often reveals performance loss, especially for MCI conversion prediction.[85-88] Thus, cognitive and neuroimaging-based prediction should be treated not only as a high-performing diagnostic domain, but also as a stress test for calibration, cohort bias, and the transportability of learned disease signatures.

Clinical Translation and Deployment Readiness

The stress test raised by cognitive and neuroimaging prediction becomes the central deployment question for the whole field. The clinical translation of multimodal AI biomarkers in Parkinson's disease and Alzheimer's disease depends not only on high diagnostic or prognostic performance in curated research datasets, but also on evidence that these models remain reliable under real-world clinical conditions. Across neurodegenerative disease research, strong internal performance often contrasts with limited external validation, heterogeneous acquisition protocols, cohort and demographic bias, preprocessing variability, missing modalities, and insufficient integration into clinical workflows. Deployment readiness therefore requires a broader evaluation framework in which discrimination is interpreted together with ecological validity, external generalizability, calibration, interpretability, and prospective clinical utility.

This challenge is not specific to neurodegenerative disease. Marconi and Cabitza found that only 3.7% of machine learning studies in selected high-impact medical informatics journals reported external validation, and only 0.01% combined external validation with uncertainty quantification.[4] In diagnostic medical imaging, Kim et al found that only 6% of 516 AI performance studies used external validation, and none of the externally validated studies combined diagnostic cohort design, multi-institutional validation, and prospective external data collection.[5] In ICU AI scoring systems, Rockenschaub et al found external validation in 84 of 572 studies (14.7%); within their publication-year trend analysis, the annual validation rate increased from 7.1% among 2018 studies to 23.9% among 2022 studies, with external validation associated with lower AUROC on average.[6] These cross-field findings support treating external validation, calibration, and uncertainty reporting as core requirements for clinically deployable AI biomarkers rather than optional methodological refinements.

Remote monitoring and external validation represent two complementary dimensions of this translational pathway. Remote monitoring tests whether digital biomarkers can capture clinically meaningful signals during daily life rather than only during standardized clinical visits. External validation tests whether AI models maintain their performance when applied to independent cohorts, different clinical sites, distinct devices, and heterogeneous patient populations. Supplementary Table 4 summarizes representative evidence across these two deployment-oriented domains. Together, these dimensions determine whether multimodal AI biomarkers can move beyond research accuracy toward reproducible and clinically useful decision-support tools.

Remote Monitoring and Ecological Validity

Remote monitoring has become a central strategy for improving the ecological validity of digital biomarkers in neurodegenerative disease. In contrast to episodic clinic-based assessments, smartphone- and wearable-based systems allow repeated measurement of motor, speech, gait, dexterity, sleep, cognitive, and behavioral signals in naturalistic settings. This is particularly relevant in PD, where motor symptoms fluctuate with medication state, fatigue, context, and time of day. Remote monitoring can therefore provide a denser and potentially more representative picture of disease burden than single in-clinic evaluations. Wearable sensor systems have been used to monitor tremor, bradykinesia, dyskinesia, and freezing of gait during activities of daily living in PD.[89]

Remote and web-based cognitive assessment can also enrich established screening instruments rather than replacing them. Tools such as the Montreal Cognitive Assessment (MoCA) are clinically useful because they summarize multiple cognitive domains in a short test, but their interpretation is affected by education, culture, practice effects, and the limited information contained in a final score. Chudzik and Przybyszewski therefore provide a useful example of how conventional cognitive assessment can be augmented with behavioral digital biomarkers: they tested machine learning classification of PD using MoCA responses combined with online task-process variables, including instrumental reaction time and submission time.[90] In this small study, adding these web-assessment variables improved the best reported diagnostic accuracy from 80.0% to 93.4% and specificity from 57.2% to 93.4%. This positions online cognitive testing as a way to recover clinically relevant response behavior that is lost when only the raw cognitive score is analyzed.

For deployment-ready remote monitoring, however, the strongest evidence currently comes from PD motor assessment. Burq et al demonstrated that smartwatch-based active motor assessments can provide frequent and reliable remote measurements of PD motor function in a longitudinal cohort of 388 early-stage patients monitored over a median of 390 days. Digital motor features showed moderate-to-strong correlations with clinical ratings, including tremor correlation of $\rho = 0.70$, bradykinesia correlation of $\rho = -0.62$, gait correlation of $\rho = -0.46$, and good-to-excellent test–retest reliability with intraclass correlation coefficients ranging from 0.75 to 0.96.[91] Importantly, in-clinic assessments often did not reflect patients' typical status at home, suggesting that home-based monitoring can capture clinically relevant variability missed by episodic clinical examination.

Similarly, Lipsmeier et al evaluated the Roche PD Mobile Application in 316 early-stage PD participants. The application combined smartphone active tests with passive smartphone and smartwatch monitoring to measure bradykinesia, bradyphrenia and speech, tremor, gait, and balance. The study reported excellent adherence of 96.29%, good-to-excellent test–retest reliability with median intraclass correlation coefficient of 0.9, and correlations between digital features and corresponding Movement Disorder Society–Unified Parkinson's Disease Rating Scale items.[92] These findings support the feasibility of smartphone- and smartwatch-based assessments as scalable tools for longitudinal monitoring.

Speech-based biomarkers also provide a mature example of remote PD assessment. Tsanas et al showed that noninvasive speech tests can estimate PD progression with clinically useful accuracy using approximately 6000 voice recordings from 42 patients.[93] Because speech can be captured remotely, repeatedly, and at low cost, it represents a promising modality for tracking disease severity outside specialized clinical environments. Smartphone-based passive and active monitoring extends this approach further by combining voice, balance, gait, dexterity, and reaction-time assessments. Zhan et al collected large-scale smartphone monitoring data from 226 participants, including more than 46,000 hours of passive monitoring and approximately 8000 structured tests, and reported medication-response detection accuracy of approximately 71%, suggesting that mobile sensing can capture clinically relevant fluctuations in PD state.[94]

Gait-based AI also links remote monitoring with clinical decision support. Wu et al reviewed AI-based gait assessment for PD diagnosis, freezing-of-gait detection and prediction, and gait-management support, emphasizing the value of wearable and non-wearable gait systems for continuous monitoring while noting that reliable clinical deployment still depends on improved data collection, labeling, generalization, and standardization.[95]

Compared with PD, the evidence base for AD-specific remote monitoring is broader but less standardized. Digital technologies for AD and mild cognitive impairment (MCI) increasingly include passive activity monitoring, digital cognitive assessment, speech and language analysis, sleep and circadian rhythm tracking, and home-based behavioral sensing. However, AD lacks the same direct sensor-accessible motor phenotype that makes PD particularly suitable for wearable monitoring. As a result, AD remote-monitoring studies more often emphasize feasibility, behavioral inference, and early cognitive change rather than deployment-ready digital endpoints validated against longitudinal clinical outcomes. This asymmetry suggests that remote monitoring is currently closer to clinical readiness in PD motor and speech assessment than in AD progression monitoring.

Thus, remote monitoring should be viewed not only as a convenient data-collection strategy, but also as a translational stress test. Digital biomarkers that remain reliable in daily life, under variable medication states, adherence patterns, sensor conditions, and environmental contexts, are more likely to support clinically meaningful deployment.

External Validation, Calibration, and Clinical Utility

External validation is essential for determining whether AI biomarker models generalize beyond the datasets in which they were developed. Internal cross-validation or random train–test splits can overestimate performance when data are drawn from the same cohort, acquisition protocol, preprocessing pipeline, or clinical environment. For clinical translation, models must be evaluated on truly independent populations that differ in recruitment criteria, disease severity, demographics, scanner or device type, clinical site, and data-processing procedures. This is particularly important in multimodal AI, where performance may depend strongly on site-specific measurement patterns or incomplete modality availability.

Accordingly, validation should be interpreted as a graded continuum rather than a binary property: random internal splits provide the weakest evidence, subject-level cross-validation is stronger, independent-cohort validation is stronger still, and prospective multi-site validation with predefined clinical endpoints provides the most relevant evidence for deployment. For this reason, deployment-ready AI biomarker studies should use standardized validation pipelines that predefine preprocessing, train–test separation, missing-modality handling, calibration assessment, subgroup analysis, and clinically meaningful endpoints.

Several AD studies illustrate both the feasibility and difficulty of external validation. Hall et al evaluated the Disease State Index for predicting progression from MCI to AD across multiple independent cohorts, including DESCRIPA, ADNI, AddNeuroMed, and Kuopio MCI, showing that model performance can vary across populations and clinical settings.[85] Birkenbihl et al similarly demonstrated that differences between cohort studies affect external validation of AI models for dementia prediction; nevertheless, external validation of an ADNI-trained model on 244 AddNeuroMed subjects remained feasible and achieved useful predictive performance.[86] Bron et al further investigated cross-cohort generalizability of deep and conventional machine learning for MRI-based AD diagnosis and prediction, highlighting the importance of evaluating models across independent datasets rather than relying only on within-cohort performance.[87] Kim et al also provided an example of imaging-environment validation by training an FDG-PET/CT model on ADNI data and testing it on an independent Severance Hospital dataset, reporting comparable classification accuracy across datasets.[88]

In PD, validation concerns are especially visible in neuroimaging-based AI. Dzialas et al reviewed 244 neuroimaging-based AI studies in PD and found substantial methodological quality problems, including data leakage, excessive model complexity, limited biological plausibility, insufficient external testing, and weak reporting of clinical utility. Only 20% of evaluated studies passed all five minimal quality criteria, and only 8% employed external test sets.[96] Such findings indicate that high reported accuracy alone is insufficient for clinical readiness. Models must also demonstrate robustness to dataset shift, transparent methodology, biologically plausible decision patterns, and reproducible performance in independent samples.

Recent PD-focused reviews converge on the same limitation. Serag et al found that prodromal PD diagnostic studies using multimodal imaging and AI-driven approaches reported sensitivity from 43–84% and specificity up to 96%, but concluded that screening use remains limited by lack of validation.[16] Jiang et al similarly reviewed multimodal PD progression prediction and found that only 8 of 59 studies used an independent external validation dataset, despite frequent performance gains from adding modalities.[84]

Performance degradation under external validation may affect not only overall accuracy or AUC, but also clinically critical operating characteristics such as specificity, sensitivity, and calibration at the decision thresholds used for referral, monitoring, or treatment stratification. Calibration is therefore another critical but often underreported requirement. A model with high discrimination may still be clinically unreliable if predicted probabilities are poorly calibrated. For example, a model may correctly rank high-risk and low-risk patients while systematically overestimating absolute risk. In clinical decision-making, this distinction matters because diagnostic thresholds, referral decisions, treatment selection, and monitoring intensity depend on the reliability of predicted probabilities. Therefore, multimodal AI biomarker evaluation should include calibration plots, Brier scores, expected calibration error, and assessment of calibration drift across external cohorts.

Explainable AI can improve transparency, but it should not be treated as a substitute for validation. Methods such as SHAP, LIME, Grad-CAM, and related attribution approaches may help identify which imaging, speech, motor, cognitive, or molecular features drive predictions. These tools can support clinician trust, reveal implausible shortcuts, and expose site- or device-specific confounding. However, feature attribution does not guarantee clinical validity. Explainability must therefore be integrated with external validation, calibration analysis, and prospective clinical utility testing.

Clinical utility represents the final step in deployment readiness. A model may be accurate and externally validated but still fail to improve clinical decisions. Utility should be evaluated using clinically meaningful endpoints, such as earlier diagnosis, improved prediction of conversion from MCI to AD, better detection of PD progression, more accurate medication-response monitoring, improved patient stratification for trials, reduced assessment burden, or improved allocation of specialist care.

Decision-curve analysis, net-benefit assessment, prospective workflow studies, and clinician-in-the-loop evaluation are therefore needed to determine whether multimodal AI biomarkers improve outcomes beyond standard clinical assessment.

Overall, the evidence suggests that remote monitoring is closest to clinical readiness in PD motor and speech assessment, whereas AD remote monitoring remains promising but less standardized. Conversely, AD provides several instructive examples of cross-cohort external validation, especially in MCI-to-AD prediction and neuroimaging-based classification. Across both diseases, the central translational challenge is no longer simply whether AI can achieve high accuracy in curated datasets, but whether multimodal biomarkers can remain reliable, calibrated, interpretable, and clinically useful across the heterogeneous conditions of real-world care.

Discussion

The preceding sections point to a consistent pattern: AI models can detect disease-relevant signals across prodromal, affective, oculomotor, temporal, multimodal, cognitive, neuroimaging, and remote-monitoring domains, but the same evidence repeatedly exposes a translational bottleneck. Strong performance in controlled settings often contrasts with limited external validation, dataset bias, incomplete calibration assessment, and uncertain real-world transportability.

The argument of this review begins with prodromal disease emergence because AD and PD are not static diagnostic categories. They unfold through long, heterogeneous, and partly compensated trajectories in which early abnormalities may appear across cognition, mood, movement, sleep, autonomic function, imaging, and molecular markers before conventional clinical thresholds are crossed. This biological structure explains why single-marker approaches are often insufficient and why pattern-based AI, including granular-computing and multimodal methods, is attractive: the clinically relevant signal may lie in the configuration of many weak abnormalities rather than in one dominant feature.

Affective and oculomotor findings extend this logic across levels of observation. Depression, apathy, anxiety, and related neuropsychiatric symptoms should not be treated only as nonspecific comorbidities, because they may reflect disrupted reward, motivation, stress, sleep, and limbic-cortical networks involved in AD and PD progression. Oculomotor signals, by contrast, provide lower-level and temporally precise measures of sensorimotor control. Together, these domains show that AI biomarkers are most informative when they connect high-level clinical expression with low-level dynamic signals and when they preserve temporal structure rather than reducing disease to a single cross-sectional label.

Temporal-structural modeling and multimodal integration represent the computational response to this biological problem. Recurrent, convolutional, graph-based, attention-based, and fusion architectures can model trajectories, combine complementary modalities, and improve classification or progression prediction in research datasets. However, the same complexity that makes these models powerful also increases the risk of overfitting, leakage, opaque decision rules, and poor generalization across cohorts, devices, acquisition protocols, and clinical populations. Multimodal AI is therefore not automatically closer to clinical readiness than unimodal AI; it becomes clinically meaningful only when additional modalities improve robust, calibrated, and interpretable prediction under realistic conditions.

Taken together, the reviewed evidence suggests that the field has moved beyond the proof-of-principle stage. The central question is no longer whether AI can identify candidate biomarkers in AD and PD, but whether these models can be converted into reliable clinical systems. AI models can detect disease-relevant structure across several biomarker domains, but the next phase requires a different evidentiary standard, focused on predefined use cases, stable data pipelines, prospective validation, calibration monitoring, subgroup performance, missing-modality handling, and integration into real clinical workflows.

This deployment phase should also clarify what AI biomarkers are expected to do. A clinically useful system may not need to replace diagnosis; it may support risk stratification, earlier referral, longitudinal monitoring, trial enrichment, medication-response assessment, or prioritization of specialist care. Each use case implies different requirements for sensitivity, specificity, calibration, interpretability, acceptable false-positive burden, and clinical actionability. Therefore, deployment readiness should be judged not only by discrimination metrics, but by whether the model provides reliable information at a defined point in the care pathway and improves decisions beyond standard assessment.

Emerging multi-disease models further indicate the potential to address overlapping clinical symptoms across neurodegenerative conditions. This is relevant because memory impairment, depression, apathy, gait disturbance, sleep disruption, and oculomotor abnormalities may cut across diagnostic boundaries. However, cross-disease modeling remains underexplored and should be developed cautiously, with explicit attention to differential diagnosis, disease-stage imbalance, comorbidity, and the risk that broad models may obscure clinically important subtype-specific mechanisms.

Overall, bridging the translational gap will require standardized methodologies, external validation, calibration assessment, uncertainty reporting, and clinically meaningful evaluation frameworks. This is the practical condition for moving from promising biomarker discovery to deployable clinical support. Multimodal AI in AD and PD should therefore be understood as a validation, workflow, and clinical-utility problem as much as a modeling problem.

Limitations

This review has two main limitations. First, it is a narrative review rather than a registered systematic review or meta-analysis. The retained evidence base was selected to support conceptual synthesis and methodological appraisal across heterogeneous biomarker domains, rather than to estimate pooled effects. The tables should therefore be read as structured summaries of representative and methodologically relevant evidence rather than as an exhaustive evidence inventory.

Second, the literature is expanding quickly, especially for remote monitoring, digital biomarkers, and multimodal foundation-model-style architectures. In these areas, publication volume often exceeds the availability of stable external validation, calibration reporting, or prospective clinical-utility evidence. The remote-monitoring section therefore emphasizes studies with clearer links to ecological validity, longitudinal measurement, or clinical deployment rather than attempting a comprehensive catalogue of every wearable, smartphone, speech, or home-sensing study.

Conclusion

The reviewed evidence indicates that artificial intelligence has already established proof-of-principle value for biomarker discovery in Parkinson's disease and Alzheimer's disease. Across digital, behavioral, neuropsychiatric, temporal, multimodal, cognitive, neuroimaging, and remote-monitoring domains, AI models can detect disease-relevant structure that is difficult to capture with isolated clinical measures.

The most important conclusion is therefore not simply that more accurate models are needed, but that the field must now enter a validation- and deployment-focused phase. Clinical adoption will depend on standardized validation pipelines, independent and prospective testing, calibration and uncertainty assessment, interpretable multimodal integration, robust handling of missing or noisy modalities, and evaluation against clinically meaningful endpoints. Multimodal AI will transform neurodegenerative-disease care only if it moves from promising signal detection toward reliable, workflow-compatible, and clinically useful decision support.

ORCID iDs

Albert Śledzianowski https://orcid.org/0000-0002-3592-6829
Artur Chudzik https://orcid.org/0000-0002-0993-4630
Andrzej W. Przybyszewski https://orcid.org/0000-0002-0156-7856

Ethical Considerations

Not applicable. This article is a narrative review of previously published literature and did not involve new studies with human participants or animals performed by the authors.

Consent for Publication

Not applicable. This article does not contain individual patient data, patient images, or participant-level material requiring consent for publication.

Author Contribution

Albert Śledzianowski: Conceptualization, Investigation, Methodology, Writing - original draft, Writing - review & editing. Artur Chudzik: Conceptualization, Investigation, Methodology, Visualization, Writing - original draft, Writing - review & editing. Andrzej W. Przybyszewski: Conceptualization, Investigation, Methodology, Supervision, Writing - original draft, Writing - review & editing.

Funding

The authors received no financial support for the research, authorship, or publication of this article.

Declaration of Conflicting Interests

The authors declare that they have no conflicts of interest.

Data Availability Statement

Not applicable. This article is a narrative review and does not report original datasets or newly generated data.

Supplemental Material

Supplemental material for this article is available online.

References

1. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:280-292. doi:10.1016/j.jalz.2011.03.003.
2. Postuma RB, Berg D. Advances in markers of prodromal Parkinson disease. *Nat Rev Neurol*. 2016;12:622-634. doi:10.1038/nrneurol.2016.152.
3. Rees RN, Noyce AJ, Schrag A. The prodromes of Parkinson's disease. *Eur J Neurosci*. 2019;49:320-327. doi:10.1111/ejn.14269.
4. Marconi L, Cabitza F. Show and tell: A critical review on robustness and uncertainty for a more responsible medical AI. *Int J Med Inform*. 2025;202:105970. doi:10.1016/j.ijmedinf.2025.105970.
5. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. *Korean J Radiol*. 2019;20:405-410. doi:10.3348/kjr.2019.0025.
6. Rockenschaub P, Akay EM, Carlisle BG, et al. External validation of AI-based scoring systems in the ICU: a systematic review and meta-analysis. *BMC Med Inform Decis Mak*. 2025;25:5. doi:10.1186/s12911-024-02830-7.
7. Baethge C, Goldbeck-Wood S, Mertens S. SANRA-a Scale for the Quality Assessment of Narrative Review Articles. *Res Integr Peer Rev*. 2019;4:5. doi:10.1186/s41073-019-0064-8.
8. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement*. 2018;14:535-562. doi:10.1016/j.jalz.2018.02.018.
9. Vos SJ, Xiong C, Visser PJ, et al. Preclinical Alzheimer's disease and its outcome: a longitudinal cohort study. *Lancet Neurol*. 2013;12:957-965. doi:10.1016/S1474-4422(13)70194-7.
10. Berg D, Borghammer P, Fereshtehnejad S-M, et al. Prodromal Parkinson disease subtypes — key to understanding heterogeneity. *Nat Rev Neurol*. 2021;17:349-361. doi:10.1038/s41582-021-00486-9.
11. Schrag A, Zhelev SS, Hotham S, Merritt RD, Khan K, Graham L. Heterogeneity in progression of prodromal features in Parkinson's disease. *Parkinsonism Relat Disord*. 2019;64:275-279. doi:10.1016/j.parkreldis.2019.05.013.
12. Stern Y. Cognitive Reserve and Alzheimer Disease. *Alzheimer Dis Assoc Disord*. 2006;20:S69-S74. doi:10.1097/00002093-200607001-00010.
13. Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press; 1998. https://mitpress.mit.edu/9780262193986/reinforcement-learning/
14. Legaz A, Abrevaya S, Dottori M, et al. Multimodal mechanisms of human socially reinforced learning across neurodegenerative diseases. *Brain*. 2022;145:1052-1068. doi:10.1093/brain/awab345.
15. Markello RD, Shafiei G, Tremblay C, Postuma RB, Dagher A, Misic B. Multimodal phenotypic axes of Parkinson's disease. *NPJ Parkinsons Dis*. 2021;7:6. doi:10.1038/s41531-020-00144-9.
16. Serag I, Azzam AY, Hassan AK, et al. Multimodal diagnostic tools and advanced data models for detection of prodromal Parkinson's disease: a scoping review. *BMC Med Imaging*. 2025;25:103. doi:10.1186/s12880-025-01620-5.
17. Przybyszewski AW. SI: SCA Measures – Fuzzy rough set features of cognitive computations in the visual system. *Journal of Intelligent & Fuzzy Systems*. 2019;36:3155-3167. doi:10.3233/JIFS-18401.
18. Przybyszewski AW, Śledzianowski A. Parkinson's disease development prediction by c-granule computing compared to different AI methods. *Journal of Information and Telecommunication*. 2020;4:425-439. doi:10.1080/24751839.2020.1749410.
19. Przybyszewski AW. Fuzzy RST and RST Rules Can Predict Effects of Different Therapies in Parkinson's Disease Patients. In: *Foundations of Intelligent Systems*. Springer; 2018:409-416. doi:10.1007/978-3-030-01851-1_39.
20. Pawlak Z. *Rough sets: Theoretical aspects of reasoning about data*. Boston, MA: Kluwer Academic Publisher; 1991. doi:10.1007/978-94-011-3534-4.
21. Przybyszewski AW. Multi-granular Computing Can Predict Prodromal Alzheimer's Disease Indications in Normal Subjects. In: *Computational Science – ICCS 2023*. Springer; 2023:278-285. doi:10.1007/978-3-031-36024-4_22.
22. Przybyszewski AW, Nowacki JP, Drabik A, et al. Pattern Classification Can Detect Prodromal Alzheimer's Neurodegenerations. In: *Pattern Recognition. ICPR 2024 International Workshops and Challenges*. Springer; 2025:224-239. doi:10.1007/978-3-031-87660-8_17.
23. O'Bryant SE, Waring SC, Cullum CM, et al. Staging dementia using Clinical Dementia Rating Scale Sum of Boxes scores: a Texas Alzheimer's Research Consortium study. *Arch Neurol*. 2008;65(8):1091-1095. doi:10.1001/archneur.65.8.1091.
24. Cheng C-H, Liu W-X. Identifying Degenerative Brain Disease Using Rough Set Classifier Based on Wavelet Packet Method. *J Clin Med*. 2018;7:124. doi:10.3390/jcm7060124.
25. Mazarbhuiya FA, Shenify M, Wungreiphi AS, Kez J, Paul R. Detecting Alzheimer's Disease using Rough Set-based Classification Approach. *Engineered Science*. 2025;37:1736. doi:10.30919/es1736.

26. Chen T, Su P, Shen Y, et al. A Dominant Set-informed Interpretable Fuzzy System for Automated Diagnosis of Dementia. *Front Neurosci*. 2022;16:867664. doi:10.3389/fnins.2022.867664.
27. Tanveer M, Sajid M, Akhtar M, et al. Fuzzy Deep Learning for the Diagnosis of Alzheimer's Disease: Approaches and Challenges. *IEEE Trans Fuzzy Syst*. 2024;32:5477-5492. doi:10.1109/TFUZZ.2024.3409412.
28. Revett K, Gorunescu F, Salem A-BM. Feature Selection in Parkinson's Disease: A Rough Sets Approach. In: *2009 international multiconference on computer science and information technology*; 2009:425-428. doi:10.1109/IMCSIT.2009.5352688.
29. Sarin K, Bardamova M, Svetlakov M, et al. A Three-stage Fuzzy Classifier Method for Parkinson's Disease Diagnosis Using Dynamic Handwriting Analysis. *Decision Analytics Journal*. 2023;8:100274. doi:10.1016/j.dajour.2023.100274.
30. Reijnders JSAM, Ehrt U, Weber WEJ, Aarsland D, Leentjens AFG. A systematic review of prevalence studies of depression in Parkinson's disease. *Mov Disord*. 2008;23:183-189. doi:10.1002/mds.21803.
31. Jellinger KA. The pathobiological basis of depression in Parkinson disease: challenges and outlooks. *J Neural Transm (Vienna)*. 2022; 129:1397-1418. doi:10.1007/s00702-022-02559-5.
32. Darweesh SKL, Verlinden VJA, Stricker BH, Hofman A, Koudstaal PJ, Ikram MA. Trajectories of prediagnostic functioning in Parkinson's disease. *Brain*. 2017;140:429-441. doi:10.1093/brain/aww291.
33. Dlay JK, Duncan GW, Khoo TK, et al. Progression of Neuropsychiatric Symptoms over Time in an Incident Parkinson's Disease Cohort (ICICLE-PD). *Brain Sci*. 2020;10:78. doi:10.3390/brainsci10020078.
34. Rodriguez-Oroz MC, Jahanshahi M, Krack P, et al. Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms. *Lancet Neurol*. 2009;8:1128-1139. doi:10.1016/S1474-4422(09)70293-5.
35. Jaunarajs KLE, Angoa-Perez M, Kuhn DM, Bishop C. Potential mechanisms underlying anxiety and depression in Parkinson's disease: Consequences of l-DOPA treatment. *Neurosci Biobehav Rev*. 2011;35:556-564. doi:10.1016/j.neubiorev.2010.06.007.
36. Brockman S, Jayawardena B, Starkstein S. The diagnosis of depression in Alzheimer's disease: review of the current literature. *Neuropsychiatry (London)*. 2011;1:377-384. doi:10.2217/npy.11.39.
37. Mourao RJ, Mansur G, Malloy-Diniz LF, Castro Costa E, Diniz BS. Depressive symptoms increase the risk of progression to dementia in subjects with mild cognitive impairment: systematic review and meta-analysis. *Int J Geriatr Psychiatry*. 2016;31:905-911. doi:10.1002/gps.4406.
38. Engedal K, Barca ML, Laks J, Selbaek G. Depression in Alzheimer's disease: specificity of depressive symptoms using three different clinical criteria. *Int J Geriatr Psychiatry*. 2011;26:944-951. doi:10.1002/gps.2631.
39. Lyketsos CG, Breitner JCS, Rabins PV. An evidence-based proposal for the classification of neuropsychiatric disturbance in Alzheimer's disease. *Int J Geriatr Psychiatry*. 2001;16:1037-1042. doi:10.1002/gps.440.
40. Migliorelli R, Teson A, Sabe L, Petracchi M, Leiguarda R, Starkstein SE. Prevalence and correlates of dysthymia and major depression among patients with Alzheimer's disease. *Am J Psychiatry*. 1995;152:37-44. doi:10.1176/ajp.152.1.37.
41. Starkstein SE, Chemerinski E, Sabe L, et al. Prospective longitudinal study of depression and anosognosia in Alzheimer's disease. *Br J Psychiatry*. 1997;171:47-52. doi:10.1192/bjp.171.1.47.
42. Barnes DE, Alexopoulos GS, Lopez OL, Williamson JD, Yaffe K. Depressive Symptoms, Vascular Disease, and Mild Cognitive Impairment. *Arch Gen Psychiatry*. 2006;63:273. doi:10.1001/archpsyc.63.3.273.
43. Zhao Q-F, Tan L, Wang H-F, et al. The prevalence of neuropsychiatric symptoms in Alzheimer's disease: Systematic review and meta-analysis. *J Affect Disord*. 2016;190:264-271. doi:10.1016/j.jad.2015.09.069.
44. Teng E, Lu PH, Cummings JL. Neuropsychiatric Symptoms Are Associated with Progression from Mild Cognitive Impairment to Alzheimer's Disease. *Dement Geriatr Cogn Disord*. 2007;24:253-259. doi:10.1159/000107100.
45. Ruthirakuhan M, Herrmann N, Vieira D, Gallagher D, Lanctôt KL. The Roles of Apathy and Depression in Predicting Alzheimer Disease: A Longitudinal Analysis in Older Adults With Mild Cognitive Impairment. *Am J Geriatr Psychiatry*. 2019;27:873-882. doi:10.1016/j.jagp.2019.02.003.
46. Johansson M, Stomrud E, Lindberg O, et al. Apathy and anxiety are early markers of Alzheimer's disease. *Neurobiol Aging*. 2020;85: 74-82. doi:10.1016/j.neurobiolaging.2019.10.008.
47. Johansson M, Stomrud E, Johansson PM, et al. Development of Apathy, Anxiety, and Depression in Cognitively Unimpaired Older Adults: Effects of Alzheimer's Disease Pathology and Cognitive Decline. *Biol Psychiatry*. 2022;92:34-43. doi:10.1016/j.biopsych.2022.01.012.
48. Belleville S, Fouquet C, Duchesne S, Collins DL, Hudon C. Detecting Early Preclinical Alzheimer's Disease via Cognition, Neuropsychiatry, and Neuroimaging: Qualitative Review and Recommendations for Testing. *J Alzheimers Dis*. 2014;42:S375-S382. doi:10.3233/JAD-141470.
49. Przybyszewski AW, Nowacki JP, Drabik A, et al. Granular Computing (GC) Demonstrates Interactions Between Depression and Symptoms Development in Parkinson's Disease Patients. In: *Intelligent Information and Database Systems*. Springer; 2019:591-601. doi:10.1007/978-3-030-14802-7_51.

50. Shah J, Krell-Roesch J, Forzani E, et al. Predicting cognitive decline from neuropsychiatric symptoms and Alzheimer's disease biomarkers: A machine learning approach to a population-based data. *J Alzheimers Dis*. 2025;103:833-843. doi:10.1177/13872877241306654.
51. Patel SB, Bredemeyer OB, FitzGerald JJ, et al. *Hierarchical Machine Learning Classification of Parkinsonian Disorders using Saccadic Eye Movements: A Development and Validation Study*; 2024. 10.48550/arXiv.2407.16063.
52. Uribarri G, von Huth SE, Waldthaler J, et al. *Deep Learning for Time Series Classification of Parkinson's Disease Eye Tracking Data*; 2023. doi:10.48550/arXiv.2311.16381
53. Alcala-Durand MF, Puerta-Acevedo JC, Arias-Londoño JD, et al. *Automatic Screening of Parkinson's Disease from Visual Explorations*; 2025. 10.48550/arXiv.2509.01326.
54. Brien DC, Riek HC, Yep R, et al. Classification and staging of Parkinson's disease using video-based eye tracking. *Parkinsonism Relat Disord*. 2023;110:105316. doi:10.1016/j.parkreldis.2023.105316.
55. Niño J, Guayacán L, Gómez S, et al. *A digital eye-fixation biomarker using a deep anomaly scheme to classify Parkisonian patterns*; 2025. 10.48550/arXiv.2502.17762.
56. Bredemeyer O, Patel S, FitzGerald JJ, Antoniades CA. Oculomotor deficits in Parkinson's disease: Increasing sensitivity using multivariate approaches. *Front Digit Health*. 2022;4:939677. doi:10.3389/fdgth.2022.939677.
57. Buonocore J, Facchin A, Crasà M, et al. Eye movement abnormalities in Parkinson's disease motor subtypes: a video-oculographic study. *Neurol Sci*. 2025;46:3677-3684. doi:10.1007/s10072-025-08184-w.
58. Wang M, Li Y, Yao J, et al. Memory-guided saccade latency as a potential biomarker for axial symptoms and freezing of gait in Parkinson's disease. *Neurobiol Dis*. 2026;218:107238. doi:10.1016/j.nbd.2025.107238.
59. Koch NA, Voss P, Cisneros-Franco JM, et al. Eye movement function captured via an electronic tablet informs on cognition and disease severity in Parkinson's disease. *Sci Rep*. 2024;14:9082. doi:10.1038/s41598-024-59750-9.
60. Balaji E, Brindha D, Elumalai VK, et al. Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network. *Appl Soft Comput*. 2021;108:107463. doi:10.1016/j.asoc.2021.107463.
61. Yoon S, Kim M, Lee WW. Long Short-Term Memory-Based Deep Learning Models for Screening Parkinson's Disease Using Sequential Diagnostic Codes. *J Clin Neurol*. 2023;19(3):270-279. doi:10.3988/jcn.2022.0160.
62. Nguyen M, He T, An L, Alexander DC, Feng J, Yeo BT. Predicting Alzheimer's disease progression using deep recurrent neural networks. *Neuroimage*. 2020;222:117203. doi:10.1016/j.neuroimage.2020.117203.
63. Zhao A, Qi L, Li J, Dong J, Yu H. A hybrid spatio-temporal model for detection and severity rating of Parkinson's disease from gait data. *Neurocomputing (Amst)*. 2018;315:1-8. doi:10.1016/j.neucom.2018.03.032.
64. Farhan S, Haq YU, Khaliq MA, et al. Enhancing the early detection of Alzheimer's disease using an integrated CNN-LSTM framework: A robust approach for fMRI-based multi-stage classification. *PLoS One*. 2025;20(8):e0317968. doi:10.1371/journal.pone.0317968.
65. Sunnetci KM, Balci M, Ekersular MN, Oguz FE, Alkan A. Alzheimer's and Parkinson's Detection with Video-Based Hybrid Deep Learning from Brain MRI. *J Imaging Inform Med*. 2026. doi:10.1007/s10278-026-01951-x.
66. Ebrahimi A, Luo S, Chiong R. Deep sequence modelling for Alzheimer's disease detection using MRI. *Comput Biol Med*. 2021;134:104537. doi:10.1016/j.compbiomed.2021.104537.
67. Salehi W, Baglat P, Gupta G, et al. An Approach to Binary Classification of Alzheimer's Disease Using LSTM. *Bioengineering (Basel)*. 2023;10:950. doi:10.3390/bioengineering10080950.
68. Pedrero-Sánchez JF, Belda-Lois J-M, Serra-Añó P, Inglés M, López-Pascual J. Classification of healthy, Alzheimer and Parkinson populations with a multi-branch neural network. *Biomed Signal Process Control*. 2022;75:103617. doi:10.1016/j.bspc.2022.103617.
69. Bringas S, Salomón S, Duque R, Lage C, Montaña JL. Alzheimer's Disease stage identification using deep learning models. *J Biomed Inform*. 2020;109:103514. doi:10.1016/j.jbi.2020.103514.
70. Mariathangam PR, Alaudin VBS, Raja JE. Identification and Analysis of Neurodegenerative Diseases with Twin Layered CNN Using Gait Dynamics. *International Journal of Intelligent Engineering and Systems*. 2022;15:68-78. doi:10.22266/ijies2022.0430.07.
71. Yuan N, Guan D, Li S, Zhang L, Zhu Q. Enhancing Neurodegenerative Disease Diagnosis Through Confidence-Driven Dynamic Spatio-Temporal Convolutional Network. *IEEE Trans Neural Syst Rehabil Eng*. 2025;33:1715-1728. doi:10.1109/TNSRE.2025.3564983.
72. Vidya B, Sasikumar P. Parkinson's disease diagnosis and stage prediction based on gait signal analysis using EMD and CNN–LSTM network. *Eng Appl Artif Intell*. 2022;114:105099. doi:10.1016/j.engappai.2022.105099.
73. Huang G, Li R, Bai Q, Alty J. Multimodal learning of clinically accessible tests to aid diagnosis of neurodegenerative disorders: a scoping review. *Health Inf Sci Syst*. 2023;11:32. doi:10.1007/s13755-023-00231-0.
74. Odusami M, Maskeliunas R, Damasevicius R, Misra S. Explainable Deep-Learning-Based Diagnosis of Alzheimer's Disease Using Multimodal Input Fusion of PET and MRI Images. *J Med Biol Eng*. 2023;43:291-302. doi:10.1007/s40846-023-00801-3.
75. Venugopalan J, Tong L, Hassanzadeh HR, Wang MD. Multimodal deep learning models for early detection of Alzheimer's disease stage. *Sci Rep*. 2021;11:3254. doi:10.1038/s41598-020-74399-w.

76. Leony F, Lin C, Alzheimer's Disease Neuroimaging Initiative. Multimodal fusion architectures for Alzheimer's disease diagnosis: An experimental study. *J Biomed Inform*. 2025;166:104834. doi:10.1016/j.jbi.2025.104834.
77. Dentamaro V, Impedovo D, Musti L, Pirlo G, Taurisano P. Enhancing early Parkinson's disease detection through multimodal deep learning and explainable AI: insights from the PPMI database. *Sci Rep*. 2024;14:20941. doi:10.1038/s41598-024-70165-4.
78. Jha D, Shreshtha AD, Sarker U, Thakur LN, Biswas A. Early Prediction of Neurodegenerative Disorders Using Multimodal Deep Learning. *Int J Innov Sci Res Technol*. 2026;11:2455-2464. doi:10.38124/ijisrt/26feb1296.
79. Singh A, Shi W, Wang MD. Multi-Modal Deep Feature Integration for Alzheimer's Disease Staging. In: *2023 IEEE international conference on bioinformatics and biomedicine (BIBM)*; 2023:1-6. 10.1109/BIBM58861.2023.10431906.
80. Ahmed SR, Hadi A, Baddour LS, et al. Alzheimer's Disease Prediction Using MRI-PET Fusion Through Multi-Modal Deep Learning With Late Fusion Techniques. In: *2025 3rd international conference on cyber resilience (ICCR)*; 2025:1-8. doi:10.1109/ICCR67387.2025.11292326.
81. Zhang Y, He X, Chan YH, Teng Q, Rajapakse JC. Multi-modal graph neural network for early diagnosis of Alzheimer's disease from sMRI and PET scans. *Comput Biol Med*. 2023;164:107328. doi:10.1016/j.compbiomed.2023.107328.
82. Kaliappan S, Kamal MDR, Balaji V, et al. Advanced Neural Network Models for Predictive Analytics and Healthcare Management in Neurodegenerative Diseases. In: *2024 international conference on advancements in smart, secure and intelligent computing (ASSIC)*; 2024:1-6. doi:10.1109/ASSIC60049.2024.10508005.
83. Vijay Anand R, Shanmuga Priyan T, Guru BM, Balusamy B, Benedetto F. IMNMAGN: Integrative Multimodal Approach for Enhanced Detection of Neurodegenerative Diseases Using Fusion of Multidomain Analysis With Graph Networks. *IEEE Access*. 2024;12:73095-73112. doi:10.1109/ACCESS.2024.3403860.
84. Jiang Y, Yang H-J, Kim J, Tang Z, Ruan X. Power of Multi-Modality Variables in Predicting Parkinson's Disease Progression. *IEEE J Biomed Health Inform*. 2025;29:1343-1356. doi:10.1109/JBHI.2024.3482180.
85. Hall A, Muñoz-Ruiz M, Mattila J, et al. Generalizability of the Disease State Index Prediction Model for Identifying Patients Progressing from Mild Cognitive Impairment to Alzheimer's Disease. *J Alzheimers Dis*. 2015;44:79-92. doi:10.3233/JAD-140942.
86. Birkenbihl C, Emon MA, Vrooman H, et al. Differences in cohort study data affect external validation of artificial intelligence models for predictive diagnostics of dementia - lessons for translation into clinical practice. *EPMA J*. 2020;11:367-376. doi:10.1007/s13167-020-00216-z.
87. Bron EE, Klein S, Papma JM, et al. Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease. *Neuroimage Clin*. 2021;31:102712. doi:10.1016/j.nicl.2021.102712.
88. Kim HW, Lee HE, Lee S, Oh KT, Yun M, Yoo SK. Slice-selective learning for Alzheimer's disease classification using a generative adversarial network: a feasibility study of external validation. *Eur J Nucl Med Mol Imaging*. 2020;47:2197-2206. doi:10.1007/s00259-019-04676-y.
89. Thorp JE, Adamczyk PG, Ploeg H-L, Pickett KA. Monitoring Motor Symptoms During Activities of Daily Living in Individuals With Parkinson's Disease. *Front Neurol*. 2018;9:1036. doi:10.3389/fneur.2018.01036.
90. Chudzik A, Przybyszewski AW. Classification of Parkinson's Disease Using Machine Learning with MoCA Response Dynamics. *Applied Sciences*. 2024;14:2979. doi:10.3390/app14072979.
91. Burq M, Rainaldi E, Ho KC, et al. Virtual exam for Parkinson's disease enables frequent and reliable remote measurements of motor function. *NPJ Digit Med*. 2022;5:65. doi:10.1038/s41746-022-00607-8.
92. Lipsmeier F, Taylor KI, Postuma RB, et al. Reliability and validity of the Roche PD Mobile Application for remote monitoring of early Parkinson's disease. *Sci Rep*. 2022;12:12081. doi:10.1038/s41598-022-15874-4.
93. Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate Telemonitoring of Parkinson's Disease Progression by Noninvasive Speech Tests. *IEEE Trans Biomed Eng*. 2010;57:884-893. doi:10.1109/TBME.2009.2036000.
94. Zhan A, Little MA, Harris DA, et al. *High Frequency Remote Monitoring of Parkinson's Disease via Smartphone: Platform Overview and Medication Response Detection*; 2016. 10.48550/arXiv.1601.00960.
95. Wu P, Cao B, Liang Z, Wu M. The advantages of artificial intelligence-based gait assessment in detecting, predicting, and managing Parkinson's disease. *Front Aging Neurosci*. 2023;15:1191378. doi:10.3389/fnagi.2023.1191378.
96. Dzialas V, Doering E, Eich H, et al. Houston, We Have AI Problem! Quality Issues with Neuroimaging-Based Artificial Intelligence in Parkinson's Disease: A Systematic Review. *Mov Disord*. 2024;39:2130-2143. doi:10.1002/mds.30002.

Appendix

Abbreviations

AD	Alzheimer's disease
ADL	activity of daily living
ADNI	Alzheimer's Disease Neuroimaging Initiative

AI	artificial intelligence
AIBL	Australian Imaging, Biomarkers and Lifestyle study
ALS	amyotrophic lateral sclerosis
ANFIS	adaptive neuro-fuzzy inference system
AUC	area under the receiver operating characteristic curve
AUROC	area under the receiver operating characteristic curve
BEGAN	boundary equilibrium generative adversarial network
CDRSUM	Clinical Dementia Rating sum of boxes
CI	confidence interval
CN	cognitively normal
CNN	convolutional neural network
CSF	cerebrospinal fluid
CT	computed tomography
DAT	dopamine transporter
DAT-SPECT	dopamine transporter single-photon emission computed tomography
DBS	deep brain stimulation
DESCRIPA	Development of Screening Guidelines and Criteria for Predementia Alzheimer's Disease
DTI	diffusion tensor imaging
ECG	electrocardiography
EEG	electroencephalography
EHR	electronic health record
EMD	empirical mode decomposition
EMCI	early mild cognitive impairment
EN	Excitation Network
FDG-PET	fluorodeoxyglucose positron emission tomography
F1	F1-score
fMRI	functional magnetic resonance imaging
FOG	freezing of gait
GNN	graph neural network
Grad-CAM	gradient-weighted class activation mapping
GRU	gated recurrent unit
HC	healthy control
ICA	independent component analysis
ICD	International Classification of Diseases
ICICLE-PD	Incidence of Cognitive Impairment in Cohorts with Longitudinal Evaluation-Parkinson's Disease
ICC	intraclass correlation coefficient
ICU	intensive care unit
IMU	inertial measurement unit
KNN	k-nearest neighbors
LEM2	Learning from Examples Module, version 2
LIME	Local Interpretable Model-agnostic Explanations
LMCI	late mild cognitive impairment
LSTM	long short-term memory
MCI	mild cognitive impairment
MDS-UPDRS	Movement Disorder Society-Unified Parkinson's Disease Rating Scale
MEG	magnetoencephalography
ML	machine learning
MLP	multilayer perceptron
MoCA	Montreal Cognitive Assessment
MRI	magnetic resonance imaging
NACC	National Alzheimer's Coordinating Center
NC	normal control
ND	neurodegenerative disease
NPS	neuropsychiatric symptoms

OASIS	Open Access Series of Imaging Studies
PD	Parkinson's disease
PET	positron emission tomography
PIGD	postural instability and gait difficulty
PLS	partial least squares
pMCI	progressive mild cognitive impairment
PPMI	Parkinson's Progression Markers Initiative
PSP	progressive supranuclear palsy
QLPO	quicksort leave-pair-out
REM	rapid eye movement
RNN	recurrent neural network
ROCKET	random convolutional kernel transform
SANRA	Scale for the Assessment of Narrative Review Articles
SCD	subjective cognitive decline
SHAP	SHapley Additive exPlanations
SMC	significant memory concern
SNF	similarity network fusion
sMCI	stable mild cognitive impairment
SNP	single-nucleotide polymorphism
sMRI	structural magnetic resonance imaging
SVM	support vector machine
TADPOLE	Alzheimer's Disease Prediction of Longitudinal Evolution challenge
TCN	temporal convolutional network
TD	tremor-dominant
UPDRS	Unified Parkinson's Disease Rating Scale
VGRF	vertical ground reaction force
ViT	vision transformer
VOG	video-oculography