

Artificial Intelligence and Human Expertise in Neurodegenerative Care: A Synergistic Approach to Diagnosis and Clinical Decision-Making

Aishiq Hussain Ganie^{1*}, Bobby Chouhan², Archita Samanta², Neha², Tanmay Paul², Dipika Khatri³, Jofrin Zakirhusen Patel⁴

¹ Assistant Professor, Department of Pharmacology, Nargund College of Pharmacy, Bengaluru, Karnataka – 560085, India.

Email: aishaqganie72@gmail.com

² Bachelor of Pharmacy Student, Nargund College of Pharmacy, Bengaluru, Karnataka – 560085, India.

³ Master of Pharmacy Student, Department of Pharmaceutics, Shree Dhanvantary Pharmacy College, Kim, Gujarat – 394110, India.

⁴ M.Pharm Student, Department of Chemistry, Shree Dhanvantary Pharmacy College, Kim, Gujarat, India – 394110.

Corresponding Author:

Aishiq Hussain Ganie

Assistant Professor, Department of Pharmacology, Nargund College of Pharmacy, Bengaluru, Karnataka – 560085, India.

Email: aishaqganie72@gmail.com

Received: 2nd Dec, 2025; Revised: 20th Dec 2025; Accepted: 14th Mar, 2026; Available Online: 26th August, 2026

ABSTRACT

Background. Neurodegenerative diseases are placing an increasing burden on healthcare systems worldwide, while artificial intelligence is becoming integrated into an expanding range of clinical and research applications, from neuroimaging interpretation to biomarker analysis and drug discovery. **Objective.** This narrative review critically evaluates the applications of artificial intelligence across Parkinson disease, Alzheimer disease, amyotrophic lateral sclerosis, Huntington disease, and multiple sclerosis, with particular emphasis on the areas in which human clinical expertise remains essential. **Main findings.** Deep learning and other artificial intelligence models applied to neuroimaging, gait, speech, and biomarker data frequently demonstrate high accuracy in controlled research cohorts. However, their clinical translation remains constrained by limited external validation, inadequate demographic representation, and persistent concerns regarding model interpretability across the neurodegenerative diseases examined. Evidence of algorithmic bias in healthcare more broadly further highlights the need for systematic evaluation of fairness within this field. **Clinical significance.** Human expertise remains central to aspects of neurodegenerative care that cannot be reliably automated, including longitudinal clinical judgment, ethical decision making, pharmacological safety oversight, interpretation of patient-specific circumstances, and meaningful communication with patients and families. **Future perspectives.** Sustainable progress will require more representative and secure data infrastructures, genuinely interpretable artificial intelligence models, and robust governance frameworks that position clinicians as active decision makers and overseers rather than passive recipients of algorithmic recommendations

How to cite this article: Ganie AH, Chouhan B, Samanta A, Neha, Paul T, Khatri D, Patel JZ..., Artificial Intelligence and Human Expertise in Neurodegenerative Care: A Synergistic Approach to Diagnosis and Clinical Decision-Making. Int J Drug Deliv Technol. 2026;16(84s): 963-980; DOI: 10.25258/ijddt.16.84s.104

Source of support: Nil.

Conflict of interest: Nil.

Introduction

1.1 Global burden of neurodegenerative diseases

Neurodegenerative diseases are emerging as an increasingly important global health challenge, with their burden rising more rapidly than that of many other chronic conditions. Parkinson disease illustrates this trend clearly, with its prevalence increasing from approximately 3.15 million cases in 1990 to 11.77 million in 2021, representing a 274% increase. This rise has been driven largely by population aging rather than by a single newly emerging risk factor, and projections based on the Global Burden of Disease 2021 data indicate that the number of affected individuals is likely to continue increasing through 2050

[1]. A similar pattern is evident for Alzheimer disease and related dementias. More than 46 million people are currently estimated to be living with dementia worldwide, and this number is projected to exceed 131 million by 2050, placing substantial and growing demands on diagnostic services, long-term care, and caregiving infrastructure [2]. Although amyotrophic lateral sclerosis is considerably less prevalent, its clinical burden is particularly severe because survival is typically limited to two to four years following symptom onset. Consequently, timely recognition and diagnosis are especially important in this disease, where even relatively short delays can have meaningful consequences for patient management and care planning [3].

Importantly, this growing burden is not distributed evenly across populations or healthcare settings. Beyond the increasing number of affected individuals, delays in diagnosis represent an additional and often underappreciated challenge. A study that reconstructed the amyotrophic lateral sclerosis diagnostic pathway using real-world electronic health records reported a median interval of 11 months between symptom onset and formal diagnosis, with fewer than 40% of patients having consulted a neurologist before receiving the diagnosis [4]. Such delays extend beyond a diagnostic inconvenience; they can reduce the time available for timely multidisciplinary care and supportive planning. In amyotrophic lateral sclerosis, earlier recognition may facilitate interventions such as nutritional support and respiratory planning, as well as more informed discussions regarding disease progression and future care needs. Together, the increasing prevalence of neurodegenerative diseases and persistent delays in diagnosis highlight the need for approaches that can support earlier recognition and more efficient clinical assessment without compromising the role of expert clinical judgment.

1.2 Challenges in current clinical practice

Several structural challenges recur across almost all neurodegenerative diseases and continue to limit timely, individualized care. Diagnostic delay remains a major concern, as illustrated by the amyotrophic lateral sclerosis data discussed above [4]. This delay is particularly consequential in diseases with a relatively short survival period, where each month spent in diagnostic uncertainty represents lost time for clinical planning, supportive interventions, and appropriate patient and family counseling [3]. At the same time, disease-modifying therapies remain limited or unavailable for many neurodegenerative conditions. Consequently, the principal benefit of earlier detection is often not immediate disease reversal, but the opportunity to initiate symptomatic management, facilitate timely referral and multidisciplinary support, consider clinical trial participation, and allow patients and families to plan for the anticipated course of disease.

Another major challenge is the substantial clinical heterogeneity observed within individual neurodegenerative diseases, which can make disease progression difficult to predict at the level of an individual patient. Huntington disease provides a clear example of this variability. Despite the presence of a confirmed CAG repeat expansion, unsupervised clustering of longitudinal data from the Enroll-HD cohort identified three distinct progression trajectories characterized as rapid, moderate, and slow. Notably, disease trajectory was more strongly associated with a composite measure incorporating age and repeat length than with repeat length alone [5]. This finding illustrates that even when a defining genetic abnormality is known, it may not be sufficient to accurately determine how the disease will evolve in a particular individual. As a result, reliable tools for precision medicine remain limited, and treatment decisions in most neurodegenerative diseases continue to rely predominantly on population-level evidence rather than on individualized biological and

clinical signatures. This gap between population-based evidence and patient-specific decision making represents an important area in which advanced analytical approaches, including artificial intelligence, may provide additional value while still requiring clinical interpretation.

1.3 Emergence of artificial intelligence in medicine

Artificial intelligence has progressed from a largely research-oriented concept to an increasingly active component of neurodegenerative disease research and care. Its growing relevance is closely linked to the nature and volume of data generated by these disorders. Neuroimaging, gait recordings, speech samples, and multi-omics datasets contain high-dimensional information that can be difficult to analyze comprehensively through conventional clinical assessment alone. Artificial intelligence offers the ability to process these complex datasets, identify patterns that may be difficult to recognize manually, and support the extraction of clinically relevant information from large volumes of data. However, its value is best understood as an extension of clinical assessment rather than a replacement for it.

Multiple sclerosis provides a practical example of where such computational support can be particularly useful. MRI-based assessment of lesion burden is central to both diagnosis and longitudinal monitoring, but manually delineating lesions across serial scans can be time-consuming and may introduce inter-rater variability. Recent work examining artificial intelligence applications in multiple sclerosis MRI analysis has described the increasing use of automated segmentation tools to identify and quantify lesions, thereby reducing the manual workload and improving the consistency of measurements in some studies [6]. Such applications illustrate an important early role for artificial intelligence in neurodegenerative and related neurological care: converting complex, repetitive data-analysis tasks into more standardized measurements that can subsequently be interpreted within the broader clinical context by healthcare professionals.

1.4 Why human expertise remains essential

Despite the rapid development of artificial intelligence, no current system has replaced the clinician in neurodegenerative care, and several well-recognized limitations explain why this remains the case. One of the most important barriers is interpretability. Although artificial intelligence models can identify complex patterns and generate highly accurate predictions in specific settings, translating those predictions into a transparent and clinically meaningful explanation of how a particular conclusion was reached remains challenging. A widely cited review of explainable artificial intelligence has highlighted the substantial methodological challenges that remain before predictions generated by complex models, particularly deep learning systems, can be consistently translated into explanations that clinicians can understand and appropriately evaluate [7]. This issue becomes even more important when artificial intelligence is applied to high-stakes clinical decisions. Rudin has argued that, rather than relying primarily on post hoc methods to explain inherently opaque models, high-stakes applications should

favor models that are interpretable by design. Post hoc explanations may be unstable and may not necessarily reflect the actual processes through which a model arrived at its prediction [8]. In neurodegenerative disease, where a diagnostic or prognostic prediction can influence treatment choices, long-term care planning, and major decisions affecting patients and their families, reliance on opaque reasoning is particularly difficult to justify.

Fairness represents a second, closely related concern. The performance of an artificial intelligence system is not determined solely by its algorithmic architecture; it can also reflect biases present in the data, variables, and healthcare systems from which the model learns. Evidence from healthcare outside neurology illustrates the potential consequences. Obermeyer and colleagues demonstrated that a widely used commercial healthcare risk algorithm systematically assigned lower levels of predicted health need to Black patients than to similarly ill White patients because healthcare expenditure was used as a proxy for medical need. However, differences in healthcare costs reflected unequal access to care rather than differences in underlying disease burden [9]. This example demonstrates how an apparently objective algorithm can reproduce or amplify existing inequities when the variables used for prediction do not adequately represent the clinical outcome of interest. Comparable systematic bias audits remain limited in neurodegenerative disease research. Therefore, the absence of documented large-scale bias in this field should not be interpreted as evidence that such bias is absent; rather, it represents an important evidence gap that requires systematic investigation before these technologies can be confidently integrated into routine clinical practice.

1.5 Aim and scope of the review

This narrative review evaluates the current applications of artificial intelligence across five major neurodegenerative diseases: Alzheimer disease, Parkinson disease, amyotrophic lateral sclerosis, Huntington disease, and multiple sclerosis. Rather than providing a descriptive summary of published studies, the review critically examines the strength, consistency, and clinical relevance of the available evidence, with particular attention to the limitations that may influence its translation into routine practice. It further identifies the major gaps between promising model performance in controlled research settings and meaningful clinical deployment, including challenges related to external validation, interpretability, bias, data privacy, and integration into healthcare workflows. Finally, the review considers how artificial intelligence can be incorporated into a collaborative model of care in which computational capabilities such as pattern recognition and data integration complement, rather than replace, human clinical expertise, judgment, ethical reasoning, and patient-centered decision making.

2. Artificial Intelligence in Healthcare

2.1 Evolution of artificial intelligence

The development of artificial intelligence in medicine has progressed through several distinct stages, with each generation of technology expanding the range and

complexity of tasks that can be supported. Early medical expert systems were primarily based on hand-coded clinical rules and operated within narrowly defined conditions anticipated by their designers. The emergence of machine learning shifted this approach toward statistical pattern recognition, enabling algorithms to learn relationships from labeled datasets using techniques such as support vector machines, random forests, and gradient boosting. Deep learning subsequently extended these capabilities by allowing models to automatically learn hierarchical representations directly from raw images, signals, and other complex data, thereby reducing the dependence on manually engineered features required by earlier machine learning approaches.

More recently, large language models have introduced another important transition by enabling artificial intelligence to process and generate information from unstructured clinical text and conversational inputs. This development has expanded the potential applications of artificial intelligence beyond structured prediction and image analysis toward tasks involving clinical documentation, information extraction, and interaction with healthcare professionals and patients. However, increasing flexibility also introduces new risks. A recent survey of large language models in medicine highlights their potential usefulness for clinical documentation and other healthcare tasks while emphasizing the persistent risk of generating plausible but factually incorrect information with unwarranted confidence [10]. In a clinical environment, particularly in diagnostic decision making, this limitation is important because fluent or convincing output does not necessarily indicate clinical accuracy. Thus, the evolution of artificial intelligence has increased its potential contribution to healthcare while simultaneously making appropriate human oversight more important.

2.2 Major artificial intelligence technologies used in healthcare

Machine learning methods remain widely used for clinical prediction tasks involving structured and sensor-derived data. These approaches are particularly useful when clinically meaningful features can be extracted from relatively well-defined datasets. A representative example is provided by Parkinson disease gait research, in which inertial sensor data collected while patients walked along a defined corridor were analyzed using classical machine learning models. These models were able to distinguish patients with Parkinson disease from healthy controls based on gait characteristics, including variations in stride patterns [11]. Such applications demonstrate how machine learning can convert quantitative movement parameters into objective measures that may complement conventional clinical assessment.

Deep learning, particularly convolutional neural networks, has become a major approach for image-based analysis in neurodegenerative disease research. Unlike conventional machine learning methods that generally depend on predefined features, deep learning models can learn complex hierarchical representations directly from imaging data. Systematic review evidence examining functional

MRI applications in Alzheimer disease demonstrates the growing use of convolutional architectures for extracting diagnostically relevant information from both structural and functional brain imaging [12]. The ability to analyze high-dimensional neuroimaging data at scale makes deep learning particularly attractive for disease classification and imaging-based assessment, although the clinical usefulness of these models ultimately depends on their reproducibility and performance beyond the datasets on which they were developed.

Computer vision techniques extend automated analysis to gait and movement assessment using video or depth-camera data, potentially eliminating the need for wearable sensor attachment. In a study of early-stage Parkinson disease, a non-contact gait assessment system was used to characterize differences in gait between patients and controls, after which machine learning models were trained to classify disease status and estimate motor severity from the resulting vision-based measurements [13]. This approach illustrates the potential of computer vision to provide objective and relatively unobtrusive assessments of motor function, particularly when repeated measurements are required. However, the usefulness of such systems in routine practice remains dependent on consistent acquisition conditions and reliable performance across different clinical environments.

Natural language processing enables artificial intelligence systems to extract clinically relevant information from unstructured medical records that would otherwise remain difficult to analyze systematically. This capability is particularly valuable when important aspects of disease progression are distributed across multiple clinical encounters. The amyotrophic lateral sclerosis diagnostic-delay study described earlier used natural language processing to analyze electronic health records from multiple hospitals and reconstruct symptom timelines [4]. By organizing information embedded within longitudinal clinical documentation, this approach can identify patterns and time intervals that would be difficult to reconstruct through manual review of large numbers of patient records. Explainable artificial intelligence seeks to address the limited transparency of complex predictive models by providing information about the factors contributing to their outputs. These approaches range from post hoc feature-attribution methods applied after model development to inherently interpretable architectures designed with transparency in mind from the outset. This distinction is particularly relevant in healthcare because an explanation generated after a prediction is not necessarily equivalent to understanding how the model actually arrived at that prediction [7,8]. Greater interpretability may therefore support clinician confidence and accountability, but the reliability and clinical validity of explanatory methods themselves must also be established before they can be treated as sufficient justification for clinical decisions.

Clinical decision support systems represent an important step in translating artificial intelligence predictions into practical healthcare workflows. Rather than generating an isolated prediction, these systems are intended to integrate

algorithmic outputs with clinical information at the point of care, potentially supporting diagnostic assessment, risk stratification, or treatment planning. However, the transition from model development to functional clinical decision support remains limited. A systematic review of artificial intelligence applications in diagnostic and prognostic neuroimaging for dementia evaluated numerous published models and found that very few had progressed beyond retrospective research validation to anything resembling a deployed decision support tool [14]. This gap highlights an important distinction between demonstrating algorithmic performance in a research environment and establishing genuine clinical utility.

Robotics remains one of the less mature artificial intelligence-related technologies within neurodegenerative disease care, although available evidence indicates potential value, particularly in rehabilitation. A pilot randomized controlled trial reported that robot-assisted gait training improved a composite gait index in individuals with mild Parkinson disease compared with treadmill training, while the intervention was considered feasible, safe, and well tolerated throughout the treatment period [15]. These findings suggest that robotic systems may provide a standardized and measurable approach to rehabilitation. Nevertheless, the evidence base remains substantially smaller than that for imaging, gait analysis, and other data-driven applications, and whether improvements achieved during robotic training translate into sustained long-term functional benefits remains insufficiently established.

The broader deep learning literature in Alzheimer disease further illustrates both the promise of these technologies and the challenges associated with their clinical translation. A systematic literature review evaluating deep learning and machine learning approaches for Alzheimer disease prediction synthesized numerous individual studies and found that reported accuracy generally declined when models were evaluated outside their original training cohorts [16]. Similarly, an earlier systematic review of deep learning approaches for detecting Alzheimer disease from neuroimaging reported that most published models had not been evaluated using independent external datasets [17]. A subsequent systematic review focused specifically on early diagnosis reinforced these concerns and additionally noted that relatively few studies reported confidence intervals or other measures of uncertainty alongside their headline performance estimates [18]. The absence of uncertainty measures can make apparently high accuracy difficult to interpret, particularly when models are considered for application to heterogeneous patient populations.

A twelve-year systematic review of artificial intelligence applications in Alzheimer disease brain MRI analysis further showed that increasing architectural complexity over time was not accompanied by a proportional improvement in real-world validation practices [19]. This observation is important because increasingly sophisticated models may improve performance under controlled conditions without necessarily improving their reliability when transferred to routine clinical settings. Evidence from Parkinson disease also supports the potential advantage of

combining complementary data types. A wearable-sensor study evaluating gait and posture in patients with Parkinson disease and cognitive impairment found that incorporating cognitive status alongside gait measures improved classification compared with gait information alone [20]. Collectively, these findings support the growing interest in

multimodal artificial intelligence, in which information from different clinical domains is integrated rather than analyzed independently, while also emphasizing the need for rigorous external validation before such models can be considered clinically dependable.

Table 1. Major Artificial Intelligence Technologies Used in Neurodegenerative Care

AI Technology	Principle	Major Clinical Applications	Advantages	Limitations	References
Machine Learning	Statistical pattern learning from structured features	Gait classification, disease trajectory clustering	Interpretable relative to deep learning, works with modest sample sizes	Requires manual feature engineering, limited with high dimensional raw data	[1,5,11]
Deep Learning	Hierarchical feature learning from raw imaging or signal data	MRI and PET based diagnosis and staging	High accuracy on image tasks, minimal manual preprocessing	Data hungry, poor generalization across scanners and cohorts	[12,16,17,18,19]
Computer Vision	Automated extraction of spatial or motion features from video	Non-contact gait and posture assessment	Objective, continuous, no sensor attachment required	Sensitive to lighting, camera angle and clinical setting	[13,20]
Natural Language Processing	Extraction of structured meaning from clinical text	Diagnostic timeline reconstruction from records	Recovers information otherwise locked in free text	Performance depends heavily on documentation quality	[4]
Explainable AI	Methods that expose or approximate model reasoning	Clinician facing interpretation, trust building	Improves accountability and regulatory acceptance	Explanations can be unstable or misleading if poorly validated	[7,8]
Clinical Decision Support Systems	Integration of predictive output into clinical workflow	Risk stratification, diagnostic flagging	Brings prediction to the point of care	Alert fatigue, workflow disruption, unclear liability	[14]
Robotics	Automated physical assistance and rehabilitation	Gait training, assistive devices for advanced disease	Reduces caregiver burden, standardizes therapy dosage	Cost, limited evidence of long-term functional benefit	[15]

3. Overview of Neurodegenerative Diseases

3.1 Alzheimer disease

Alzheimer disease is the most common cause of dementia and represents the largest area of artificial intelligence research among the neurodegenerative diseases examined in this review. Its underlying pathology is characterized primarily by the accumulation of amyloid- β plaques and tau neurofibrillary tangles, accompanied by progressive synaptic loss that can begin years, and in some cases more than a decade, before clinically apparent cognitive decline. Traditionally, diagnosis has relied on a combination of

clinical history, cognitive assessment, and, increasingly, molecular and functional neuroimaging. The application of artificial intelligence to Alzheimer disease imaging has expanded considerably, particularly in the analysis of positron emission tomography (PET) data. Interest in artificial intelligence-assisted PET analysis increased following regulatory approval of the tau-specific tracer Tauvid, which contributed to greater availability of tau imaging data for computational model development. A recent narrative review synthesizing high-impact literature from this post-approval period documented a corresponding

increase in published artificial intelligence models focused on tau and related PET imaging applications [21].

Despite these advances in diagnosis and disease characterization, therapeutic options remain limited. Current approved approaches are largely directed toward symptomatic management, while a smaller number of disease-modifying therapies aim to slow aspects of amyloid-related pathology rather than reverse the broader neurodegenerative cascade. This therapeutic landscape further increases the importance of accurate disease detection, characterization, and monitoring, while also highlighting the need to distinguish promising artificial intelligence-based research findings from tools that have demonstrated meaningful clinical utility.

3.2 Parkinson disease

Parkinson disease is the second most common neurodegenerative disorder and, as noted earlier, is currently among the fastest-growing neurological conditions in terms of global disease burden [1]. Its characteristic motor manifestations, including resting tremor, rigidity, and bradykinesia, are primarily associated with the progressive loss of dopaminergic neurons in the substantia nigra. Despite advances in neuroimaging and biomarker research, diagnosis remains predominantly clinical and relies largely on neurological examination, as no single laboratory or imaging biomarker is currently accepted as a standalone diagnostic test. This reliance on clinical assessment becomes particularly challenging during the early stages of disease, when symptoms may be subtle or incompletely expressed. Moreover, Parkinson disease extends beyond its classical motor manifestations. Non-motor features, including cognitive decline, sleep disturbances, and autonomic dysfunction, may emerge years before the onset of motor symptoms and therefore represent potentially valuable indicators of early disease. However, these manifestations are considerably more heterogeneous and difficult to standardize, making their use for early detection and prediction more challenging than the assessment of established motor signs.

3.3 Amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis is a fatal neurodegenerative disorder affecting both upper and lower motor neurons and is characterized by relentlessly progressive muscle weakness, ultimately leading to respiratory failure in many patients. Survival is typically limited to approximately two to four years from symptom onset, making timely diagnosis and appropriate supportive management particularly important [3]. The clinical presentation is heterogeneous, with patients commonly exhibiting spinal or bulbar onset phenotypes that may differ in their patterns and rates of progression as well as in prognosis. This clinical variability creates challenges not only for accurate and timely diagnosis but also for patient stratification and the design of clinical trials [4]. Electrophysiological investigations provide important diagnostic information, particularly when clinical features overlap with other neuromuscular disorders. In this context, a recent large-scale study applying artificial intelligence to F-wave responses obtained during routine nerve conduction studies

demonstrated strong discrimination between amyotrophic lateral sclerosis and clinical mimics, including inclusion body myositis and radiculopathy. Notably, the study used training and validation cohorts comprising tens of thousands of patients, representing a scale that remains uncommon in the artificial intelligence literature on amyotrophic lateral sclerosis [22]. These findings illustrate the potential of artificial intelligence to extract diagnostically useful patterns from routinely acquired electrophysiological data and support its possible role in addressing one of the major challenges in amyotrophic lateral sclerosis: distinguishing the disease from clinically overlapping conditions at an earlier stage.

3.4 Huntington disease

Huntington disease is an autosomal dominant neurodegenerative disorder caused by a CAG repeat expansion in the *HTT* (huntingtin) gene and is characterized by progressive motor, cognitive, and psychiatric deterioration. Unlike many other neurodegenerative diseases, the underlying genetic cause can be identified through genetic testing before the appearance of clinical symptoms. This distinctive feature provides an important opportunity to study the premanifest stage of disease and to develop models aimed at predicting individual disease trajectories before substantial clinical impairment becomes evident. As discussed earlier, machine learning approaches applied to longitudinal Enroll-HD data have begun to capture the considerable variability in disease progression by identifying distinct progression subgroups among genetically confirmed carriers [5]. Nevertheless, the presence of a known causative mutation does not eliminate the challenge of individualized prediction. Determining precisely when a particular individual will convert from the premanifest stage to clinically apparent disease, and how rapidly symptoms will subsequently progress, remains difficult. Thus, Huntington disease provides a particularly relevant setting for evaluating whether artificial intelligence can move beyond genetic risk identification toward meaningful, individualized prediction of disease trajectory.

3.5 Multiple sclerosis

Multiple sclerosis is an autoimmune demyelinating disease of the central nervous system and differs from the other conditions discussed in this review in several important clinical respects. In many patients, the disease follows a relapsing–remitting course and typically begins during early adulthood rather than later life. Diagnosis relies heavily on magnetic resonance imaging (MRI), particularly the demonstration of lesions disseminated in space and time, with cerebrospinal fluid analysis obtained through lumbar puncture often providing additional diagnostic support. Because lumbar puncture is an invasive procedure, there is considerable interest in approaches that could extract more diagnostically useful information from non-invasive imaging data and potentially reduce the need for invasive investigations in appropriate patients. A systematic review examining textural analysis and artificial intelligence as decision-support tools in multiple sclerosis diagnosis found consistent evidence that automated textural and radiomic features can differentiate multiple sclerosis

lesions from other white matter abnormalities [23]. However, the review also highlighted an important limitation: most of the included studies were based on small, single-center cohorts, which limits the confidence with which their findings can be generalized to broader and more diverse patient populations [23]. Thus, although artificial intelligence-based MRI analysis shows promise for improving lesion characterization, its clinical value will ultimately depend on validation across larger, multicenter, and more representative datasets.

4. Applications of Artificial Intelligence in Neurodegenerative Care

4.1 Early diagnosis

Neuroimaging remains the most extensively studied application of artificial intelligence in neurodegenerative disease, with Alzheimer disease accounting for a particularly large proportion of the published literature. The systematic reviews discussed above consistently report classification accuracies above 90% when convolutional neural networks are applied to structural MRI or PET data obtained from the Alzheimer's Disease Neuroimaging Initiative and similarly curated research cohorts [12,16,17,18,19]. However, such performance figures should be interpreted cautiously rather than accepted as direct evidence of clinical readiness. These research cohorts are often demographically narrow and are generated using standardized imaging protocols within specialized and well-resourced research settings. Consequently, they may not adequately represent the broader and more heterogeneous populations encountered in routine clinical practice. This distinction becomes particularly important when models are evaluated on independent, real-world clinical datasets, where performance has been shown to decline across a substantial proportion of the studies reviewed [14]. An early and still relevant comparison by Klöppel and colleagues found that a computerized method could achieve diagnostic accuracy comparable to that of experienced radiologists when distinguishing Alzheimer disease from healthy aging using structural MRI, but did not clearly demonstrate superiority over expert clinical assessment [24]. This finding remains instructive because it emphasizes that high algorithmic performance should not automatically be interpreted as evidence that artificial intelligence can outperform trained clinicians.

In Parkinson disease, research into early diagnosis has focused substantially on quantitative gait assessment using wearable inertial sensors, depth cameras, and markerless video-based systems, as discussed in Section 2.2 [1,11,13,20]. These approaches seek to capture subtle alterations in movement that may be difficult to quantify consistently through conventional clinical examination. A further study applied machine learning to gait severity classification by training models to estimate Hoehn and Yahr staging directly from movement-derived data, with the aim of reducing reliance on subjective clinical rating scales that may vary between examiners [25]. Collectively, these findings demonstrate the potential of automated movement analysis to provide more objective and reproducible measures of motor dysfunction. Nevertheless,

the evidence remains methodologically fragmented. Differences in sensor placement, walking protocols, patient populations, and outcome definitions are substantial across studies, making direct comparison of reported accuracy values difficult and often of limited interpretive value, even when individual studies are internally well designed.

Speech-based analysis represents another emerging approach to early detection, particularly in Alzheimer disease, where subtle alterations in spontaneous language and communication may appear before a formal diagnosis. Recent advances in natural language processing and large language models have made it possible to analyze speech transcripts at a level of detail that was previously difficult to achieve using conventional methods. One recent study used large language models to predict dementia status from transcripts of spontaneous speech and reported discriminatory performance that the authors suggested could support low-cost, remote screening, particularly in populations with limited access to specialist cognitive assessment [26]. Such an approach could potentially extend early screening beyond conventional clinical settings and reduce some of the practical barriers associated with specialist evaluation. However, the evidence remains at an early, proof-of-concept stage. The generalizability of speech-based models across different languages, accents, educational backgrounds, and communication styles has not yet been adequately established, and these factors may substantially influence model performance. Therefore, promising results from controlled speech datasets should be considered an indication of potential rather than definitive evidence for routine clinical deployment.

4.2 Biomarker discovery

Blood-based biomarkers have substantially changed the practical landscape of Alzheimer disease research in recent years. A review of blood biomarkers for Alzheimer disease indicates that plasma measures of phosphorylated tau and related biomarkers can approach the diagnostic performance of cerebrospinal fluid analysis in several validation cohorts, offering a considerably less invasive approach to early disease detection than lumbar puncture [27]. The increasing availability of these markers also creates opportunities for artificial intelligence to integrate biochemical signals with clinical, imaging, and other patient-level data, potentially improving disease characterization and risk stratification. Nevertheless, translation into routine clinical practice remains incomplete. Important challenges include standardization of assays across laboratories and manufacturers, establishment of appropriate reference ranges across diverse patient populations, and clear integration of blood-based biomarkers into existing diagnostic criteria. Thus, although the diagnostic potential of these markers is substantial, their real-world variability warrants careful interpretation, particularly because enthusiasm surrounding their clinical utility has at times progressed faster than the evidence supporting standardized implementation [27].

Biomarker research in amyotrophic lateral sclerosis has followed a notably different, cell-based approach. One study developed an artificial intelligence-based prediction

model using motor neurons derived from induced pluripotent stem cells, applying a convolutional neural network to classify cell images according to whether they originated from patients with amyotrophic lateral sclerosis or healthy controls. The model demonstrated strong discrimination in this early proof-of-concept setting [28]. This represents a particularly innovative application of artificial intelligence because it transforms complex cellular morphology into quantifiable patterns that would otherwise require detailed and time-consuming manual assessment. Such an approach may eventually contribute to the identification of disease-associated cellular phenotypes and potentially support biomarker discovery. However, its current clinical applicability remains limited because the workflow requires patient-derived cell culture and specialized laboratory procedures rather than a readily obtainable clinical specimen. Therefore, despite its methodological novelty, this application currently represents a research-level biomarker strategy rather than a practical bedside diagnostic tool.

4.3 Disease prediction and risk stratification

Huntington disease provides a distinctive example of how machine learning can be applied to disease progression rather than diagnosis, because the underlying genetic mutation can be identified before the onset of clinical manifestations, as discussed previously [5]. In this setting, the objective is to distinguish individuals according to their likely disease trajectory and rate of progression. Such stratification could eventually have practical value in clinical research, particularly by facilitating more efficient trial recruitment and identifying participants who are more likely to demonstrate measurable clinical changes within a realistic study period. However, these potential applications remain dependent on prospective validation in cohorts that are independent of the original training datasets before such models can be considered clinically established.

Amyotrophic lateral sclerosis prognosis research has generated a substantial body of machine learning studies, with much of this work based on the Pooled Resource Open-Access Amyotrophic Lateral Sclerosis Clinical Trials dataset, one of the largest publicly available clinical trial resources for the disease. A systematic review of machine learning and brain-computer interface approaches to amyotrophic lateral sclerosis prognosis identified numerous models designed to predict outcomes such as survival and functional decline. Although many reported reasonable predictive performance, very few had undergone external validation using cohorts that were independent of the datasets used for model development [29]. A separate review focusing specifically on machine learning approaches for amyotrophic lateral sclerosis prognosis reached a similar conclusion and further highlighted substantial variation in model performance according to the clinical, demographic, and genetic features included [30]. Importantly, no consistent combination of features has yet emerged as the optimal representation of the prognostic signal. This variability underscores a broader challenge in prognostic modeling: a model may perform well within a particular dataset without necessarily identifying predictors

that remain robust across different patient populations and clinical settings. Against this background, the large validation cohort used in the F-wave-based diagnostic model discussed earlier is particularly notable because validation at this scale remains uncommon within the broader amyotrophic lateral sclerosis artificial intelligence literature [22].

In Alzheimer disease, the potential of artificial intelligence for prognosis has also been explored alongside its more established application in diagnostic classification. A widely cited review of deep learning applications using neuroimaging data provided an early framework for considering prognostic prediction as a distinct objective from diagnostic classification and helped establish the conceptual basis for subsequent research in this area [31]. Much of the later literature continues to follow this general framework, reflecting a gradual shift from simply identifying whether disease is present toward estimating how the disease may progress in an individual patient. This distinction is clinically important because accurate risk stratification could ultimately provide information that supports longitudinal monitoring and more individualized care, provided that such predictions demonstrate sufficient reliability and external validity.

4.4 Personalized medicine

Personalized medicine in neurodegenerative disease remains more aspirational than established in current clinical practice. Although artificial intelligence-based risk stratification models have been developed for several of the diseases discussed in this review, their translation into individualized treatment decisions remains limited. Existing models have predominantly focused on predicting disease progression, survival, or other clinical outcomes rather than prospectively determining which treatment is most likely to benefit a specific patient [29,30]. Consequently, an important gap remains between identifying an individual who is predicted to progress rapidly and determining how that prediction should alter clinical management. Predictive performance alone does not establish clinical utility; the clinical value of a model ultimately depends on whether acting on its predictions leads to better decisions and meaningful improvements in patient outcomes. Among the studies examined in this review, no randomized clinical trial was identified in which an artificial intelligence-derived risk score was prospectively used to guide an individual treatment decision and subsequently demonstrated improved patient outcomes. Thus, although artificial intelligence provides a promising foundation for precision medicine, its transition from prognostic prediction to evidence-based, treatment-guiding personalization remains an important unresolved challenge.

4.5 Drug discovery and drug repurposing

Artificial intelligence has increasingly been applied across several stages of drug discovery relevant to neurodegenerative diseases, including molecular property prediction, virtual screening, and therapeutic target identification. These computational approaches can process large chemical and biological datasets and prioritize candidate compounds before they proceed to more

resource-intensive experimental testing. A recent review focusing specifically on artificial intelligence in natural product drug discovery for neurodegenerative diseases describes the growing use of computational screening to identify potentially neuroprotective compounds from natural product libraries. The authors suggest that such approaches may be particularly valuable for natural products because of their considerable chemical diversity and structural complexity, which can make conventional screening and optimization more challenging [32].

Drug repurposing has attracted particular interest in rare neurodegenerative diseases because it offers the possibility of identifying new therapeutic uses for compounds that have already undergone some degree of safety evaluation. By starting with existing drugs rather than developing entirely new chemical entities, this strategy may reduce the time and resources required during the early stages of drug development, an advantage that is particularly relevant for diseases affecting relatively small patient populations. A mini-review of artificial intelligence applications in drug repurposing for rare diseases describes several case studies and computational platforms illustrating how these approaches can be used to identify potential therapeutic candidates [33]. A broader review of artificial intelligence in drug repurposing similarly highlights the expanding range of computational strategies being applied, including network-based methods and deep learning approaches that analyze molecular structures and drug–target relationships [34].

Despite this promise, the transition from computational prediction to a clinically actionable therapy remains a substantial challenge. A review specifically examining drug repositioning in neurodegenerative diseases emphasizes that most computationally identified candidates still require extensive preclinical and clinical validation before their therapeutic potential can be established. The problem is particularly relevant to rare neurodegenerative diseases, where disease-specific datasets are often limited and may increase the risk of false-positive target or compound predictions [35]. Therefore, an algorithmically prioritized compound should be regarded as a starting point for experimental investigation rather than as evidence of therapeutic efficacy. The distance between identifying a promising candidate through artificial intelligence and establishing its safety, efficacy, and clinical benefit remains considerable. This distinction is important when interpreting claims of artificial intelligence-driven drug discovery, because computational success in identifying or prioritizing a candidate does not by itself demonstrate that the candidate will translate into an effective treatment for patients.

4.6 Remote patient monitoring

Wearable accelerometers, smartphone-based assessments, voice analysis, and passive sensors used in the home are increasingly being explored for monitoring neurodegenerative disease beyond the clinical setting. These technologies can generate repeated or continuous measurements of symptoms and functional changes during patients' everyday activities, potentially providing a more

detailed picture of disease progression than intermittent clinic visits. A systematic review of wearable GPS and accelerometer technologies for monitoring mobility and physical activity in neurodegenerative disorders found consistent evidence that these devices can capture patterns of gait, mobility, and physical activity with a level of temporal detail that periodic clinical assessments cannot readily provide [36]. Whereas a clinic visit represents a limited snapshot of a patient's condition at a particular point in time, wearable monitoring can record changes continuously over days or months. A complementary narrative review examining machine learning approaches for developing remotely monitored central nervous system biomarkers using wearable sensors further described how these continuous data streams can be transformed into composite digital biomarkers, in which multiple underlying signals are integrated into a single measure intended to represent clinically meaningful changes in disease activity [37]. Such approaches could potentially complement conventional clinical assessments by providing longitudinal, real-world measures of patient function rather than relying exclusively on observations obtained during scheduled appointments.

However, the clinical value of remote monitoring depends not only on the quality of the data or the accuracy of the algorithms analyzing them, but also on whether patients are willing and able to use these technologies consistently. An editorial introducing a special collection on digital remote patient monitoring in neurodegenerative diseases highlighted an important gap between interest in monitoring technologies and their actual adoption. Survey data cited in the editorial indicated that although more than 90% of patients with Parkinson disease expressed interest in new monitoring technologies, only 24% reported using consumer wearable devices for disease management, while just 8% used medical-grade devices [38]. This disparity provides an important counterpoint to technology-centered narratives that focus primarily on algorithmic performance. A highly accurate model cannot provide meaningful clinical benefit if the patient does not consistently use the device required to generate the underlying data. Sustained engagement, usability, comfort, accessibility, and adherence may therefore represent equally important determinants of the clinical effectiveness of remote monitoring, particularly when meaningful disease-related changes need to be observed over months or years.

4.7 Clinical decision support systems

Clinical decision support systems that integrate artificial intelligence outputs into neurology workflows remain considerably less mature than the predictive models on which they are based. Although many models demonstrate promising diagnostic or prognostic performance under research conditions, translating these predictions into tools that can be reliably embedded within routine clinical practice represents a separate and substantially more demanding step. As noted earlier, a systematic review of artificial intelligence-based diagnostic and prognostic neuroimaging models in dementia found that the great majority of published tools remained research prototypes

evaluated retrospectively, rather than fully integrated, regulator-approved decision support systems operating in routine clinical settings [14]. This distinction is important because technical performance alone does not establish clinical usefulness; a model must also function reliably within existing workflows, provide information that clinicians can meaningfully interpret, and support decisions without creating additional sources of error or burden. Moreover, evidence that artificial intelligence-based decision support systems actually change clinician behavior or improve patient outcomes remains limited across nearly all disease categories examined in this review. Most published studies have focused on whether a model can

accurately classify disease, predict progression, or identify a clinical pattern, while comparatively fewer have evaluated what happens after the model's output is incorporated into real clinical decision making. Thus, the field continues to face a critical translational gap between predicting an outcome accurately in a research dataset and demonstrating that using that prediction leads to better clinical decisions and, ultimately, better patient outcomes. This gap reinforces the need for prospective, workflow-integrated evaluation of artificial intelligence systems rather than relying solely on retrospective measures of model accuracy.

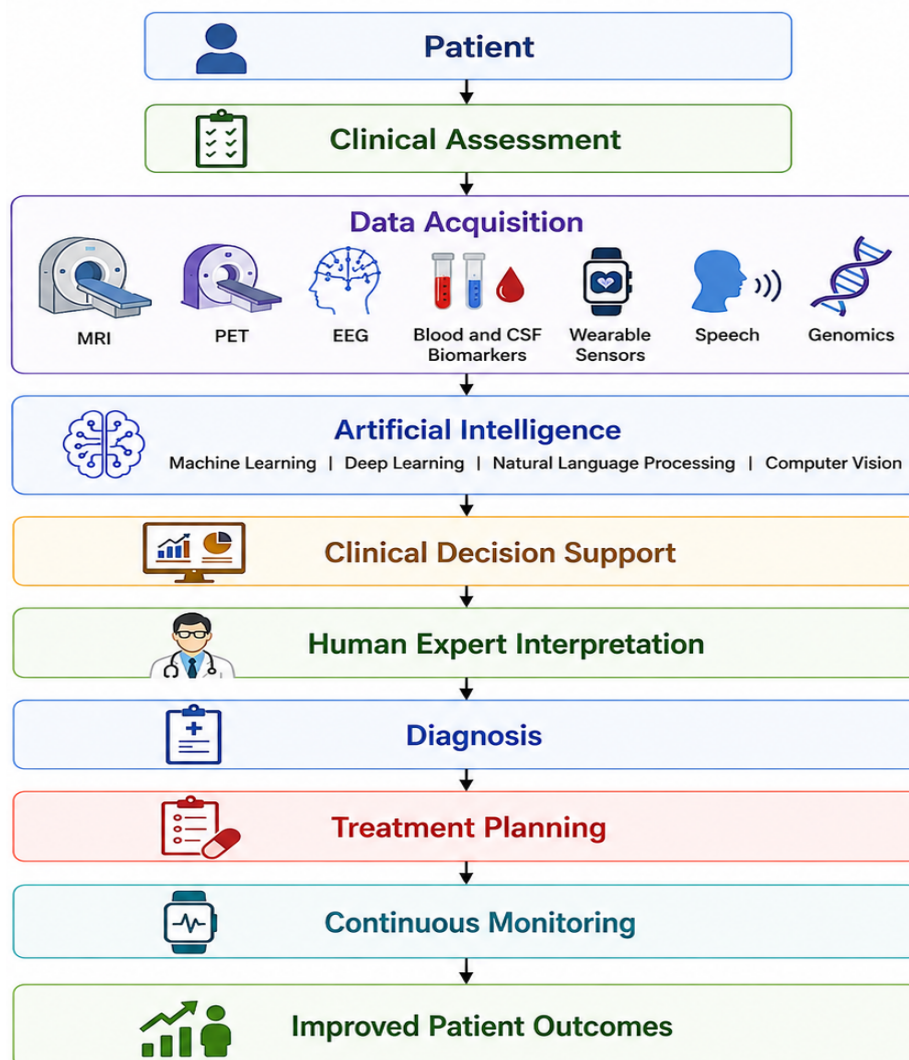


Figure 1. Artificial Intelligence Across the Neurodegenerative Care Continuum

5. Human Expertise in Neurodegenerative Care

5.1 Clinical judgment

Clinical judgment integrates information that no current artificial intelligence model can observe and contextualize in its entirety. A caregiver's seemingly minor observation of a subtle behavioral change, a discrepancy between performance on a formal cognitive assessment and functioning in everyday life, or the accumulated pattern of

symptoms observed across several years may each influence clinical interpretation. These elements are often distributed across multiple consultations rather than captured at a single point of assessment. Consequently, diagnosis in neurodegenerative disease is rarely a single decision based on one test or one algorithmic output. Instead, it is a longitudinal process in which evidence is continuously gathered, interpreted, and reconciled over

time. This broader clinical context remains fundamentally the clinician's responsibility, even when an imaging-, gait-, speech-, or biomarker-based artificial intelligence model contributes one component of the overall assessment.

5.2 Patient-centered care

Patients with neurodegenerative diseases may face years of uncertainty, progressive loss of independence, changing functional abilities, and complex decisions involving family members and caregivers. Within this setting, patient-centered care extends beyond establishing a diagnosis or predicting disease progression. Communicating a diagnosis honestly while preserving appropriate hope, explaining uncertainty without creating unnecessary fear, and adapting discussions to a patient's cognitive capacity and circumstances at a particular stage of disease require empathy, flexibility, and moment-to-moment clinical judgment. These interactions are also shaped by individual values, expectations, family circumstances, and preferences, which may change as the disease progresses. Although artificial intelligence may provide information to support these conversations, it cannot independently assume the relational and contextual responsibilities involved in communicating difficult clinical information and supporting patients and families through its consequences.

5.3 Ethical decision making

Genetic testing in Huntington disease raises particularly important ethical considerations because an individual can learn that they carry the causative mutation years, and sometimes decades, before clinical symptoms become apparent [5]. The decision to undergo testing therefore has implications that extend beyond establishing a diagnosis. Decisions regarding whether to pursue genetic testing, whether and how genetic information should be communicated to family members who may also be at risk, and how future care should be planned around an uncertain but potentially foreseeable disease trajectory require careful consideration of autonomy, personal values, family circumstances, and long-term consequences. Such decisions cannot be reduced to a probability generated by a predictive algorithm and instead require ethical reasoning that remains fundamentally human.

5.4 Interpretation of AI outputs

Even a highly accurate artificial intelligence model produces a probability, classification, or risk estimate rather than a clinical decision. Translating such an output into an actionable clinical plan requires the clinician to consider it alongside the patient's comorbidities, clinical history, preferences, functional status, and broader life circumstances. Explainable artificial intelligence methods may improve transparency by helping clinicians understand which features contributed to a model's prediction, but explanation alone does not eliminate the need for professional interpretation [7,8]. Human expertise therefore remains essential not only for determining whether an AI output is clinically plausible, but also for deciding how much weight it should carry within the context of the complete clinical picture.

5.5 Shared decision making

Shared decision making is particularly important in neurodegenerative diseases in which disease-modifying treatment options remain limited. In conditions such as Alzheimer disease and amyotrophic lateral sclerosis, the value of an artificial intelligence-supported diagnostic or prognostic tool extends beyond its numerical performance. Its practical value also depends on how the information is communicated and incorporated into discussions between the patient, family, and healthcare team. A prediction may help clarify the clinical situation, but decisions concerning treatment priorities, acceptable risks, quality of life, goals of care, and future planning remain deeply individual. Accordingly, artificial intelligence should function as an information-supporting component of shared decision making rather than as a substitute for the dialogue through which these decisions are reached.

5.6 Multidisciplinary team approach

Neurodegenerative care typically involves neurologists, radiologists, pharmacists, rehabilitation specialists, nurses, and family caregivers working together across a prolonged disease course. The contribution of each member may change as the patient's symptoms, functional status, and care requirements evolve. Artificial intelligence tools may intersect with these roles in different ways, for example by supporting image interpretation, monitoring movement, identifying clinically relevant patterns, or assisting with risk assessment. However, the usefulness of these individual outputs depends on how they are interpreted and coordinated within the broader care pathway. Integrating information from different professionals and technologies, resolving conflicting assessments, prioritizing patient-specific needs, and translating multiple inputs into a coherent care plan remain fundamentally collaborative processes. Thus, even as artificial intelligence becomes increasingly integrated into neurodegenerative care, coordinating its outputs with multidisciplinary clinical expertise remains a distinctly human responsibility.

6. Human–AI Synergy in Clinical Practice

Artificial intelligence is most likely to provide meaningful value in neurodegenerative care when it complements, rather than replaces, the expertise of different members of the clinical team. Radiologists, neurologists, rehabilitation specialists, pharmacists, nurses, and caregivers may each interact with AI-derived information in different ways, with the ultimate objective of improving the completeness, consistency, and timeliness of clinical care.

Radiologists can benefit from automated lesion segmentation and quantitative assessment, particularly in multiple sclerosis, where artificial intelligence-based algorithms can assist in measuring lesion burden and changes in lesion volume across serial MRI examinations. Such automated measurements may improve consistency when scans are reviewed over time and reduce some of the variability associated with manual assessment by different radiologists [6,23]. Neurologists may similarly benefit from gait- and speech-based screening tools that can identify subtle abnormalities and potentially flag patients who warrant closer evaluation, complementing conventional

symptom-driven referral and clinical assessment [1,11,13,20,25,26]. In rehabilitation, wearable sensor data can provide objective measures of movement and functional performance that may help specialists tailor therapy intensity and monitor response more systematically. To a more limited extent, robotic gait-training approaches have also shown potential as adjuncts to conventional rehabilitation in Parkinson disease [15,36,37].

Clinical pharmacists may benefit from artificial intelligence-assisted drug safety surveillance, particularly through the analysis of large electronic health record datasets to identify patterns associated with adverse drug events. A systematic review and meta-analysis of machine learning models for predicting adverse drug events from electronic health record data found that many models demonstrated good discrimination during retrospective evaluation; however, relatively few had undergone prospective validation within real-time clinical workflows [39]. This limitation is important because drug safety systems must perform reliably under the conditions of routine prescribing and monitoring rather than solely within retrospective datasets. Ethical oversight represents another layer of clinical practice that cannot be delegated to an algorithm. Char, Shah, and Magnus highlighted ethical challenges associated with the implementation of machine learning in healthcare, including concerns surrounding bias, transparency, and accountability, emphasizing the continuing responsibility of healthcare professionals and institutions in addressing these issues [40].

Patient communication presents a more nuanced example of human-AI interaction. A cross-sectional study comparing responses generated by physicians and an AI

chatbot to patient questions posted on a public online forum found that trained evaluators rated chatbot responses as higher in quality and empathy than physician responses in most of the comparisons examined [41]. However, the findings should not be interpreted as evidence that AI can replace clinicians in patient communication. The study itself cautioned that such systems may be better positioned as tools for drafting or supporting communication that remains subject to clinician review [41]. This distinction is particularly important in neurodegenerative care, where communication often involves uncertainty, progressive disability, treatment limitations, and emotionally complex decisions that require contextual understanding beyond the content of an individual question.

Nurses and family caregivers may benefit most directly from remote monitoring technologies that reduce the need for continuous in-person observation and provide longitudinal information about changes in mobility, activity, or other clinically relevant measures. However, these potential benefits depend heavily on sustained patient engagement and adherence to the monitoring technology. As discussed earlier, low adoption and continued-use barriers can substantially limit the practical value of otherwise capable digital monitoring systems [38]. Therefore, the most effective model of human-AI synergy is not one in which technology progressively removes human involvement, but one in which AI assumes selected tasks that benefit from continuous measurement, large-scale data processing, or pattern recognition while healthcare professionals retain responsibility for interpretation, communication, ethical oversight, and final clinical decisions.

Table 2. Comparative Roles of Artificial Intelligence and Human Expertise in Neurodegenerative Care

Clinical Activity	Role of AI	Role of Human Expert	Synergistic Outcome	References
Early Diagnosis	Pattern detection in imaging, gait and speech data	Integration with history, exam and caregiver report	Earlier, more consistent flagging without replacing clinical confirmation	[12,13,16,17,18,19,24,25,26]
MRI Interpretation	Automated lesion segmentation and volumetric quantification	Contextual interpretation and differential diagnosis	Faster, more reproducible measurement supporting expert review	[6,23]
Biomarker Analysis	High throughput screening of blood, CSF and cellular markers	Clinical correlation and pretest probability judgment	More accessible pathways toward earlier detection	[27,28]
Disease Prediction	Trajectory clustering and progression modeling	Communicating uncertainty to patients and families	Better informed trial design and patient counseling	[5,22,29,30,31]
Treatment Selection	Candidate compound and repurposing suggestions	Pharmacological judgment and safety oversight	Shortened discovery timelines with maintained safety standards	[32,33,34,35]

Drug Monitoring	Detects adverse drug event patterns in longitudinal EHR data	Causal judgment, dose adjustment, prospective oversight	Earlier signal detection combined with sound clinical action	[39]
Patient Counseling	Can draft quality, empathetic responses to routine written questions	Full clinical responsibility, relationship, complex or sensitive communication	AI assisted drafting with mandatory clinician review	[41]
Rehabilitation	Objective movement tracking, robot assisted gait training	Therapy design and motivational support	Individualized, measurable rehabilitation programs	[15,36,37]
Ethical Decision Making	None	Full responsibility for bias, transparency and accountability	Preserves patient autonomy and informed consent	[40]
Long Term Follow up	Continuous remote symptom tracking	Longitudinal relationship and trend interpretation	Earlier detection of decline between scheduled visits	[38]

7. Current Challenges and Limitations

7.1 Algorithm bias

Bias in clinical algorithms is not merely a theoretical concern; it has already been demonstrated in real-world healthcare applications. As discussed in Section 1.4, a widely used commercial risk-prediction algorithm was shown to systematically underestimate the healthcare needs of Black patients because it relied on healthcare expenditure as a proxy for medical need, even though differences in healthcare costs reflected disparities in access to care rather than differences in illness burden [9]. Although an equivalent large-scale systematic audit of algorithmic bias in neurodegenerative disease has not yet been established, there is little basis for assuming that models developed in this field are inherently protected from similar sources of bias. In fact, the composition of the datasets used for model development represents an important potential vulnerability. Many neuroimaging and gait datasets underlying the studies reviewed in this article are derived from research cohorts concentrated in high-income countries and specialized clinical centers. Consequently, patients from underrepresented demographic, geographic, and socioeconomic groups may be insufficiently represented during model development and validation. This limitation is particularly relevant to Alzheimer disease neuroimaging research, where demographic underrepresentation has been repeatedly recognized as a concern across the literature [12,16,17,18,19]. If such models are subsequently deployed in populations that differ substantially from their training cohorts, apparently high performance may not translate into equally reliable predictions across all patient groups. Therefore, demographic diversity should be considered not simply a desirable feature of future datasets, but an important component of the clinical validity and fairness of neurodegenerative disease AI systems.

7.2 Data privacy

Neurodegenerative disease data can be highly sensitive, particularly when it includes genetic information, as in Huntington disease, or detailed longitudinal imaging and

clinical records, as in Alzheimer disease and multiple sclerosis. Protecting these data while still enabling sufficiently large and diverse datasets for artificial intelligence development therefore represents an important challenge. Federated learning has emerged as one potential technical approach, allowing models to be trained across multiple institutions while keeping patient-level data within the institution where they were generated. A foundational paper examining the future of digital health with federated learning demonstrated the feasibility of this approach in multi-institutional medical imaging and highlighted its potential to facilitate collaborative model development without requiring centralized pooling of sensitive patient data [42].

Despite this potential, the use of federated learning specifically in neurodegenerative disease research remains limited. Most of the models reviewed in this manuscript have been developed using data pooled at a single institution or within a small number of established research consortia rather than through federated infrastructures spanning multiple independent centers. Wider adoption of such approaches could potentially improve access to more diverse datasets while reducing some of the privacy concerns associated with centralized data sharing. However, federated learning should be viewed as a means of addressing one component of the data-sharing problem rather than as a complete solution to privacy and governance. Secure data handling, appropriate consent, institutional governance, and responsible management of sensitive clinical and genetic information remain necessary even when the underlying patient-level data are not directly centralized.

7.3 Explainability

The black-box nature of deep learning models remains one of the most consistently cited barriers to their clinical adoption across the literature reviewed in this manuscript [7,8,14]. Although post hoc explanation methods, including saliency maps and feature-attribution techniques, are increasingly used to make model predictions more interpretable, their ability to provide a reliable

representation of the model's actual reasoning remains uncertain. As discussed earlier, such explanations may change in response to relatively minor alterations in the input data, raising concerns about their stability and reproducibility [8]. This is particularly important in neurodegenerative disease, where an AI-generated prediction may influence diagnostic interpretation, prognostic assessment, or decisions made over a prolonged course of care. An explanation that is unstable or poorly aligned with the model's underlying decision process may therefore create a false sense of transparency rather than genuine clinical understanding. Explainability should consequently be treated not simply as a visual add-on to an otherwise opaque model, but as an important component of trustworthy clinical AI that requires careful validation before it can be relied upon in high-stakes decision making [7,8].

7.4 Limited clinical validation

A consistent pattern emerges across nearly every area examined in this review: impressive accuracy figures are frequently reported from retrospective studies conducted within single cohorts, whereas prospective, multicenter validation remains comparatively uncommon. This distinction is important because performance demonstrated within a controlled research cohort does not necessarily translate into equivalent performance when a model encounters patients, imaging protocols, clinical workflows, and demographic characteristics that differ from those represented during development. The amyotrophic lateral sclerosis F-wave model stands out precisely because of its unusually large validation cohort, which included tens of thousands of patients [22]. The fact that validation at this scale remains uncommon elsewhere in the literature, including in the amyotrophic lateral sclerosis prognostic studies discussed in Section 4.3, is itself informative about the current maturity of the evidence base [29,30]. Until models are consistently evaluated prospectively across independent, multicenter, and clinically diverse populations, reported research accuracy should be interpreted as evidence of potential rather than definitive evidence of clinical effectiveness.

7.5 Regulatory challenges

Regulatory pathways for artificial intelligence-based diagnostic and monitoring tools in neurodegenerative disease remain less established than in some more mature areas of clinical AI. Most of the tools reviewed in this manuscript have not yet undergone the type of prospective, regulator-facing clinical evaluation that would be required to support formal approval for routine diagnostic or monitoring use. This limitation is particularly evident in dementia neuroimaging, where the systematic review discussed in Section 4.7 found that the majority of published artificial intelligence tools remained retrospective research prototypes rather than clinically deployed decision-support systems [14]. Consequently, the transition from a promising research model to a regulated clinical technology remains an important translational step for the field. Demonstrating high accuracy in a research dataset is therefore only one component of regulatory

readiness; evidence of prospective performance, reproducibility, clinical utility, and safe integration into real-world clinical workflows is also necessary before widespread clinical adoption can be justified.

7.6 Cost and accessibility

Advanced neuroimaging, PET-based molecular imaging, and genetic testing remain relatively expensive and are geographically concentrated in high-income healthcare settings. Artificial intelligence tools that depend on these technologies inevitably inherit some of the same limitations in availability and access. This creates a potential equity concern: if increasingly sophisticated AI-driven diagnostic and prognostic tools are developed primarily around resource-intensive investigations, their benefits may remain concentrated in healthcare systems that can afford the underlying infrastructure. In such circumstances, artificial intelligence could unintentionally widen rather than reduce existing disparities in neurodegenerative care. Future development should therefore consider accessibility from the outset, including the feasibility of deploying AI tools in lower-resource settings, rather than treating equitable implementation as a secondary consideration after technical development has already been completed.

7.7 Integration into healthcare systems

Even a validated, explainable, and appropriately evaluated artificial intelligence model cannot provide meaningful clinical benefit unless it can be integrated into the healthcare environment in which it is intended to operate. Effective implementation requires compatibility with electronic health record systems, incorporation into existing clinical workflows, appropriate clinician training, and sustainable reimbursement and governance structures. These organizational and infrastructural requirements are often less visible than the technical challenges of model development, yet they can determine whether a clinically promising system is actually adopted and used effectively. The gap between developing a technically successful model and embedding that model into routine care may therefore represent one of the most difficult translational challenges in clinical artificial intelligence. Ultimately, successful implementation requires not only a model that performs well, but also a healthcare system capable of incorporating its outputs into everyday clinical practice without creating additional burden or disrupting established patient-care processes.

8. Future Perspectives

The bibliometric landscape indicates that artificial intelligence in neurodegenerative disease is developing as a rapidly expanding research field rather than representing a transient technological trend. A bibliometric analysis of research published since 2000 documented an exponential increase in publication volume after 2017, driven largely by advances in deep learning and the growing integration of multimodal data, with the United States and China together accounting for approximately half of the global research output [43]. However, increasing publication volume should not be equated with increasing clinical readiness. The more important question for the coming years is

whether this expanding research activity can be translated into systems that are independently validated, clinically useful, equitable, and safely integrated into routine care. Several directions appear particularly important in closing this translational gap.

Explainable artificial intelligence should move beyond post hoc attempts to rationalize predictions and increasingly favor models that incorporate interpretability into their design, particularly when they are intended to support high-stakes diagnostic or prognostic decisions. This direction is consistent with Rudin's argument that inherently interpretable models may be preferable to attempting to explain opaque models after they have already generated a prediction [7,8]. For neurodegenerative diseases, where AI outputs may influence diagnosis, prognosis, and long-term treatment planning, meaningful interpretability should therefore be regarded as a component of clinical safety rather than simply a desirable technical feature.

Federated learning provides a potential pathway toward developing larger and more representative datasets while reducing the need to centralize sensitive genetic, imaging, and clinical information. Building on frameworks demonstrated in broader medical imaging research, federated approaches could allow institutions to contribute to model development while retaining patient-level data locally [42]. Wider implementation could be particularly valuable in neurodegenerative research, where individual centers may lack sufficient numbers of patients or adequate demographic diversity to develop robust models independently. However, federated learning should complement rather than replace appropriate data governance, consent, security, and institutional oversight. Multimodal artificial intelligence represents another important direction. Neurodegenerative diseases manifest across multiple domains, including structural and functional brain changes, movement abnormalities, speech alterations, cognition, and molecular biomarkers. Models capable of integrating these complementary data streams may therefore provide a more comprehensive representation of disease than approaches based on a single modality in isolation. Early evidence from Parkinson disease, in which combining gait measures with cognitive status improved classification compared with gait information alone, supports the potential value of this strategy [20]. Future multimodal models should nevertheless be evaluated carefully to determine whether additional data genuinely improve clinical utility rather than simply increasing model complexity.

Foundation models and large language models are beginning to expand the role of artificial intelligence beyond conventional image and signal analysis. Their ability to process unstructured clinical text and spontaneous speech creates opportunities for extracting clinically relevant information that may otherwise remain difficult to analyze at scale. The speech-based dementia prediction work discussed in Section 4.1 illustrates this emerging direction, although the reliability, generalizability, and safety of such systems remain insufficiently established for routine clinical use [10,26]. Related research has also

explored prompting large language models to reproduce structured diagnostic reasoning rather than simply providing a final answer, potentially making their reasoning process more inspectable and improving clinician interaction with these systems [44]. Such approaches should, however, be evaluated on clinically meaningful outcomes rather than judged solely by the apparent plausibility of their generated reasoning.

Digital twins, which seek to model and simulate an individual's disease trajectory over time, remain largely conceptual within neurodegenerative medicine. If trajectory models can become sufficiently accurate and generalizable, they could eventually support individualized clinical trial design, prediction of disease progression, and exploration of potential treatment strategies. Huntington disease provides a particularly relevant foundation for this concept because machine learning has already demonstrated the ability to identify distinct progression patterns among genetically confirmed individuals [5]. The challenge will be determining whether such population-derived trajectories can be reliably translated into predictions that are sufficiently accurate for individual patients.

AI-assisted clinical trials may provide another avenue for improving the efficiency of neurodegenerative research. Patient stratification based on predicted disease trajectory, rate of progression, or other clinically meaningful characteristics could potentially help investigators identify participants more likely to demonstrate measurable change within a defined trial period. The trajectory modeling demonstrated in Huntington disease provides an early example of the type of stratification that could eventually support this approach [5]. However, prospective validation will be essential before AI-derived stratification can be relied upon for trial eligibility or treatment allocation.

Human-in-the-loop AI is likely to represent a more realistic near-term model of clinical implementation than fully autonomous decision making. Under this approach, clinicians can interrogate, contextualize, override, and correct model outputs rather than receiving an algorithmic prediction as a fixed conclusion. Such systems preserve professional accountability while allowing artificial intelligence to contribute computational capabilities that complement human expertise. This approach is also consistent with the concerns surrounding interpretability, reliability, and clinical responsibility discussed throughout this review [7,8]. Importantly, the goal should not simply be to place a human somewhere in the workflow, but to design the interaction so that human oversight genuinely improves the safety and quality of AI-supported decisions. Ethical AI governance will become increasingly important as the field expands. Governance frameworks should include systematic bias auditing, transparent reporting of dataset composition and model performance, accountability for AI-assisted decisions, and standards for monitoring performance after deployment. The ethical challenges surrounding bias, transparency, and accountability identified by Char, Shah, and Magnus provide a broader framework for these concerns in healthcare machine learning [40]. The documented disparities associated with

healthcare algorithms in adjacent clinical domains further demonstrate why the absence of a comparable large-scale bias assessment in neurodegenerative AI should not be interpreted as evidence that such bias does not exist [9]. Given the rapid expansion of research output in this field [43], establishing governance and evaluation standards alongside technological development—not after widespread deployment—will be essential. Overall, the future of artificial intelligence in neurodegenerative care is unlikely to be determined by whether algorithms can outperform clinicians on isolated research tasks. The more consequential question is whether

these systems can become reliable, interpretable, equitable, clinically validated, and meaningfully integrated into human-led care. Progress will therefore depend not only on better models, but also on better datasets, stronger prospective validation, responsible governance, and deliberate collaboration between artificial intelligence and the healthcare professionals who remain accountable for patient care.

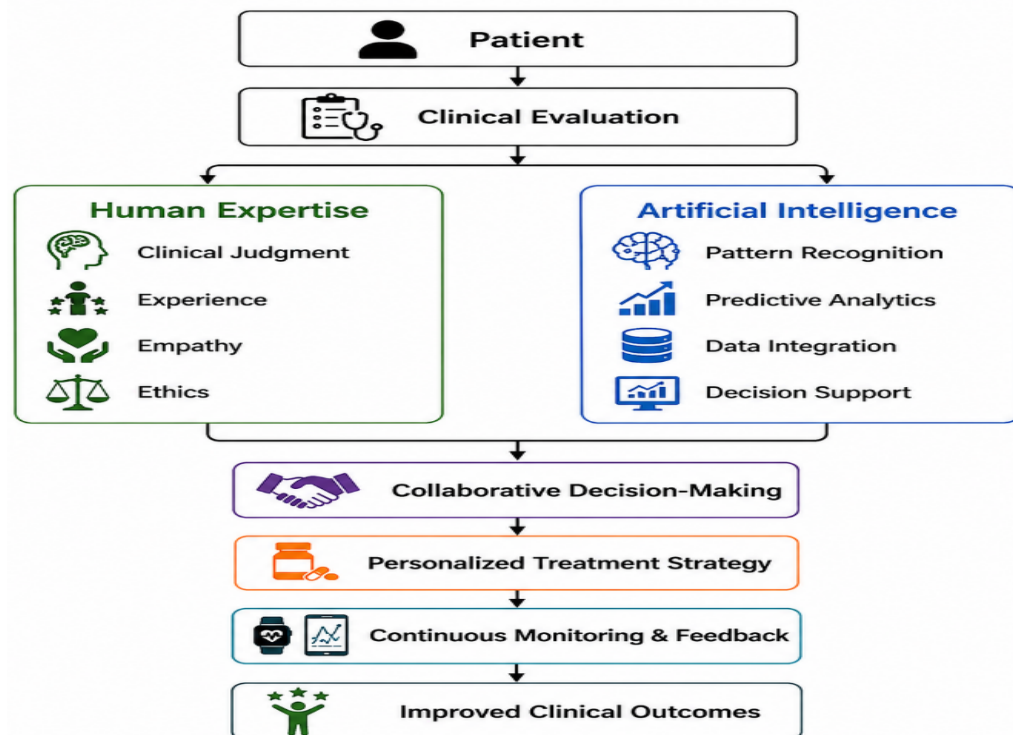


Figure 2. Human-AI Collaborative Framework for Precision Neurodegenerative Care

9. Conclusion

Artificial intelligence has substantially expanded the ability to recognize complex patterns across neurodegenerative disease imaging, gait and movement analysis, biomarker discovery, drug safety surveillance, and remote patient monitoring, while research activity in this field continues to grow rapidly [43]. However, the current evidence base remains uneven, and several limitations continue to separate promising research findings from clinically established applications. Across multiple disease areas, many published models have been evaluated retrospectively within single cohorts, whereas prospective and multicenter validation remains comparatively uncommon [14,16,17,18,19,29,30]. This limitation is particularly important because high performance within a controlled research dataset does not necessarily guarantee equivalent performance in diverse, real-world clinical populations.

Algorithmic fairness represents another unresolved concern. Although large-scale systematic bias audits

specific to neurodegenerative disease AI remain limited, evidence from healthcare more broadly demonstrates that algorithmic bias can have substantial and clinically consequential effects [9]. Similarly, interpretability remains both a technical and clinical challenge rather than a problem that can be considered solved simply by deploying existing models more widely [7,8]. At the same time, responsibilities such as ethical oversight, medication safety judgment, and the relational aspects of communicating with patients and families remain fundamentally human, even when artificial intelligence can assist with surveillance, information synthesis, or communication drafting [39,40,41].

Ultimately, the value of artificial intelligence in neurodegenerative care should not be measured solely by whether an algorithm can outperform a clinician on an isolated prediction task. Human expertise remains essential for longitudinal clinical judgment, interpretation of uncertain findings, ethical reasoning surrounding genetic risk and future care, and the communication and support

required throughout a prolonged and often unpredictable disease course. The most realistic path forward is therefore not a competition between artificial intelligence and clinicians, but a collaborative model in which algorithmic pattern recognition extends what clinicians can detect and analyze, while human expertise provides the clinical context, patient values, ethical judgment, and accountability required to translate those predictions into appropriate care. In this model, artificial intelligence does not replace the clinician; it strengthens the clinician's ability to make informed, individualized, and accountable decisions.

10. Conflict of Interest

The authors declare that they have no conflicts of interest....

REFERENCE

1. Peng S, Liu P, Wang X, Li K. Global, regional and national burden of Parkinson's disease in people over 55 years of age: a systematic analysis of the global burden of disease study, 1991-2021. *BMC Neurol.* 2025;25:178. DOI: 10.1186/s12883-025-04191-8.
2. Bevilacqua R, Barbarossa F, Fantechi L, Fornarelli D, Paci E, Bolognini S, et al. Radiomics and artificial intelligence for the diagnosis and monitoring of Alzheimer's disease: a systematic review of studies in the field. *J Clin Med.* 2023;12(16):5432. DOI: 10.3390/jcm12165432.
3. Vidovic M, Müschen LH, Brakemeier S, Machetanz G, Naumann M, Castro-Gomez S. Current state and future directions in the diagnosis of amyotrophic lateral sclerosis. *Cells.* 2023;12(5):736. DOI: 10.3390/cells12050736.
4. Segura T, Medrano IH, Collazo S, Maté C, Sguera C, Del Rio-Bermudez C, et al. Symptoms timeline and outcomes in amyotrophic lateral sclerosis using artificial intelligence. *Sci Rep.* 2023;13:702. DOI: 10.1038/s41598-023-27863-2.
5. Ko J, Furby H, Ma X, Long JD, Lu XY, Slowiejko D, et al. Clustering and prediction of disease progression trajectories in Huntington's disease: an analysis of Enroll-HD data using a machine learning approach. *Front Neurol.* 2023;13:1034269. DOI: 10.3389/fneur.2022.1034269.
6. Falet JPR, Nobile S, Szpindel A, Barile B, Kumar A, Durso-Finley J, et al. The role of AI for MRI-analysis in multiple sclerosis: a brief overview. *Front Artif Intell.* 2025;8:1478068. DOI: 10.3389/frai.2025.1478068.
7. Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access.* 2018;6:52138-52160. DOI: 10.1109/ACCESS.2018.2870052.
8. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* 2019;1(5):206-215. DOI: 10.1038/s42256-019-0048-x.
9. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447-453. DOI: 10.1126/science.aax2342.
10. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med.* 2023;29(8):1930-1940. DOI: 10.1038/s41591-023-02448-8.
11. Țîrnăucă C, Stan D, Meissner JM, Salas-Gómez D, Fernández-Gorgojo M, Infante J. A machine learning approach to detect Parkinson's disease by looking at gait alterations. *Mathematics.* 2022;10(19):3500. DOI: 10.3390/math10193500.
12. Warren SL, Moustafa AA. Functional magnetic resonance imaging, deep learning, and Alzheimer's disease: a systematic review. *J Neuroimaging.* 2023;33(1):5-18. DOI: 10.1111/jon.13063.
13. Yin W, Zhu W, Gao H, Niu X, Shen C, Fan X, et al. Gait analysis in the early stage of Parkinson's disease with a machine learning approach. *Front Neurol.* 2024;15:1472956. DOI: 10.3389/fneur.2024.1472956.
14. Borchert RJ, Azevedo T, Badhwar AP, Bernal J, Betts M, Bruffaerts R, et al. Artificial intelligence for diagnostic and prognostic neuroimaging in dementia: a systematic review. *Alzheimers Dement.* 2023;19(12):5885-5904. DOI: 10.1002/alz.13412.
15. Sale P, De Pandis MF, Le Pera D, Sova I, Cimolin V, Ancillao A, et al. Robot-assisted walking training for individuals with Parkinson's disease: a pilot randomized controlled trial. *BMC Neurol.* 2013;13:50. DOI: 10.1186/1471-2377-13-50.
16. Kaur A, Mittal M, Bhatti JS, Thareja S, Singh S. A systematic literature review on the significance of deep learning and machine learning in predicting Alzheimer's disease. *Artif Intell Med.* 2024;154:102928. DOI: 10.1016/j.artmed.2024.102928.
17. Ebrahimighahnavieh MA, Luo S, Chiong R. Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review. *Comput Methods Programs Biomed.* 2020;187:105242. DOI: 10.1016/j.cmpb.2019.105242.
18. Fathi S, Ahmadi M, Dehnad A. Early diagnosis of Alzheimer's disease based on deep learning: a systematic review. *Comput Biol Med.* 2022;146:105634. DOI: 10.1016/j.combiomed.2022.105634.
19. Frizzell TO, Glashutter M, Liu CC, Zeng A, Pan D, Hajra SG, et al. Artificial intelligence in brain MRI analysis of Alzheimer's disease over the past 12 years: a systematic review. *Ageing Res Rev.* 2022;77:101614. DOI: 10.1016/j.arr.2022.101614.
20. Ma L, Lin S, Jin J, Wang Z, Wang X, Chen Z, et al. Objective assessment of gait and posture symptoms in Parkinson's disease using wearable sensors and machine learning. *Front Aging Neurosci.* 2025;17:1618764. DOI: 10.3389/fnagi.2025.1618764.
21. Christodoulou RC, Woodward A, Pitsillos R, Ibrahim R, Georgiou MF. Artificial intelligence in Alzheimer's disease diagnosis and prognosis using PET-MRI: a narrative review of high-impact literature post-Tauvid approval. *J Clin Med.* 2025;14(16):5913. DOI: 10.3390/jcm14165913.
22. Martinez-Thompson JM, Mazurek KA, Parra-Cantu C, Naddaf E, Gogineni V, Botha H, et al. Artificial intelligence models using F-wave responses predict amyotrophic lateral sclerosis. *Brain.* 2025;148(7):2320-2330. DOI: 10.1093/brain/awaf014.
23. Orzan F, Iancu ȘD, Dioșan L, Bálint Z. Textural analysis and artificial intelligence as decision support tools

- in the diagnosis of multiple sclerosis: a systematic review. *Front Neurosci.* 2025;18:1457420. DOI: 10.3389/fnins.2024.1457420.
24. Klöppel S, Stonnington CM, Barnes J, Chen K, Chu C, Good CD, et al. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. *Brain.* 2008;131(11):2969-2974. DOI: 10.1093/brain/awn239.
25. Mittal R, Agarwal N, Dubey M, Pathak V, Shukla P, Rani G, et al. Machine learning approach to gait analysis for Parkinson's disease detection and severity classification. *Front Robot AI.* 2025;12:1623529. DOI: 10.3389/frobt.2025.1623529.
26. Agbavor F, Liang H. Predicting dementia from spontaneous speech using large language models. *PLOS Digit Health.* 2022;1(12):e0000168. DOI: 10.1371/journal.pdig.0000168.
27. Hansson O, Blennow K, Zetterberg H, Dage J. Blood biomarkers for Alzheimer's disease in clinical practice and trials. *Nat Aging.* 2023;3(5):506-519. DOI: 10.1038/s43587-023-00403-3.
28. Imamura K, Yada Y, Izumi Y, Morita M, Kawata A, Arisato T, et al. Prediction model of amyotrophic lateral sclerosis by deep learning with patient induced pluripotent stem cells. *Ann Neurol.* 2021;89(6):1226-1233. DOI: 10.1002/ana.26047.
29. Kew SYN, Mok SY, Goh CH. Machine learning and brain-computer interface approaches in prognosis and individualized care strategies for individuals with amyotrophic lateral sclerosis: a systematic review. *MethodsX.* 2024;13:102765. DOI: 10.1016/j.mex.2024.102765.
30. Papaiz F, Dourado MET, Valentim RAM, de Moraes AHF, Arrais JP. Machine learning solutions applied to amyotrophic lateral sclerosis prognosis: a review. *Front Comput Sci.* 2022;4:869140. DOI: 10.3389/fcomp.2022.869140.
31. Jo T, Nho K, Saykin AJ. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. *Front Aging Neurosci.* 2019;11:220. DOI: 10.3389/fnagi.2019.00220.
32. Fontanella F, D'Alessandro T, Nardone E, De Stefano C, Vicidomini C, Roviello GN. Artificial intelligence for natural products drug discovery in neurodegenerative therapies: a review. *Biomolecules.* 2026;16(1):129. DOI: 10.3390/biom16010129.
33. Cortial L, Montero V, Tourlet S, Del Bano J, Blin O. Artificial intelligence in drug repurposing for rare diseases: a mini-review. *Front Med.* 2024;11:1404338. DOI: 10.3389/fmed.2024.1404338.
34. Wan Z, Sun X, Li Y, Chu T, Hao X, Cao Y, Zhang P. Applications of artificial intelligence in drug repurposing. *Adv Sci.* 2025;12(14):e2411325. DOI: 10.1002/advs.202411325.
35. Kakoti BB, Bezbaruah R, Ahmed N. Therapeutic drug repositioning with special emphasis on neurodegenerative diseases: threats and issues. *Front Pharmacol.* 2022;13:1007315. DOI: 10.3389/fphar.2022.1007315.
36. Ó Breasail M, Biswas B, Smith MD, Mazhar MKA, Tenison E, Cullen A, et al. Wearable GPS and accelerometer technologies for monitoring mobility and physical activity in neurodegenerative disorders: a systematic review. *Sensors.* 2021;21(24):8261. DOI: 10.3390/s21248261.
37. ZhuParris A, de Goede AA, Yocarini IE, Kraaij W, Groeneveld GJ, Doll RJ. Machine learning techniques for developing remotely monitored central nervous system biomarkers using wearable sensors: a narrative literature review. *Sensors.* 2023;23(11):5243. DOI: 10.3390/s23115243.
38. Khanna A, Stephenson D, Jones G. Editorial: digital remote patient monitoring in neurodegenerative diseases. *Front Digit Health.* 2025;7:1725819. DOI: 10.3389/fdgth.2025.1725819.
39. Hu Q, Li J, Li X, Zou D, Xu T, He Z. Machine learning to predict adverse drug events based on electronic health records: a systematic review and meta-analysis. *J Int Med Res.* 2024;52(12):3000605241302304. DOI: 10.1177/03000605241302304.
40. Char DS, Shah NH, Magnus D. Implementing machine learning in health care: addressing ethical challenges. *N Engl J Med.* 2018;378(11):981-983. DOI: 10.1056/NEJMp1714229.
41. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med.* 2023;183(6):589-596. DOI: 10.1001/jamainternmed.2023.1838.
42. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *NPJ Digit Med.* 2020;3:119. DOI: 10.1038/s41746-020-00323-1.
43. Zhang Y, Yu L, Lv Y, Yang T, Guo Q. Artificial intelligence in neurodegenerative diseases research: a bibliometric analysis since 2000. *Front Neurol.* 2025;16:1607924. DOI: 10.3389/fneur.2025.1607924.
44. Savage T, Nayak A, Gallo R, Rangan E, Chen JH. Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. *NPJ Digit Med.* 2024;7:20. DOI: 10.1038/s41746-024-01010-1...