Int. j. adv. multidisc. res. stud. 2026; 6(6):625-630

Received: 24-09-2026
Accepted: 04-11-2026

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

In-Silico Prediction of Alzheimer's Disease Using Computational Approaches

[1] Sreeja Nannapaneni, [2] G Kalyani Akshita, [3] Takshil Kishore Jukunta, [4] T Jyotsna Chowdary, [5] Amara Sai Sri Nimilitha

[1] Independent Researcher, Troy, Michigan, United States of America
[2, 3, 4, 5] Department of Biotechnology, KL University, Guntur, Andhra Pradesh, India

DOI: https://doi.org/10.62225/2583049X.2026.6.6.7169
Corresponding Author: **Sreeja Nannapaneni**

Abstract

Alzheimer's disease (AD) is a progressive neurodegenerative disorder in which biological changes may appear prior to cognitive decline. The increasing availability of clinical, cognitive, neuroimaging, genetic, and fluid biomarker data has made machine learning (ML) a potentially suitable approach for identifying patterns associated with AD diagnosis and future progression. This review focuses on how the choice of data, prediction task, and validation strategy influences the reported performance of certain ML models in AD research. Conventional approaches, including logistic regression, support vector machines, and ensemble methods, remain useful for structured or engineered features, whereas deep learning models, particularly convolutional neural networks, are widely used in neuroimaging. Evidence from systematic reviews and cross-cohort studies suggests that high performance is obtainable in distinguishing established AD from cognitively unimpaired individuals, but predicting progression from mild cognitive impairment (MCI) is comparatively complex. Direct comparison of reported accuracy or area under the receiver operating characteristic curve (AUC) becomes difficult due to differences in cohorts, preprocessing, feature selection, data splitting, and external validation. Recent methodological evidence also shows that data leakage and inadequate control of confounding can substantially inflate reported performance. Therefore, the central challenge for ML is no longer demonstrating that prediction is possible, but establishing whether predictions remain reliable across independent populations and clinically relevant settings. Future research should prioritise subject-level validation, external and prospective testing, multimodal but clinically feasible data integration, interpretable modeling and transparent reporting. These considerations are essential if ML is to move from promising retrospective research applications toward reproducible clinical decision support.

Keywords: Alzheimer's Disease, In-silico, PET, CNN

1. Introduction

Alzheimer's disease (AD) is the most common cause of dementia and represents a major and increasing public-health burden. In the United States, an estimated 7.4 million people aged 65 years or older were living with clinical Alzheimer's dementia in 2026 [1]. AD is now increasingly understood as a biological process rather than a condition defined only by the presence of dementia. The 2024 revised Alzheimer's Association criteria incorporated biomarkers of amyloid-β and tau into the biological definition and staging of disease [2]. This shift has increased interest in identifying individuals at earlier stages and estimating how their disease is likely to progress.

Machine learning has become significant in this area because AD research generates data at multiple levels, including cognitive testing, demographic and clinical variables, magnetic resonance imaging (MRI), positron emission tomography (PET), genetic information, cerebrospinal fluid (CSF), and blood-based biomarkers [3, 4]. ML can accommodate combinations of these variables and identify associations that may be difficult to capture using a small number of manually specified predictors. The clinical question, available sample size, representation of the target population, and intended use of the prediction are equally important in justifying the use of complex ML models.

A further distinction between diagnostic classification and prediction of future disease is needed. Classifying an individual with established AD against a cognitively unimpaired control is not equal to predicting which person with MCI will progress to AD dementia. The latter requires information that enables prediction of a future event in individuals who may have relatively

subtle abnormalities. Consequently, reported performance values should be interpreted in relation to the prediction task rather than treated as interchangeable measures of algorithmic quality.

This review, therefore, focuses on a narrower question: How data, model choice, and validation strategy shape the reported usefulness of ML for AD detection and progression prediction.

2. Data Modalities and Prediction Tasks

The efficiency of an ML model is closely linked to the information it receives. In AD research, commonly used inputs can be grouped into clinical and cognitive measures, neuroimaging, genetic information, and fluid biomarkers. These modalities differ not only in biological content but also in cost, availability, dimensionality and susceptibility to technical variation.

2.1 Clinical and cognitive data

Clinical and cognitive variables are among the most accessible predictors. Age, education, medical history and cognitive measures such as the Mini-Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) are frequently used in risk and progression models. Longitudinal resources such as the National Alzheimer's Coordinating Center (NACC) Uniform Data Set were developed to standardise clinical and cognitive assessments across Alzheimer Disease Centers [5]. These variables are beneficial for ML models because they can be collected in routine settings, but they are not specific to AD pathology. Cognitive performance can also be influenced by education, language, cultural background, and other conditions. Therefore, models based only on clinical measures may have limited ability to distinguish AD pathology from other causes of cognitive impairment.

2.2 Neuroimaging

MRI and PET have dominated the ML literature for AD, particularly because imaging provides information about structural or molecular changes in the brain. MRI-based models commonly use regional volumes and cortical thickness, while PET can provide metabolic or molecular information. Systematic reviews show that MRI has been the most frequently investigated imaging modality, with substantial use of both conventional ML and deep learning [6, 4]. Three-dimensional CNNs are particularly suited to volumetric imaging because they can learn spatial representations directly from voxels [7].

However, imaging also introduces sources of variation that can affect generalizability. Scanner characteristics, acquisition protocols, preprocessing pipelines and differences between imaging centers can become associated with the outcome. A model may therefore learn site-related characteristics alongside disease-related patterns. This is one reason why performance within a single imaging cohort should not be interpreted as evidence that the model will perform similarly in another clinical environment.

2.3 Genetic and fluid biomarkers

Genetic information, including APOE ε4 and broader genetic profiles, can provide information about susceptibility, whereas CSF and blood-based biomarkers provide information more closely related to AD-associated biological processes. Multimodal dementia datasets increasingly combine biomarkers with imaging, cognitive and demographic information [3]. Blood-based biomarkers are particularly relevant to future clinical implementation because blood collection is less invasive than CSF sampling. Nevertheless, biomarker concentrations can vary with assay methods, sample handling and disease stage, so models incorporating these measurements also require careful validation.

2.4 Prediction tasks

Three tasks are particularly important when evaluating the ML literature. First, diagnostic classification seeks to distinguish established AD from cognitively unimpaired individuals and other clinical groups. Second, early-risk modelling attempts to identify individuals who are likely to develop clinically meaningful disease. Third, longitudinal prediction estimates future cognitive decline or conversion from MCI to AD dementia. The distinction matters because a model can perform very well on a relatively easy classification problem while providing considerably less accurate predictions of future progression.

3. Machine Learning Approaches: Matching the Model to the Data

AD ML studies use a broad range of methods, but the literature does not support the idea that a single algorithm is inherently superior. Instead, model choice is largely determined by the representation of the data and the prediction task [3]. For structured clinical, cognitive, or biomarker data, logistic regression, support vector machines (SVMs), random forests, and gradient-boosting methods remain useful. For high-dimensional images, CNNs and related deep-learning architectures are more commonly used.

3.1 Conventional machine learning

Logistic regression is useful as an interpretable baseline because its coefficients can be interpreted as the contribution of predictors under the model's assumptions. SVMs have been particularly common in AD studies because they can operate in high-dimensional feature spaces and have been widely applied to imaging-derived features. Random forests and gradient-boosting methods can represent nonlinear relationships and interactions among structured predictors. These methods can therefore be competitive when the dataset contains engineered imaging measurements, clinical variables, or biomarkers rather than raw images.

An important point is that preprocessing and feature selection can influence results as much as the classifier itself. For example, an OASIS-2 study comparing SVM, random forest and logistic regression found that feature-selection strategy materially affected performance, with logistic regression combined with minimum redundancy maximum relevance feature selection producing high reported accuracy [8]. Such findings argue against interpreting the classifier name alone as the main determinant of performance.

3.2 Deep learning

Deep learning differs mainly in its ability to learn representations from complex inputs rather than relying entirely on manually engineered features. CNNs have consequently become prominent in MRI- and PET-based

AD research [4, 7]. Deep models can represent spatial relationships that are difficult to capture through a small set of predefined imaging features, but this flexibility comes with greater computational and data requirements and a greater need for rigorous validation.

Importantly, greater model complexity has not consistently produced better external performance. In a cross-cohort study, conventional and deep-learning models showed similar performance for AD-versus-control classification within the development cohort, while performance declined in an independent cohort [9]. The same study found substantially weaker external performance for MCI conversion than for established AD classification. This illustrates a broader point: The value of an architecture cannot be separated from the difficulty of the prediction problem or the population in which it is evaluated.

3.3 Multimodal and hybrid models

Multimodal models attempt to combine complementary information, such as MRI with cognitive measures, PET with clinical variables, or imaging with genetic and fluid biomarkers. The rationale is biologically plausible: Structural imaging may capture neurodegeneration, biomarkers may capture pathology, and cognitive measures may reflect functional consequences. However, multimodal integration is not automatically beneficial. Complete-case analyses may substantially reduce the sample size because not every participant has every modality. Different modalities may also be measured at different time points and with different levels of reliability [3].

Hybrid approaches can also combine deep-learning feature extraction with a conventional classifier. These systems may be useful, but they add another layer of complexity and can make it harder to determine which component is responsible for the observed performance. For a clinical system, the practical question is therefore not how many data types or algorithms can be combined, but whether the additional information produces a reproducible improvement that is feasible to obtain in routine care.

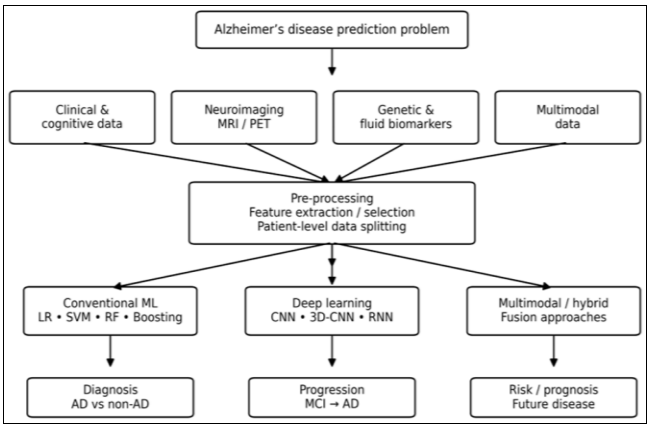


Fig 1: Conceptual workflow for machine-learning-based Alzheimer's disease prediction

4. What Does the Current Evidence Show

Reported performance across AD ML studies is often high, particularly for AD-versus-cognitively-unimpaired classification. However, the evidence becomes less consistent when the target is future progression. This difference is important because diagnostic classification is often based on established disease, whereas progression prediction must distinguish future outcomes among people who may currently have similar clinical profiles.

4.1 AD versus cognitively unimpaired classification

A systematic review of AI-based brain MRI studies found that CNN approaches were associated with high reported classification accuracy, although the included studies differed substantially in datasets and methodology [6]. A separate review of ML and deep-learning studies likewise identified MRI and PET as dominant modalities and ADNI as the most frequently used dataset [4]. These findings demonstrate the technical feasibility of image-based classification, but they do not constitute a controlled comparison between algorithms because the models were developed under different experimental conditions.

The cross-cohort study by Bron *et al.* provides a more informative example because conventional and deep-learning approaches were evaluated using the same general imaging problem and then tested across cohorts [9]. Within ADNI, the reported AUC values for AD-versus-control classification were 0.940 for SVM and 0.933 for CNN. In an independent cohort, the corresponding values fell to 0.896 and 0.876. The important observation is not which model had the larger value, but that both showed a reduction when transferred outside the development environment.

4.2 Prediction of MCI conversion

Prediction of conversion from MCI to AD dementia presents a more difficult and clinically relevant problem. A systematic review of 116 neuroimaging studies found that SVM was the most commonly used algorithm, with a mean accuracy of 75.4%, while CNN approaches had a mean accuracy of 78.5%. Studies combining MRI and PET generally reported higher classification accuracy than single-modality approaches [10]. These results suggest that additional information can be useful, but they should not be interpreted as evidence that CNNs or multimodal imaging are universally superior.

A systematic review and meta-analysis of prediction models identified AUC values of 0.67–0.98 for development models and 0.62–0.96 for validation models. MRI, APOE ε4, age, MMSE and ADAS-Cog were among the predictors most frequently included [11]. However, the review also found that many models lacked external validation, limiting confidence in their applicability to broader populations [11]. The gap between development and independent validation is therefore central to interpreting the literature.

4.3 Why performance numbers cannot be compared directly

Accuracy, sensitivity, specificity and AUC are useful metrics, but they are conditional on the population, outcome definition and validation strategy. High accuracy obtained from a carefully selected research cohort with clearly separated AD and control groups may reflect a different level of difficulty than predicting future conversion in a clinically heterogeneous MCI population. Even within the same disease category, differences in preprocessing, feature selection, class balance and threshold selection can change the reported values.

For this reason, a meaningful review should emphasise patterns across studies rather than compile the highest accuracy reported by each algorithm. The available evidence supports the feasibility of ML-based prediction, but it is less

convincing as evidence that one model family consistently outperforms others across independent populations.

5. Validation, Bias and Sources of Inflated Performance

The most important limitation in the AD ML literature is the difficulty of separating genuine generalisable prediction from performance that is specific to the development dataset. Research cohorts such as ADNI have been invaluable for advancing the field, but repeated use of the same highly characterised cohort can reduce variation in participant characteristics, imaging protocols and clinical environments [4]. NACC provides a broader longitudinal clinical resource, but recent work has also demonstrated temporal and centre-level differences in participant representation, reinforcing the need to consider cohort composition when interpreting results [12].

5.1 Data leakage

Data leakage occurs when information that should be unavailable to the model at evaluation influences model development. In imaging studies, one of the clearest examples is allowing scans from the same participant to appear in both training and test sets. Feature selection, normalisation or preprocessing performed before participant-level splitting can also allow information from the evaluation data to enter the training process.

A 2025 scoping review specifically examining data leakage in deep-learning studies of AD found substantial methodological concerns and showed that reported performance can fall markedly when appropriate validation is applied [13]. The review identified inadequate confounder control and limited external validation as recurring problems. These findings reinforce the need for participant-level separation and transparent reporting of the complete modelling pipeline.

5.2 External validation

Internal cross-validation can estimate performance within the distribution represented by the development dataset, but external validation addresses a different question: Whether the model transfers to an independent population. The distinction is particularly important in AD because datasets can differ in age distribution, diagnostic procedures, disease severity, ethnicity, comorbidity, scanner characteristics and referral patterns. The cross-cohort findings reported by Bron *et al*. illustrate how apparently strong within-cohort performance can decline outside the development setting [9].

5.3 Class imbalance and confounding

Accuracy can also be misleading when diagnostic groups are unbalanced. Sensitivity, specificity, precision, recall, F1-score and AUC should therefore be considered together, with calibration reported when the model is intended to estimate individual risk. Confounding is equally important. Models may exploit age, sex, site, scanner or other variables that differ systematically between groups rather than learning disease-specific patterns. These issues are not merely technical details; they determine whether a prediction reflects the intended biological or clinical signal.

5.4 Representativeness

Large sample size does not automatically imply representativeness. UK Biobank, for example, provides population-scale data but was not designed as an AD-specific clinical cohort. Similarly, NACC participants are recruited through specialised Alzheimer Disease Research Centers [12]. Consequently, researchers should report who was included, how the outcome was defined and how the study population relates to the population in which the model is intended to be used.

6. Interpretability and Clinical Translation

Clinical translation requires more than a high AUC. A useful model should provide information that is relevant to a clinical decision, remain sufficiently reliable when patient characteristics differ from the training population and operate within realistic workflow and data-availability constraints [3, 14]. This creates a distinction between technical explainability and clinical interpretability.

Methods such as feature importance measures, SHAP values and imaging saliency maps can help describe which variables or regions influence a prediction. However, an explanation of model behaviour should not automatically be treated as evidence of disease mechanism. A model may rely on a correlated variable, acquisition artefact or demographic characteristic. Interpretability should therefore be evaluated alongside robustness and external validity rather than used as a substitute for them.

Clinical feasibility also places limits on multimodal modelling. A research model requiring MRI, PET, CSF, genetic testing and repeated cognitive assessments may be scientifically informative but difficult to deploy where only a subset of these measurements is routinely available. Models based on blood biomarkers and routinely collected clinical data may ultimately have a different implementation pathway from highly specialised imaging systems. The appropriate comparison is therefore not simply predictive accuracy, but predictive performance relative to the information, cost and workflow required to obtain it.

Ethical considerations are also relevant. ML systems can reproduce or amplify biases present in their training data, while the collection and integration of clinical and genetic information raise privacy concerns. These issues are especially important when models are applied to populations that were underrepresented in the development data [14]. Transparent reporting of cohort composition and subgroup performance should therefore be considered part of model evaluation rather than an optional addition.

7. Future Direction

Future progress in AD ML is likely to depend less on increasingly elaborate architectures and more on better study design. Multicentre datasets with consistent definitions, participant-level splitting and independent validation should receive greater emphasis. External validation should become routine, followed where feasible by prospective evaluation in the clinical environment in which a model is intended to operate.

Longitudinal modelling is another important direction. Rather than treating AD as a static classification problem, future systems should estimate clinically meaningful outcomes such as probability of MCI-to-AD conversion, rate of cognitive decline or time to progression. Such models must respect the timing of observations and avoid incorporating information that would not have been available at the point at which the prediction is supposed to be made.

Multimodal modelling is likely to remain important,

particularly as blood-based biomarkers become more widely available. However, future studies should explicitly test whether each additional modality provides incremental value and whether the resulting model remains useful when some measurements are missing. Models that are slightly less complex but easier to obtain and validate may ultimately be more useful than models requiring extensive research-only measurements.

Finally, reporting standards need to become more consistent. Studies should clearly describe participant selection, outcome definitions, sample size, preprocessing, feature selection, train-test separation, hyperparameter tuning, class balancing, performance metrics, confidence intervals, calibration, external validation and subgroup performance. Such reporting would make the literature easier to compare and would reduce the risk that apparently strong results are driven by methodological differences rather than genuine improvements in prediction.

8. Conclusion

Machine learning has considerable potential for AD research because it can integrate clinical, cognitive, imaging, genetic and biomarker information. Conventional methods remain useful for structured and engineered features, while deep learning has expanded the analysis of high-dimensional neuroimaging. However, the literature does not establish a universally superior algorithm. Instead, performance depends strongly on the prediction task, data representation and validation strategy.

The clearest distinction in the evidence is between established-disease classification and future progression prediction. High performance is frequently reported for AD-versus-control classification, whereas MCI-to-AD prediction is more uncertain and is often limited by inadequate external validation [9-11]. Across both areas, data leakage, cohort selection, confounding, class imbalance and differences between acquisition sites can produce performance that does not transfer to independent populations [13].

The next stage of AD ML research should therefore focus on reproducibility and clinical relevance rather than algorithmic complexity alone. Independent and prospective validation, clinically feasible multimodal data, transparent reporting and interpretable predictions are likely to be more important for translation than another incremental increase in model complexity. The field has moved beyond demonstrating that ML can identify patterns associated with AD; the remaining challenge is demonstrating that those predictions remain reliable when they are used on the patients and clinical environments for which they are intended.

9. References

1. Alzheimer's Association. 2026 Alzheimer's disease facts and figures. Alzheimer's & Dementia. 2026; 22(4):e71345. Doi: 10.1002/alz.71345
2. Jack CR Jr, Andrews JS, Beach TG, *et al*. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. Alzheimer's & Dementia. 2024; 20(8):5143-5169. Doi: 10.1002/alz.13859
3. Wang Y, Liu S, Spiteri AG, *et al*. Understanding machine learning applications in dementia research and clinical practice: A review for biomedical scientists and clinicians. Alzheimer's Research & Therapy. 2024; 16:175. Doi: 10.1186/s13195-024-01540-6
4. 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. Artificial Intelligence in Medicine. 2024; 154:102928. Doi: 10.1016/j.artmed.2024.102928
5. Morris JC, Weintraub S, Chui HC, *et al*. The Uniform Data Set (UDS): Clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. Alzheimer Disease & Associated Disorders. 2006; 20(4):210-216. Doi: 10.1097/01.wad.0000213865.09806.92
6. Frizzell TO, Glashutter M, Liu CC, *et al*. Artificial intelligence in brain MRI analysis of Alzheimer's disease over the past 12 years: A systematic review. Ageing Research Reviews. 2022; 77:101614. Doi: 10.1016/j.arr.2022.101614
7. Zia-ur-Rehman, Awang MK, Ali G, Faheem M. Deep learning techniques for Alzheimer's disease detection in 3D imaging: A systematic review. Health Science Reports. 2024; 7(9):e70025. Doi: 10.1002/hsr2.70025
8. Alshamlan H, Alwassel A, Banafa A, Alsaleem L. Improving Alzheimer's disease prediction with different machine learning approaches and feature selection techniques. Diagnostics. 2024; 14(19):2237. Doi: 10.3390/diagnostics14192237
9. Bron EE, Klein S, Papma JM, *et al*. Cross-cohort generalizability of deep and conventional machine learning for MRI-based diagnosis and prediction of Alzheimer's disease. NeuroImage: Clinical. 2021; 31:102712. Doi: 10.1016/j.nicl.2021.102712
10. Grueso S, Viejo-Sobera R. Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: A systematic review. Alzheimer's Research & Therapy. 2021; 13(1):162. Doi: 10.1186/s13195-021-00900-w
11. Chen Y, Qian X, Zhang Y, Su W, Huang Y, Wang X, *et al*. Prediction models for conversion from mild cognitive impairment to Alzheimer's disease: A systematic review and meta-analysis. Frontiers in Aging Neuroscience. 2022; 14:840386. Doi: 10.3389/fnagi.2022.840386
12. Chan KCG, Xia F, Kukull WA. NACC data: Who is represented over time and across centers, and implications for generalizability. Alzheimer's & Dementia. 2025; 21(9):e70657. Doi: 10.1002/alz.70657
13. Young VM, Gates S, Garcia LY, Salardini A. Data leakage in deep learning for Alzheimer's disease diagnosis: A scoping review of methodological rigor and performance inflation. Diagnostics. 2025; 15(18):2348. Doi: 10.3390/diagnostics15182348
14. Hofweber T, Walker RL. Machine learning in health care: Ethical considerations tied to privacy, interpretability, and bias. North Carolina Medical Journal. 2024; 85(4):240-245. Doi: 10.18043/001c.120562
15. Aisen PS, Petersen RC, Donohue MC, *et al*. Clinical core of the Alzheimer's Disease Neuroimaging Initiative: Progress and plans. Alzheimer's & Dementia. 2010; 6(3):239-246. Doi: 10.1016/j.jalz.2010.03.006
16. Marcus DS, Wang TH, Parker J, *et al*. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive

Neuroscience. 2007; 19(9):1498-1507. Doi: 10.1162/jocn.2007.19.9.1498
17. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. New England Journal of Medicine. 2019; 380(14):1347-1358. Doi: 10.1056/NEJMra1814259