Title: AI genetics for Alzheimer's disease prediction.

Alzheimer's disease (AD) affects millions of people worldwide and is the most common cause of dementia in older adults. The disease has a strong genetic component, with studies identifying numerous genetic variants that can increase or decrease the risk of developing AD. Traditional genetic-based methods for predicting AD risk have limited accuracy because the disease involves complex interactions between multiple genes and environmental factors. Artificial intelligence (AI) and machine learning offer new ways to analyze large amounts of genetic data and find patterns that have not been identified before. These AI systems can process genetic information from thousands of individuals to identify combinations of genetic markers that predict AD risk more accurately than single gene tests. Early prediction of AD is crucial as it can help practitioners to start treatments sooner and help patients in time. Recent advances in AI have made it possible to combine genetic data with other types of information like brain scans and cognitive test results to improve prediction accuracy.

# Machine learning approaches and algorithms

Machine learning offers several powerful approaches for predicting AD from genetic data, with performance varying significantly across different methods and datasets [1], [2]. Traditional machine learning algorithms like Support Vector Machine (SVM), Random Forest, and logistic regression have been widely applied to genetic datasets, with reported classification accuracies ranging from 0.65 to 0.975, though realistic performance expectations are more modest [3], [4], [5], [6]. Recent studies have achieved notable results, with SVM models reaching 89% accuracy in detecting AD using genome-wide association study data [7]. Deep learning approaches are increasingly being used to handle the complexity of genetic data, though they face challenges when working with genetic variants alone [8]. Novel deep learning frameworks like Deep-Block use multi-stage approaches that incorporate biological knowledge, combining genome segmentation based on linkage patterns with attention mechanisms and ensemble methods to identify genetic regions associated with AD [9]. Other innovative approaches include transformer-based models that preserve the sequence structure of genetic variants and use uncertainty estimation to improve prediction reliability [10]. The trend in machine learning for AD prediction is moving toward multi-feature datasets rather than single biomarker approaches, with AI systems trained on combinations of genetic, neuroimaging, and clinical data [11], [12]. Recent work has also developed specialized AI tools like AD-GPT, which combines large language models with biomedical data to enhance genetic information retrieval and analysis for AD research [13]. However, researchers acknowledge that current dataset sizes limit expected performance to between 0.55 and 0.7 AUC for genetic-only prediction models, with higher reported accuracies often resulting from overfitting [3].

# Performance and accuracy metrics

Performance metrics for genetic AI prediction of AD vary widely across studies, with reported accuracies ranging from 59% to 99% AUC depending on the methodology and data types used [14]. However, studies suggest that realistic expectations for genetic only prediction models should be between 0.55 and 0.7 AUC given current datasets sizes., with higher reported accuracies likely results from overfitting [3]. Traditional machine learning approaches using genetic data alone typically achieve more modest results, with one comprehensive study reporting approximately 72% area under the ROC curve as the best performance for predicting late-onset AD from genetic variation data [1]. Recent advances in AI have achieved notably higher performance when combining multiple data types. Deep learning models using epigenomic data from blood samples have achieved AUC values of 0.93-0.99 with 97% sensitivity and specificity [15]. Multimodal approaches that integrate genetic data with neuroimaging show promising results, with one study achieving 83.78% classification accuracy and 0.924 AUC-ROC using both MRI scans and genetic sequencing data [16]. Advanced AI models can predict Alzheimer's disease up to 75.8 months before final diagnosis using neuroimaging, achieving 82% specificity at 100% sensitivity [11], [12]. More recent studies continue to show encouraging results across different approaches. Network-based models that integrate brain connectivity with genetic data achieve AUC values of 0.684 for combined approaches, improving to 0.778 when including clinical covariates like sex and APOE genotypes [17]. Genetic risk scoring approaches demonstrate that individuals in the top decile of genetic risk scores have ten-fold increased odds compared to those in the bottom decile [18].  Some specialized models using gene expression data and machine learning have reported perfect 100% accuracy in cross-validation studies [19], while more recent transformer-based models achieve 99% accuracy when combining RNA sequencing data with brain imaging [20].

# Genetic markers and features

Key Genetic Markers and Features Used in AI Alzheimer's Prediction:

• *APOE gene variants* - The primary genetic risk factor for Alzheimer's disease, with APOE ε4 being the most significant risk allele and APOE ε2 providing protective effects [21], [22].

• *Specific SNPs identified through AI* - Including rs429358 and rs769449 within APOE, rs4821510 (most important SNP for detection), and rs429358 (indication for Alzheimer's disease) [7], [10].

• *Novel gene targets from AI analysis* - THAP9-AS1 identified as a top noncoding region target, and ORAI2 discovered as a shared biomarker across frontal, hippocampal, and temporal brain regions [19], [23]

• *Calcium signaling pathway genes* - SNPs in PRKCZ, PLCB1, and ITPR2, along with genes related to Ca2+ ion release in affected brain regions [24].

• *Gene expression biomarkers* - HLA-DQB1, EIF1AY, HLA-DQA1, and ZFP57 expression levels identified through machine learning analysis of blood samples [24].

• *Hub genes for disease progression* - ACBD5, GABARAPL1, and HSPA8 identified as key genes associated with Alzheimer's progression [25].

• *Epigenetic markers* - DNA methylation patterns (CpG sites) across the genome, with hundreds of new significant brain CpGs predicted using machine learning approaches like EWASplus [26], [27].

*• Polygenic risk scores* - Combinations of multiple SNPs that provide stronger predictive power than single gene tests, with optimal performance achieved using fewer than 100 causal SNPs [18], [28].

• *Protein-level genetic predictors* - Genetically predicted protein levels in plasma used as instruments to investigate associations with Alzheimer's risk [29].

• *Familial Alzheimer's mutations* - APP, PSEN1, and APOE4 mutations that contribute to amyloid beta and tau pathology [30].

Table 1. Literature Comparison

|  |  |  |  |
| --- | --- | --- | --- |
| Papers | Key Genetic Markers | Ai And Machine Learning Techniques | Innovative Methodologies |
| [Gupta et al, 2024](https://www.semanticscholar.org/p/274744373) | APOE genotype | N/A | AI techniques for senhanced precision in biomarker analysis and neuroimaging standardization. |
| ECSOC 2024 |
| [Jo et al, 2024](https://www.semanticscholar.org/p/272768425) | APOE rs429358, rs769449, novel SNPs in top 1,500 LD blocks. | Deep learning framework with TabNet and Random Forest algorithms. | Deep-Block framework: multi-stage deep learning with biological knowledge for SNP feature importance quantification. |
| medRxiv (2 citations) |
| [Sekaran et al, 2023](https://www.semanticscholar.org/p/257075958) | ORAI2, STIM1, TRPC3, and TPI1 identified as significant genetic markers. | Supervised ML classification algorithms, explainable AI techniques, Naive Bayes classifier. | Explainable AI with supervised ML classification algorithms for biomarker identification. |
| Metabolic brain disease (12 citations) |
| [Jemimah et al, 2023](https://www.semanticscholar.org/p/263911838) | SNPs in PRKCZ, PLCB1, ITPR2; expression of HLA-DQB1, EIF1AY, HLA-DQA1, ZFP57. | Deep learning classifier, constrained neural network, SHAP explanations. | c-Diadem model with KEGG pathway constraints for genetic marker identification. |
| BMC Medical Genomics (5 citations) |
| [Huang et al, 2021](https://www.semanticscholar.org/p/236199068) | Hundreds of new significant brain CpGs associated with AD. | Supervised machine learning strategy. | EWASplus, a supervised machine learning tool extending EWAS coverage to entire genome for AD analysis. |
| Nature Communications (39 citations) |
| [Zhu et al, 2024](https://www.semanticscholar.org/p/266935058) | 69 proteins with genetically predicted concentrations associated with AD risk. | Genetic prediction models for protein levels in plasma. | Genetic prediction models for plasma protein levels in AD analysis. |
| Alzheimer's Research & Therapy (9 citations) |

# Data integration and multimodal approaches

The integration of genetic data with other biomarkers represents a major advancement in AI-based Alzheimer's prediction, as researchers have found that multimodal approaches consistently outperform single-data-type models [11]. AI algorithms can now analyze massive quantities of data from numerous sources, including medical images, proteins in blood and cerebrospinal fluid, genetic information, clinical records, and even behavioral data [31].

One of the most successful combinations involves integrating neuroimaging with genetic data. Recent deep learning models like IGnet achieve 83.78% classification accuracy and 0.924 AUC-ROC by combining MRI scans with genetic sequencing data from chromosome 19 [16]. Studies using both brain MRI and genetic data from 543 patients show that genetic information better predicts disease progression while MRI data better reflects anatomical brain changes, with combined approaches outperforming either method alone [32], [33].

Advanced AI systems are now incorporating even more diverse data types for comprehensive risk assessment. Modern predictive algorithms can integrate brain imaging, genetic markers, blood biomarkers, cognitive test results, and even data from wearable technology that tracks heart rate, sleep patterns, and physical activity [34], [35], [36], [37]. Network-based models like BrainNetScore demonstrate the power of this approach by integrating brain connectivity networks with genetic associations, achieving AUC values of 0.684 for combined genetic and brain imaging data, improving to 0.778 when clinical covariates are included [17].

Epigenetic data integration has also shown remarkable results, with machine learning methods like EWASplus extending genome-wide coverage to predict hundreds of new brain methylation sites associated with Alzheimer's disease [26], [27]. The most advanced multimodal systems can achieve up to 99% accuracy by combining RNA sequencing data with brain MRI images using transformer-based models and computer vision algorithms  [20]. Future research directions emphasize the importance of developing interactive AI interfaces that allow doctors to query and adjust predictions from these integrated multimodal systems [38].

# Clinical application and future directions

[To be continue]

# References

[1] J. De Velasco Oriol, E. E. Vallejo, K. Estrada, J. G. Taméz Peña, and T. A. Disease Neuroimaging Initiative, “Benchmarking machine learning models for late-onset alzheimer’s disease prediction from genomic data,” *BMC Bioinformatics*, vol. 20, no. 1, p. 709, Dec. 2019, doi: 10.1186/s12859-019-3158-x.

[2] R. Mishra and B. Li, “The Application of Artificial Intelligence in the Genetic Study of Alzheimer’s Disease,” *Aging and disease*, vol. 11, no. 6, p. 1567, 2020, doi: 10.14336/AD.2020.0312.

[3] M. Osipowicz, B. Wilczynski, M. A. Machnicka, and for the Alzheimer’s Disease Neuroimaging Initiative, “Careful feature selection is key in classification of Alzheimer’s disease patients based on whole-genome sequencing data,” *NAR Genomics and Bioinformatics*, vol. 3, no. 3, p. lqab069, Jun. 2021, doi: 10.1093/nargab/lqab069.

[4] N. Briones and V. Dinu, “Data mining of high density genomic variant data for prediction of Alzheimer’s disease risk,” *BMC Med Genet*, vol. 13, no. 1, p. 7, Dec. 2012, doi: 10.1186/1471-2350-13-7.

[5] T.-T. Nguyen, J. Z. Huang, Q. Wu, T. T. Nguyen, and M. J. Li, “Genome-wide association data classification and SNPs selection using two-stage quality-based Random Forests,” *BMC Genomics*, vol. 16, no. S2, p. S5, Dec. 2015, doi: 10.1186/1471-2164-16-S2-S5.

[6] M. E. Stokes, M. Barmada, M. Kamboh, and S. Visweswaran, “The application of network label propagation to rank biomarkers in genome-wide Alzheimer’s data,” *BMC Genomics*, vol. 15, no. 1, p. 282, 2014, doi: 10.1186/1471-2164-15-282.

[7] T. Khater *et al.*, “Explainable Machine Learning Model for Alzheimer Detection Using Genetic Data: A Genome-Wide Association Study Approach,” *IEEE Access*, vol. 12, pp. 95091–95105, 2024, doi: 10.1109/ACCESS.2024.3410135.

[8] Computer and Information Technology Faculty, Sana’a University, Yemen, G. M. Fadhl Alqubati, G. H. Algaphari, and Computer and Information Technology Faculty, Sana’a University, Yemen, “Machine learning and deep learning-based approaches on various biomarkers for Alzheimer’s disease early detection: A review,” *IJSECS*, vol. 7, no. 2, pp. 26–43, Aug. 2021, doi: 10.15282/ijsecs.7.2.2021.4.0087.

[9] T. Jo, P. Bice, K. Nho, A. J. Saykin, and Alzheimer’s Disease Sequencing Project, “LD‐informed deep learning for Alzheimer’s gene loci detection using WGS data,” *A&D Transl Res & Clin Interv*, vol. 11, no. 1, p. e70041, Jan. 2025, doi: 10.1002/trc2.70041.

[10] T. Jo, E. H. Lee, and A. D. S. Project, “Uncertainty-Aware Genomic Classification of Alzheimer’s Disease: A Transformer-Based Ensemble Approach with Monte Carlo Dropout,” 2025, *arXiv*. doi: 10.48550/ARXIV.2506.00662.

[11] F. Ursin, C. Timmermann, and F. Steger, “Ethical Implications of Alzheimer’s Disease Prediction in Asymptomatic Individuals through Artificial Intelligence,” *Diagnostics*, vol. 11, no. 3, p. 440, Mar. 2021, doi: 10.3390/diagnostics11030440.

[12] Y. Ding *et al.*, “A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using18 F-FDG PET of the Brain,” *Radiology*, vol. 290, no. 2, pp. 456–464, Feb. 2019, doi: 10.1148/radiol.2018180958.

[13] Z. Liu *et al.*, “AD-GPT: Large Language Models in Alzheimer’s Disease,” 2025, *arXiv*. doi: 10.48550/ARXIV.2504.03071.

[14] A. S. Alatrany, A. J. Hussain, J. Mustafina, and D. Al-Jumeily, “Machine Learning Approaches and Applications in Genome Wide Association Study for Alzheimer’s Disease: A Systematic Review,” *IEEE Access*, vol. 10, pp. 62831–62847, 2022, doi: 10.1109/ACCESS.2022.3182543.

[15] R. O. Bahado-Singh *et al.*, “Artificial intelligence and leukocyte epigenomics: Evaluation and prediction of late-onset Alzheimer’s disease,” *PLoS ONE*, vol. 16, no. 3, p. e0248375, Mar. 2021, doi: 10.1371/journal.pone.0248375.

[16] J. X. Wang, Y. Li, X. Li, and Z.-H. Lu, “Alzheimer’s Disease Classification Through Imaging Genetic Data With IGnet,” *Front. Neurosci.*, vol. 16, p. 846638, Mar. 2022, doi: 10.3389/fnins.2022.846638.

[17] Y. Nam *et al.*, “BrainNetScore: Enhancing Alzheimer’s disease risk prediction using genetic‐guided brain volumetric phenotype network,” *Alzheimer’s &amp; Dementia*, vol. 20, no. S1, p. e084351, Dec. 2024, doi: 10.1002/alz.084351.

[18] Q. Zhang *et al.*, “Risk prediction of late-onset Alzheimer’s disease implies an oligogenic architecture,” *Nat Commun*, vol. 11, no. 1, p. 4799, Sep. 2020, doi: 10.1038/s41467-020-18534-1.

[19] K. Sekaran, A. M. Alsamman, C. George Priya Doss, and H. Zayed, “Bioinformatics investigation on blood-based gene expressions of Alzheimer’s disease revealed ORAI2 gene biomarker susceptibility: An explainable artificial intelligence-based approach,” *Metab Brain Dis*, vol. 38, no. 4, pp. 1297–1310, Apr. 2023, doi: 10.1007/s11011-023-01171-0.

[20] H. Anzum, N. S. Sammo, and S. Akhter, “Leveraging transformers and explainable AI for Alzheimer’s disease interpretability,” *PLoS One*, vol. 20, no. 5, p. e0322607, May 2025, doi: 10.1371/journal.pone.0322607.

[21] R. Gupta and Z. Iftekhar, “Artificial Intelligence for Alzheimer’s Disease Detection: Enhancing Biomarker Analysis and Diagnostic Precision,” in *ECSOC 2024*, MDPI, Nov. 2024, p. 25. doi: 10.3390/ecsoc-28-20206.

[22] C. Woods, X. Xing, S. Khanal, and A.-L. Lin, “Machine Learning-Driven Prediction of Brain Age for Alzheimer’s Risk: APOE4 Genotype and Gender Effects,” *Bioengineering*, vol. 11, no. 9, p. 943, Sep. 2024, doi: 10.3390/bioengineering11090943.

[23] S. Khan, K. H. Barve, and M. S. Kumar, “Recent Advancements in Pathogenesis, Diagnostics and Treatment of Alzheimer’s Disease,” *Curr Neuropharmacol*, vol. 18, no. 11, pp. 1106–1125, Nov. 2020, doi: 10.2174/1570159X18666200528142429.

[24] S. Jemimah, A. AlShehhi, and for the Alzheimer’s Disease Neuroimaging Initiative, “c-Diadem: a constrained dual-input deep learning model to identify novel biomarkers in Alzheimer’s disease,” *BMC Med Genomics*, vol. 16, no. S2, p. 244, Oct. 2023, doi: 10.1186/s12920-023-01675-9.

[25] H. Guo *et al.*, “Identification of Lipophagy-Related Gene Signature for Diagnosis and Risk Prediction of Alzheimer’s Disease,” *Biomedicines*, vol. 13, no. 2, p. 362, Feb. 2025, doi: 10.3390/biomedicines13020362.

[26] M. Vinciguerra, “The Potential for Artificial Intelligence Applied to Epigenetics,” *Mayo Clinic Proceedings: Digital Health*, vol. 1, no. 4, pp. 476–479, Dec. 2023, doi: 10.1016/j.mcpdig.2023.07.005.

[27] Y. Huang *et al.*, “A machine learning approach to brain epigenetic analysis reveals kinases associated with Alzheimer’s disease,” *Nat Commun*, vol. 12, no. 1, p. 4472, Jul. 2021, doi: 10.1038/s41467-021-24710-8.

[28] C. Bellenguez *et al.*, “New insights into the genetic etiology of Alzheimer’s disease and related dementias,” *Nat Genet*, vol. 54, no. 4, pp. 412–436, Apr. 2022, doi: 10.1038/s41588-022-01024-z.

[29] J. Zhu *et al.*, “Associations between genetically predicted plasma protein levels and Alzheimer’s disease risk: a study using genetic prediction models,” *Alz Res Therapy*, vol. 16, no. 1, p. 8, Jan. 2024, doi: 10.1186/s13195-023-01378-4.

[30] S. L. Boschen, A. A. Mukerjee, A. H. Faroqi, B. E. Rabichow, and J. Fryer, “Research models to study lewy body dementia,” *Mol Neurodegeneration*, vol. 20, no. 1, p. 46, Apr. 2025, doi: 10.1186/s13024-025-00837-w.

[31] S. Khastehband *et al.*, “The Use of Artificial Intelligence in the Management of Neurodegenerative Disorders; Focus on Alzheimer’s Disease:,” *GMJ*, vol. 12, Sep. 2023, doi: 10.31661/gmj.v12i.3061.

[32] S. Mirkin and B. C. Albensi, “Should artificial intelligence be used in conjunction with Neuroimaging in the diagnosis of Alzheimer’s disease?,” *Front. Aging Neurosci.*, vol. 15, p. 1094233, Apr. 2023, doi: 10.3389/fnagi.2023.1094233.

[33] the Alzheimer’s Disease Neuroimaging Initiative *et al.*, “Machine Learning Based Multimodal Neuroimaging Genomics Dementia Score for Predicting Future Conversion to Alzheimer’s Disease,” *JAD*, vol. 87, no. 3, pp. 1345–1365, May 2022, doi: 10.3233/JAD-220021.

[34] E. El Abiad, A. Al-Kuwari, U. Al-Aani, Y. Al Jaidah, and A. Chaari, “Navigating the Alzheimer’s Biomarker Landscape: A Comprehensive Analysis of Fluid-Based Diagnostics,” *Cells*, vol. 13, no. 22, p. 1901, Nov. 2024, doi: 10.3390/cells13221901.

[35] H. J. Tricás-Vidal, M. O. Lucha-López, C. Hidalgo-García, M. C. Vidal-Peracho, S. Monti-Ballano, and J. M. Tricás-Moreno, “Health Habits and Wearable Activity Tracker Devices: Analytical Cross-Sectional Study,” *Sensors*, vol. 22, no. 8, p. 2960, Apr. 2022, doi: 10.3390/s22082960.

[36] J. A. Yoon *et al.*, “Correlation between cerebral hemodynamic functional near-infrared spectroscopy and positron emission tomography for assessing mild cognitive impairment and Alzheimer’s disease: An exploratory study,” *PLoS ONE*, vol. 18, no. 8, p. e0285013, Aug. 2023, doi: 10.1371/journal.pone.0285013.

[37] I. Malik, A. Iqbal, Y. H. Gu, and M. A. Al-antari, “Deep Learning for Alzheimer’s Disease Prediction: A Comprehensive Review,” *Diagnostics*, vol. 14, no. 12, p. 1281, Jun. 2024, doi: 10.3390/diagnostics14121281.

[38] M. L. Raza *et al.*, “Advancements in deep learning for early diagnosis of Alzheimer’s disease using multimodal neuroimaging: challenges and future directions,” *Front. Neuroinform.*, vol. 19, p. 1557177, May 2025, doi: 10.3389/fninf.2025.1557177.