AI genetic 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]

[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.