# Review Article

## Artificial Intelligence-Driven Biomarker Discovery in Huntington's Disease: Integrating Multi-Omics and Multimodal Data

**Rekha B[1], Jeyavalli A[2], Nithesh Kumar Haridoss[3]**

*From, [1]Associate Professor, Department of Pharmacology, Thanthai Roever College of Pharmacy, Perambalur. [2]Associate Professor, Department of Pharmacology, Dhanalakshmi Srinivasan College of Pharmacy, Perambalur. [3]Assistant Professor, Department of Pharmacy practice, KG College of Pharmacy and Research Institute, Villupuram, Tamilnadu.*

## ABSTRACT

Huntington's disease (HD) is an autosomal dominant neurodegenerative disorder caused by an expanded CAG repeat in the HTT gene. Although the causal mutation is established, disease onset and progression vary among individuals, creating a need for biomarkers that capture molecular injury, neurodegeneration, disease stage, and treatment response. Artificial intelligence (AI), particularly machine learning, can integrate high-dimensional clinical, neuroimaging, electrophysiological, and molecular datasets to identify multidimensional biomarker signatures. This review examines primary research on AI-assisted biomarker discovery in HD, with emphasis on the integration of genomics, transcriptomics, proteomics, metabolomics, epigenomics, neuroimaging, EEG, gait, speech, and circulating biofluid markers. Structural and functional MRI provide established measures of disease-related brain changes, while machine-learning approaches can combine distributed imaging features into composite biomarkers. Biofluid studies support neurofilament light (NfL), mutant huntingtin, microRNAs, proteins, and metabolites as complementary biomarker candidates. Transcriptomic, epigenomic, and multi-omic studies further identify molecular and cell-specific alterations that may improve biological characterization of disease progression. AI-based analysis of EEG, speech, gait, and wearable-sensor data offers additional opportunities for objective and potentially repeated disease monitoring. However, translation remains limited by small sample sizes, cohort heterogeneity, high-dimensional datasets, overfitting, limited external validation, and challenges in interpreting model-derived features biologically.

***Key words:*** Huntington's disease, multi-omics, artificial intelligence, biomarkers, neuroimaging.

Huntington's disease (HD) is a dominantly inherited neurodegenerative disorder caused by expansion of a CAG repeat in the HTT gene. The mutation establishes genetic risk, but the clinical phenotype emerges through interacting molecular, cellular, network, and systemic processes. Motor impairment is accompanied by cognitive and psychiatric changes, and measurable abnormalities can precede overt motor diagnosis. Large longitudinal studies such as TRACK-HD established that quantitative imaging, cognitive, motor, and neuropsychiatric measures can detect abnormalities during the premanifest phase, providing a foundation for biomarker development [1,2]. The biomarker problem in HD is therefore broader than identifying a diagnostic signal. A clinically useful biomarker may need to identify disease-related change before motor diagnosis, reflect biological progression, predict subsequent clinical decline, or provide a pharmacodynamic measure of target engagement. Structural MRI has been particularly informative: longitudinal studies demonstrate progressive whole-brain and regional atrophy, with caudate and putamen measures showing substantial sensitivity to disease burden and progression [3-5]. Other modalities, including magnetic resonance spectroscopy, diffusion imaging, EEG, and biofluid assays, provide complementary information [6-9]. AI is relevant because these datasets are high-dimensional and often contain correlated features distributed across modalities. Conventional univariate analyses can identify individual associations, but multivariate machine-learning methods can combine weak signals into a composite predictor. Early HD imaging studies demonstrated that support vector machines and related multivariate approaches could classify premanifest carriers and controls and estimate disease-related variation from structural, diffusion, and functional MRI features [10].

**Access this article online**

Quick Response Code

Received – 23th November 2025

Initial Review – 25th November 2025

Accepted – 20th December 2025

**Correspondence to:** Dr. Rekha B, Associate Professor Department of Pharmacology, Dhanalakshmi Srinivasan College of Pharmacy, Perambalur – 621212, Tamil Nadu, India

**Email:** jdbrekha@gmail.com

Subsequent work extended this approach to multimodal imaging polymarkers and large longitudinal registries [11,12].This review synthesizes primary research on AI-assisted biomarker discovery in HD, emphasizing how molecular and multimodal data can be combined rather than treating each biomarker class in isolation.

## 2. NEUROIMAGING AS A FOUNDATION FOR AI-ASSISTED BIOMARKER DISCOVERY

Neuroimaging provides one of the most mature quantitative biomarker domains in HD. TRACK-HD demonstrated measurable regional and global brain changes in premanifest carriers and early HD, including caudate and whole-brain atrophy, alongside cognitive and motor changes [1-3]. Longitudinal multimodal MRI from IMAGE-HD similarly showed progressive caudate and putaminal abnormalities and associations between imaging change and clinical decline [4]. Across pooled datasets, robust anatomical markers have remained useful despite differences between multicenter cohorts, supporting the value of standardized image processing and region-based measures [5].

AI adds a multivariate layer to these established imaging markers. In an early machine-learning study, structural, diffusion-weighted, and functional MRI features were used to classify 39 premanifest gene carriers and 25 controls, with the best models achieving classification accuracy of up to 76% [10]. The important contribution was not the accuracy estimate alone but the demonstration that distributed imaging patterns contain information about disease state that is not captured by a single anatomical measurement.A later study used structural and resting-state MRI in a multivariate framework to develop an imaging polymarker for clinical diagnosis. Combining measures performed better than individual imaging variables, and the model also showed potential for identifying carriers who were closer to clinical diagnosis than expected from conventional CAG-age estimates [11]. These findings support the concept that AI can integrate complementary anatomical and functional abnormalities into a composite biomarker.

AI-based imaging models should nevertheless be interpreted cautiously. Small datasets can produce optimistic estimates if feature selection and model evaluation are not strictly separated. External validation is particularly important when imaging protocols, scanners, preprocessing pipelines, and participant characteristics differ between sites. Large pooled studies indicate that caudate and putamen volumes remain among the most reproducible progression markers, suggesting that AI models may be most useful when they add information beyond these established measures rather than simply reproducing them [5,13].

## 3. BIOFLUID BIOMARKERS AND MOLECULAR LAYERS

Biofluid biomarkers offer a complementary route because blood and cerebrospinal fluid (CSF) can be sampled repeatedly and may provide molecular information that is not directly accessible through imaging. Mutant huntingtin (mHTT) can be quantified in CSF using ultrasensitive assays, and mHTT is detectable in mutation carriers but not controls in early studies [14]. Subsequent work showed that mHTT and neurofilament light (NfL) have distinct longitudinal behavior. NfL in plasma and CSF correlated more closely with clinical and imaging measures, whereas CSF mHTT retained potential value as a pharmacodynamic marker for huntingtin-lowering interventions [15]. Plasma NfL has emerged as one of the more extensively investigated fluid markers. In a large retrospective cohort, plasma NfL was higher in mutation carriers than controls, increased across disease stages, correlated with clinical and MRI measures, and was associated with subsequent clinical onset in premanifest carriers [16]. Longitudinal studies have further characterized NfL trajectories across the HD spectrum, supporting its use as a marker of neurodegenerative burden while emphasizing that concentration and longitudinal change may provide different types of information [15,17].

Other molecular classes provide additional dimensions for multimodal models. Plasma microRNA profiling identified miR-34b as elevated in premanifest HD, suggesting that circulating regulatory RNAs may be useful as minimally invasive biomarkers [18]. CSF profiling later identified a set of microRNAs that increased across prodromal risk categories, providing evidence that RNA-based signals can be detected before clinical diagnosis [19]. Proteomic studies have identified inflammatory and neuronal proteins in plasma and CSF, while targeted work on proenkephalin showed a relationship with symptom severity and supported its investigation as a marker of medium spiny neuron integrity [20,21]. Metabolomics adds another layer. Plasma studies have identified changes involving amino acids, carnitine, phosphatidylcholine species, and other metabolic pathways [22]. CSF metabolomics has identified alterations in tyrosine, phenylalanine, purine, and related pathways that associate with disease severity or risk of onset [23]. Longitudinal and cross-sectional metabolic investigations indicate that peripheral and central metabolic abnormalities are heterogeneous, making multivariate integration attractive but also requiring careful control of nutritional, endocrine, medication, and disease-stage effects [24,25].

## 4. TRANSCRIPTOMICS, EPIGENOMICS, AND MULTI-OMICS INTEGRATION

Transcriptomic studies demonstrate that HD is associated with coordinated changes in gene expression across vulnerable brain regions. A comparative analysis of patient brain regions identified a common gene-expression signature, supporting the

existence of shared molecular responses across affected tissues [26]. Single-nucleus RNA sequencing has further resolved cell-type-specific responses. In human HD cortex, distinct astrocyte states were identified alongside changes in neurons, oligodendrocytes, and microglia [27]. More recent single-nucleus work in an ovine HD model identified a medium spiny neuron glutamate-excitotoxicity signature before detectable neuronal loss, illustrating how single-cell approaches can expose early cellular states that bulk measurements may obscure [28].

Epigenomic studies provide another layer of regulatory information. Human and animal studies have identified disease-associated DNA methylation patterns, including associations with HD mutation status and loci related to motor progression [29]. These findings are relevant to AI because epigenetic measurements can contribute thousands of correlated features, creating a setting in which regularization, dimensionality reduction, and biologically informed feature selection are essential. Integrated multi-omics studies demonstrate the value of combining molecular layers. In HD, joint analysis of transcriptomic and metabolic data has been used to identify pathways associated with protein aggregation and cellular metabolism [30]. A recent multi-omic analysis combining single-nucleus RNA sequencing, bulk RNA sequencing, lipidomics, CAG repeat measurements, and multiplexed immunofluorescence identified a compensatory astrocyte state in HD brain tissue, illustrating how different molecular layers can converge on cell-type-specific disease mechanisms [31].

The role of AI in this setting is principally integrative. Genomic information can encode inherited risk and CAG-repeat burden; transcriptomics can capture altered cellular programs; proteomics can reflect protein abundance and processing; metabolomics can capture downstream biochemical consequences; and epigenomics can represent regulatory state. Combining these layers may produce biomarkers that are more biologically informative than any single feature. However, the availability of multiple omics layers from the same individuals is limited, and cross-platform normalization remains a major methodological challenge.

## 5. AI BEYOND IMAGING: EEG, GAIT, SPEECH, AND WEARABLE DATA

AI-assisted biomarker discovery is increasingly extending beyond conventional imaging and molecular assays. EEG provides a relatively accessible measure of neural activity. A machine-learning study demonstrated that EEG features could distinguish HD from controls, supporting the feasibility of automated classification [32]. Such approaches may be particularly useful when combined with clinical or molecular measures, although EEG findings are not disease-specific and can be influenced by recording conditions and medication. Digital motor phenotyping provides another source of continuous quantitative data. Inertial-sensor studies have shown that gait variability measures correlate with clinical disease severity and functional capacity [33]. Deep-learning analysis of footstep pressure data has also been used to classify HD severity, illustrating how raw sensor signals can be transformed into objective digital biomarkers [34]. Wearable-sensor models have demonstrated the ability to classify HD status and estimate selected Unified Huntington's Disease Rating Scale motor subscores [35].

Speech is particularly attractive because recordings can be obtained remotely and repeatedly. Machine-learning models trained on speech features have predicted cognitive, motor, and functional scores, with speech features also correlating with striatal atrophy [36]. These studies suggest that digital speech measures may provide a scalable complement to clinic-based assessments. Newer models are extending this concept to automated severity indices and longitudinal prediction, but these approaches require prospective validation before they can be considered established clinical biomarkers [37].

## 6. MACHINE LEARNING FOR DISEASE PROGRESSION AND MULTIMODAL PREDICTION

Prediction of disease progression is a central application of AI in HD because conventional CAG-age models explain population-level variation but do not fully capture individual trajectories. Machine-learning models trained on large observational datasets can incorporate motor, cognitive, functional, imaging, and demographic variables simultaneously. A large progression model using data from PREDICT-HD, REGISTRY, TRACK-HD, and Enroll-HD identified nine latent disease states from 44 clinical measures and validated the state structure in an independent Enroll-HD dataset [38].

This type of model illustrates a key advantage of AI: the outcome does not have to be a single binary diagnosis. Disease can be represented as a continuous or staged trajectory, allowing different biomarkers to contribute at different phases. Imaging markers may be particularly informative for structural progression, NfL for neurodegenerative burden, mHTT for target engagement, and digital measures such as speech or gait for frequent functional monitoring. The strongest future models may therefore be multimodal rather than dependent on one biomarker class. Model performance must be interpreted in relation to the study design.

A model that performs well within a highly selected research cohort may not generalize to community populations, different ethnic groups, different scanners, or different recording environments. Prospective validation, external test sets, transparent reporting of missing data, prespecified feature selection, and calibration assessment should become standard requirements for AI-derived HD biomarkers.

## CONCLUSION

Primary research across imaging, biofluids, molecular profiling, and digital phenotyping supports a role for AI in extracting multidimensional biomarker signals from Huntington's disease. Structural MRI remains a strong foundation, while NfL, mHTT, microRNAs, proteins, metabolites, transcriptomic signatures, epigenetic features, EEG, gait, and speech provide complementary information. Multi-omics integration offers a way to connect upstream genetic and regulatory features with downstream molecular and clinical phenotypes. The next stage of the field should move from exploratory model development toward reproducible, externally validated, longitudinal multimodal biomarkers that can be tested prospectively in therapeutic studies.

## REFERENCES


1. Tabrizi SJ, Scahill RI, Durr A, Roos RAC, Leavitt BR, Jones R, et al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. Lancet Neurol. 2009;8(9):791-801. doi:10.1016/S1474-4422(09)70170-X.
2. Tabrizi SJ, Scahill RI, Owen G, Durr A, Leavitt BR, Roos RAC, et al. Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. Lancet Neurol. 2013;12(7):637-649. doi:10.1016/S1474-4422(13)70088-7.
3. Tabrizi SJ, Langbehn DR, Leavitt BR, Roos RAC, Durr A, Craufurd D, et al. Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal analysis. Lancet Neurol. 2011;10(1):31-42. doi:10.1016/S1474-4422(10)70276-3.
4. Georgiou-Karistianis N, Scahill R, Tabrizi SJ, Aylward E, Long JD, Stout JC, et al. Multimodal imaging biomarkers in premanifest and early Huntington's disease: 30-month IMAGE-HD data. Br J Psychiatry. 2016;208(1):46-56. doi:10.1192/bjp.bp.114.153536.
5. Wijeratne PA, Johnson EB, Eshaghi A, Aksman L, Gregory S, Johnson HJ, et al. Robust markers and sample sizes for multicenter trials of Huntington disease. Ann Neurol. 2020;87(5):751-762. doi:10.1002/ana.25709.
6. Sturrock A, Laule C, Decolongon J, Dar Santos R, Coleman AJ, Creighton S, et al. A longitudinal study of magnetic resonance spectroscopy Huntington's disease biomarkers. Mov Disord. 2015;30(3):393-401. doi:10.1002/mds.26118.
7. Estevez-Fraga C, Scahill R, Rees G, Tabrizi SJ, Gregory S. Diffusion imaging in Huntington's disease: comprehensive review. J Neurol Neurosurg Psychiatry. 2021;92(1):62-69. doi:10.1136/jnnp-2020-324377.
8. Odish OFF, Johnsen K, van Someren P, Roos RAC, van Dijk JG. EEG may serve as a biomarker in Huntington's disease using machine learning automatic classification. Sci Rep. 2018;8(1):16090. doi:10.1038/s41598-018-34269-y.
9. Ponomareva N, Klyushnikov S, Abramycheva N, et al. Alpha-theta border EEG abnormalities in preclinical Huntington's disease. J Neurol Sci. 2014;344(1-2):114-120. doi:10.1016/j.jns.2014.06.035.
10. Rizk-Jackson A, Stoffers D, Sheldon S, et al. Evaluating imaging biomarkers for neurodegeneration in pre-symptomatic Huntington's disease using machine learning techniques. Neuroimage. 2011;56(2):788-796. doi:10.1016/j.neuroimage.2010.04.273.
11. Mason SL, Daws RE, Soreq E, Johnson EB, Scahill RI, Tabrizi SJ, et al. Predicting clinical diagnosis in Huntington's disease: an imaging polymarker. Ann Neurol. 2018;83(3):532-543. doi:10.1002/ana.25171.
12. Mohan A, Sun Z, Ghosh S, Li Y, Sathe S, Hu J, et al. A machine-learning derived Huntington's disease progression model: insights for clinical trial design. Mov Disord. 2022;37(3):553-562. doi:10.1002/mds.28866.
13. Hyare H, Langbehn DR, Jones R, Leavitt BR, Roos RA, Durr A, et al. Clinical impairment in premanifest and early Huntington's disease is associated with regionally specific atrophy. Hum Brain Mapp. 2013 Mar;34(3):519-29. doi: 10.1002/hbm.21449.
14. Wild EJ, Boggio R, Langbehn D, et al. Quantification of mutant huntingtin protein in cerebrospinal fluid from Huntington's disease patients. J Clin Invest. 2015;125(5):1979-1986. doi:10.1172/JCI80743.
15. Byrne LM, Rodrigues FB, Johnson EB, et al. Mutant huntingtin and neurofilament light have distinct longitudinal dynamics in Huntington's disease. Sci Transl Med. 2020 Dec 16;12(574):eabc2888. doi: 10.1126/scitranslmed.abc2888.
16. Byrne LM, Rodrigues FB, Johnson EB, et al. Neurofilament light protein in blood as a potential biomarker of neurodegeneration in Huntington's disease: a retrospective cohort analysis. Lancet Neurol. 2017;16(8):601-609. doi:10.1016/S1474-4422(17)30124-2.
17. Parkin GM, Thomas EA, Corey-Bloom J. Mapping neurodegeneration across the Huntington's disease spectrum: a five-year longitudinal analysis of plasma neurofilament light. EBioMedicine. 2024 Jun;104:105173. doi: 10.1016/j.ebiom.2024.105173.
18. Gaughwin PM, Ciesla M, Lahiri N, Tabrizi SJ, Brundin P, Björkqvist M. Hsa-miR-34b is a plasma-stable microRNA that is elevated in pre-manifest Huntington's disease. Hum Mol Genet. 2011;20(8):1621-1631. doi:10.1093/hmg/ddr111.
19. Reed ER, Latourelle JC, Bockholt JH, Bregu J, Smock J, Paulsen JS, et al. MicroRNAs in CSF as prodromal biomarkers for Huntington disease in the PREDICT-HD study. Neurology. 2018;90(4):e264-e272. doi:10.1212/WNL.0000000000004844.
20. Dalrymple A, Wild EJ, Joubert R, et al. Proteomic profiling of plasma in Huntington's disease reveals neuroinflammatory activation and biomarker candidates. J Proteome Res. 2007;6(11):4579-4589. doi:10.1021/pr0700753.
21. Niemela V, Landtblom AM, Nyholm D, Kneider M, Constantinescu R, Paucar M, et al. Proenkephalin Decreases in Cerebrospinal Fluid with Symptom Progression of Huntington's Disease. Mov Disord. 2021 Feb;36(2):481-491. doi: 10.1002/mds.28391.
22. Cheng ML, Chang KH, Wu YR, Chen CM. Metabolic disturbances in plasma as biomarkers for Huntington's disease. J Nutr Biochem. 2016 May;31:38-44. doi: 10.1016/j.jnutbio.2015.12.001.
23. Patassini S, Begley P, Xu J, et al. Alterations in the tyrosine and phenylalanine pathways revealed by biochemical profiling in cerebrospinal fluid of Huntington's disease subjects. Sci Rep. 2019;9:16597. doi:10.1038/s41598-019-52928-0.
24. McGarry A, Gaughan J, Hackmyer C, et al. Cross-sectional analysis of plasma and CSF metabolomic markers in

Huntington's disease for participants of varying functional disability: a pilot study. Sci Rep. 2020;10(1):20490. doi:10.1038/s41598-020-77526-9.
25. Nambron R, Silajdžić E, Kalliolia E, et al. A metabolic study of Huntington's disease. PLoS One. 2016;11(1):e0146480. doi:10.1371/journal.pone.0146480.
26. Neueder A, Bates GP. A common gene expression signature in Huntington's disease patient brain regions. BMC Med Genomics. 2014;7:60. doi:10.1186/s12920-014-0060-2.
27. Al-Dalahmah O, Sosunov AA, Shaik A, Ofori K, Liu Y, Vonsattel JP, Adorjan I, Menon V, Goldman JE. Single-nucleus RNA-seq identifies Huntington disease astrocyte states. Acta Neuropathol Commun. 2020 Feb 18;8(1):19. doi: 10.1186/s40478-020-0880-6.
28. Jiang A, You L, Handley RR, et al. Single nuclei RNA-seq reveals a medium spiny neuron glutamate excitotoxicity signature prior to the onset of neuronal death in an ovine Huntington's disease model. Hum Mol Genet. 2024;33(17):1524-1539. doi:10.1093/hmg/ddae087.
29. Lu AT, Narayan P, Grant MJ, Langfelder P, Wang N, Kwak S, Wilkinson H, et al. DNA methylation study of Huntington's disease and motor progression in patients and in animal models. Nat Commun. 2020 Sep 10;11(1):4529. doi: 10.1038/s41467-020-18255-5.
30. Pradhan SS, Thota SM, Rajaratnam S, et al. Integrated multi-omics analysis of Huntington disease identifies pathways that modulate protein aggregation. Dis Model Mech. 2022;15(10):dmm049492. doi:10.1242/dmm.049492.
31. Paryani F, Kwon JS, Ng CW, et al. Multi-omic analysis of Huntington's disease reveals a compensatory astrocyte state. Nat Commun. 2024;15(1):6742. doi:10.1038/s41467-024-50626-0.
32. Zhang S, Poon SK, Vuong K, Sneddon A, Loy CT. A deep learning-based approach for gait analysis in Huntington disease. Stud Health Technol Inform. 2019;264:477-481. doi:10.3233/SHTI190267.
33. Gaßner H, Jensen D, Marxreiter F, Kletsch A, Bohlen S, Schubert R, et al. Gait variability as digital biomarker of disease severity in Huntington's disease. J Neurol. 2020 Jun;267(6):1594-1601. doi: 10.1007/s00415-020-09725-3.
34. Zhang S, Poon SK, Vuong K, Sneddon A, Loy CT. A deep learning-based approach for gait analysis in Huntington disease. Stud Health Technol Inform. 2019;264:477-481. doi:10.3233/SHTI190267.
35. Scheid BH, Aradi S, Pierson RM, Baldassano S, Tivon I, Litt B, et al. Predicting Severity of Huntington's Disease With Wearable Sensors. Front Digit Health. 2022 Apr 4;4:874208. doi: 10.3389/fdgth.2022.874208.
36. Riad R, Lunven M, Titeux H, et al. Predicting clinical scores in Huntington's disease: a lightweight speech test. J Neurol. 2022;269(9):5008-5021. doi:10.1007/s00415-022-11148-1.
37. Shankar, R., Goh, Z., Devi, F. et al. A systematic review of explainable artificial intelligence methods for speech-based cognitive decline detection. npj Digit. Med. 2025;8:724. doi: 10.1038/s41746-025-02105-z.
38. Ghosh S, Sun Z, Li Y, Cheng Y, Mohan A, Sampaio C, et al. An exploration of latent structure in observational Huntington's disease studies. AMIA Jt Summits Transl Sci Proc. 2017;2017:92-102.

**How to cite this article:** Rekha B, Jeyavalli A**,** Haridoss N K. Artificial Intelligence-Driven Biomarker Discovery in Huntington's Disease: Integrating Multi-Omics and Multimodal Data. Indian J Pharm Drug Studies. 2025; 4(4):141-145.

*Funding: None* *Conflicts of Interest: None Stated*