GMR
The Original

Quantum-Integrated Deep Learning Framework for Large-Scale Gene Expression Analysis and Predictive Modeling of Parkinson's and Alzheimer's disease

Dr. A.Mary Subashini[1] , Ali Bostani[2] , Muninathan N[3] , Dr. Trupti Kaushiram Wable[4] , M Aruna[5] , Dr. T. Nagalakshmi[6] , Dr. Manpreet Kaur[7]

[1]Assistant Professor, Department of Artificial Intelligence & Machine Learning, Idhaya college for women, Kumbakonam, Email: subasinghphd2020@gmail.com ORCHID ID: 0009-0005-0755-6583
[2]Associate Professor, College of Engineering and Applied Sciences, American University of Kuwait, Salmiya, Kuwait. Email: abostani@auk.edu.kw 0000-0002-7922-9857
[3]Scientist, Central Research Laboratory, Meenakshi Medical College Hospital & Research Institute, Meenakshi Academy of Higher Education and Research, Chennai, Tamilnadu, India. Email: muninathan@maher.ac.in
[4]Assistant Professor, Department of Electronics and Computer Engineering,Sir Visvesvaraya Institute of Technology, Nashik – 422102, Maharashtra, India. Email: wabletrupti@gmail.com Orcid: 0009-0000-1995-5235
[5]Assistant Professor, Dayananda Sagar Academy of Technology and Management Bangalore 560082, India. Email: arunasrini2005@gmail.com Orcid id 0000-0002-7427-1888
[6]Associate Professor, Department of Mathematics, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Avadi, Chennai-600062, Tamilnadu, India.Email: drnagalakshmit@veltech.edu.in Orcid id: 0000-0001-8212-8283
[7]Assistant Professor, Faculty of Computing, Guru Kashi University, Talwandi Sabo (BTI), PB, India. Email: apmanpreetkaur@gmail.com Orcid id: 0000-0003-1027-9239

Abstract

Parkinson disease (PD) and Alzheimer disease (AD) are progressive neurodegenerative diseases that are marked by complicated molecular changes and shared pathogenesis. Due to its significant role in the appearance and further evolution of disease, transcriptomic deregulation is quite difficult to measure; the key issue is finding solid molecular signatures among hundreds of gene activity profiles. The objectives of the current research were to isolate differentially expressed genes (DEGs) that are related to PD and AD, as well as to determine their applicability in the classification of the disease with the help of integrative computational methods. Available transcriptomic data in the general population were subjected to the analysis in order to identify important DEGs at the level of statistical providers such as false discovery rate (FDR) adjusted values. The functional enrichment analysis, which consists of Gene ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis was conducted to determine the biological processes and signalling pathways involved in neurodegeneration. Predictive modelling methods, such as L1/L2-regularised Logistic Regression, Random Forest, XGBoost, Support Vector Machine radial basis function kernel and a hybrid quantum-deep learning model, were then used to analyse the identified gene signatures. Notable DEGs were highly enriched with pathways, which dealt with neuroinflammation, synaptic transmission, mitochondrial dysfunction, and dopaminergic signalling. Analysis of comparative classification revealed that models had strong predictive performances, with the integrative hybrid structure having a better discriminative capacity in the form of better accuracy and area under ROC curve (AUC) than baseline strategies. These results should put into the limelight important transcriptomic phenotypes underlying PD and AD and show how integrative modelling schemes could further improve molecular-based disease prediction. The readings of the identified biomarkers and enriched

pathways can help in better early diagnosis and give ideas on the targeted therapeutic strategy in neurodegenerative disorders.

Keywords: Parkinson's disease; Alzheimer's disease; Gene expression; Differential expression analysis; Functional enrichment; Biomarkers; Predictive modeling; Neurodegeneration.

1. INTRODUCTION

Among the most urgent neurological issues of the twenty-first century, neurodegenerative disorders, especially, the Alzheimer's disease (AD) and the Parkinson's disease (PD), should be mentioned. AD is the predominant cause of dementia in the world and PD ranks second as the most prevalent neurodegenerative disease that affects millions of people across the globe. As the ageing populations continue to grow at a high rate, pre-empting rates, the prevalence and incidence of both the disorders are expected to increase significantly, which are increasing the healthcare expenses, care giving burden and the socioeconomic consequences. In addition to mortality, the diseases result in cognitive and motor decline that deteriorates with time, grossly affecting the quality of life and independence in the long term (Vilor-Tejedor et al., 2025; Wang et al., 2025). Despite AD and PD having different clinical manifestations, there is an accruing body of evidence based on the molecular basis to indicate that there is some overlap between the pathogenic mechanisms of the two disorders. The extent of abnormal protein aggregation, mitochondrial dysfunction, oxidative stress, synaptic degeneration, and chronic neuroinflammation are both features of the conditions. Genomic and transcriptomic analyses conducted on large scales have found abnormalities of the activation of immune, apoptotic, and cellular stress responses through pathways in and around the impacted brain regions (Zhang et al., 2025; Potemkin et al., 2025). These common molecular markers emphasize the need to investigate such systems on a systems level instead of an individual gene of interest.

The application of high-throughput RNA sequencing technologies to undertake transcriptomic profiling has revolutionised neurodegeneration research as it can take measurements of gene pattern of expression of thousands of genes at a time. Bulk and single-cell analysis of the neuronal and glial changes has become possible via bulk and single-cell RNA-seq methods, providing an in depth understanding of heterogeneity and progression of diseases in a way like never before (Huang et al., 2025; Vilor-Tejedor et al., 2025). This is because such data may be used to identify diagnostic biomarkers, reveal regulatory networks, and identify potential therapeutic targets. In spite of the mass dimensions of transcriptomic data, it is exceedingly difficult to interpret high-dimensional data of gene expression. Neurodegenerative datasets tend to have tens of thousands of features and relatively small sample sizes which causes the risk of overfitting and false discoveries. There are additional complications in the name of biological variability, batch effects and noise that make reliable inference difficult. The traditional statistical methods often do not focus on nonlinear interaction regulations of genes and their higher-order regulatory patterns that are combined into more complex biological systems (Ren et al., 2025; Wang et al., 2025).

Machine learning (ML) and artificial intelligence (AI) approaches have been embraced to overcome such shortcomings when it comes to transcriptomic analysis. The models of supervised learning have proven to be more accurate in disease classification and also the identification of biomarkers by eliciting unique gene signatures on massive datasets (Ahammad et al., 2024; Peng et al., 2024). In addition, explainable AI methods allow making sense of predictive characteristics, which offers biological plausibility and mechanistic significance instead of black-box results (Usman et al., 2025). Recent developments have gone through improvements in single-omics analysis to include integrative multi-omics approaches that include transcriptomics with genomics, epigenomics, proteomics, and clinical phenotype. The above integrative approaches are more effective at identifying convergent molecular pathways and therapeutic targets that are common between AD and PD (Alharbi et al., 2025; Li et al., 2026). Sections of the pipeline ML guided by systems biology and individual biomarker discovery platforms are also evidence of the strength of computational modeling with biological domain expertise (Mottaqi et al., 2025).

New computational models such as quantum-like regression networks, transcriptomics-neuroimaging bimetabolism, suggest the transition to more complex analytical processes capable of describing complex interaction effects of diseases (Konar et al., 2025; Huang et al., 2025). Such strategies render the need to have computational frameworks that are scalable, interpretable and that can be generalized to be able to handle heterogeneous and high-dimensional biological data and still maintain clinical relevance. Figure 1 demonstrates the conceptualization workflow of integrating transcriptomic data and the AI-based biomarker discovery in neurodegenerative diseases (Ren et al., 2025). Considering these facts and chances, the current study will build a unified computational framework of a high-dimensional transcriptomic profiling in Alzheimer and Parkinson disease. The given strategy is based on strong feature selection, predictive analysis, and biological explanability to detect both general and disease-specific molecular signatures. The intention of crossing transcriptomics with the more sophisticated AI techniques is that this work can improve diagnosis stratification, aid the precision medicine plan, and help to obtain actionable therapeutic targets in neurodegenerative studies.

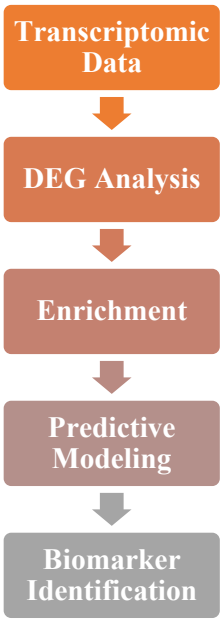

Figure 1. Integrative Study Workflow for Transcriptomics-Based Biomarker Discovery in Alzheimer's and Parkinson's disease

2. Previous Genetic and Transcriptomics Studies

2.1 Transcriptomics Alterations in Parkinson's disease

The transcriptomic analysis of PD has concurred that variousially expressed genes (DEGs) are linked to degeneration of dopaminergic neurons, mitochondrial dysfunction, and neuroinflammation. Massive transcriptomic profiling has discovered that there is dysregulation of genes that are linked to the synthesis of dopamine, its vesicular delivery, and synaptic signalling in the substantia nigra (Wang et al., 2025). The multimethod transcriptomic changes also identify the changes in the activity of oxidative phosphorylation pathways and mitochondrial complex, which further prove the impact of metabolic stress as a factor in PD development (Li et al., 2025). Besides dopaminergic pathways disturbances, inflammatory signalling has proven to be a decisive cause of PD pathogenesis. Upregulation of cytokine mediated pathways, microglial

activation markers and immune related transcription factors have been shown thus, indicating persistent neuroimmune activation by gene expression (Peng et al., 2024; Wang et al., 2025). These results would follow a paradigm where chronic neuroinflammation increases the vulnerability of the neurons and also fastens the rate of appearance of the disease. Although these progress have been reported, not all of the identified DEGs are continually verified across datasets as a result of cohort and analysis strategy heterogeneity.

2.2 Gene Expression Dysregulation in Alzheimer's disease

In Alzheimer's disease (AD), transcriptomic studies have shown that there is widespread deregulation in gene expression that relates to synaptic malfunction and cognitive impairment. It was repeatedly noted that the affected cortical and hippocampal areas downregulate the synaptic transmission genes, plasticity regulators and neuronal signalling molecules (Zhang et al., 2025). At the same time, there is an abnormal expression pattern of genes related to amyloid precursor protein (APP) processing and tau-related cytoskeleton processes, which are indicative of pathophysiological processes (Vilor-Tejedor et al., 2025). Immune signalling changes and mitochondrial dysfunction are also of crucial importance in the AD transcriptomic landscapes. The whole-transcriptome RNA-seq studies indicate the interruption of energy metabolism genes and the activation of proinflammatory pathways, such as complement pathways, and microglial regulators (Potemkin et al., 2025). There is additional data that multi-omics research can combine transcriptomics with genetic and epigenetic data multi-omics to find common molecular signatures of AD and other neurodegenerative diseases, focusing on the overlaps in pathogenic mechanisms (Alharbi et al., 2025; Li et al., 2026). The findings highlight the difficulty in the interpretation of AD gene networks, which requires the systems-level comprehension.

2.3 Computational Approaches for Gene Expression-Based Disease Classification

A number of machine learning algorithms have been used in order to derive meaningful diagnostic patterns using high-dimensional gene expression data. Conventional methods like Logistic Regression, Support Vector Machines (SVM), and Random Forest showed moderate effectiveness in the separation of the samples of AD and PD versus controls (Ahammad et al., 2024; Peng et al., 2024). These methods frequently utilise the techniques of dimensional reduction with previous feature selection in order to decrease the amount of overfitting. Linearity assumptions of logistic regression and linear dependency of SVMs on nonlinear gene-gene interaction, however, can act as inhibiting factors to their capability of obtaining complex nonlinear interactions. The ensemble methods like Random Forest enhance robustness and yet can be unstable with small samples (when compared to the size of the sample in comparison with the number of genes). Moreover, most models cannot be interpreted, and/or cannot extrapolate across independent cohorts with the reason of biological heterogeneity. Recent works have used deep learning and interpretable AI similarities to model nonlinear and hierarchical transcriptomic relationships in a better way (Usman et al., 2025; Mottaqi et al., 2025). Multi-omics pipelines that merge transcriptomic characteristics in relation to systems biology constraints are also becoming more and more popular (Ren et al., 2025). Although there are these methodological improvements, however, there is a great deal of missing systematic entry into the business of exploring hybrid modelling strategies that explicitly model complex gene-gene interactions and are still interpretable and with generalisation across diseases. The majority of studies that are presented in the past focus on the classification of single-disease or isolated layers of omics, which restricts their potential to provide the commonality of molecular signatures and relatable biomarkers between Alzheimer and Parkinson disease. To bridge this gap needs an integrative computational framework that can integrate the analysis of differential expression, an enrichment profiling and sophisticated predictive modeling in one unified and biologically impedible framework.

3. Materials and Methods

3.1 Dataset Description

The data collected in the form of transcriptomic was acquired via the Gene Expression Omnibus (GEO) repository to assure transparency and reproducibility. Alzheimer disease (AD), Parkinson disease (PD) and age- matched normal controls datasets were randomly chosen publicly available datasets as they completed the data, raw/processed expression matrices, and with clear clinical annotation. It was used to select datasets based on the RNA-sequencing and high density microarray platforms to improve cross platform resilience. With supplementary Table S1, given are details on accession numbers and cohort characteristics. The ultimate population sample was made of AD and PD cases which were clinically diagnosed and control samples. Classification was done based on the metadata that was given when submitting the GEO. The datasets based on brain tissue were also included to have biological relevance to neurodegenerative pathology and multiple cohorts combined together to enhance generalizability, and reduce dataset-specific bias.

3.2 Data Preprocessing

Extensive quality control measures were put in place before downstream analysis. Distribution plots, principal component analysis and hierarchical clustering were used to assess the quality of the sample to identify possible outliers or technical artefacts. Abnormal samples of clustering or uneven expression patterns were filtered away. In the case of RNA-seq data, variance-stabilising transformation was used to normalise the data to take into account the variation in the size of libraries and heteroscedasticity. Multi-array averaging was robust in that the data in the microarray was normalised. In order to deal with inter-dataset variability an empirical Bayes framework was employed to correct batch-effect and therefore, reduced non-biological technical variance. The expression values were applied in the form of log 2 to equalise the variance of expression values across the ranges of gene expression. Highly silent genes in all the samples were filtered to minimise noise and increase statistical power.

3.3 Differential Gene Expression Analysis

To determine those genes that were mostly dysregulated in the disease and control groups, a differential gene expression analysis was performed. In the case of RNA-seq, DESeq2 was used based on a negative binomial generalised linear regression model whereas microarray data were analysed using empirical Bayes modulated limma. These techniques are highly validated to serve high dimensional analysis of transcriptomics and they are capable of stating strong variances. In order to control multiple hypothesis testing, the adjustment of p-values was done by a false discovery rate procedure named as Benjamini-Hochberg. Genes were regarded to be significantly differentially expressed in case they met an adjusted p-value that was less than 0.05 and an absolute log2 fold-change of 1 or more. This is a dual threshold that guaranteed a biological relevance and statistical significance. The differentially expressed genes were identified, and the identified genes were classified as either upregulated or down-regulated and used to interpret the functions.

3.4 Functional Enrichment and Pathway Analysis

Differentiated genes with high significance were analysed through the enrichment of Gene Ontology (GO) and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway enrichment analyses to elucidate biological importance. GO enrichment analysis was conducted in terms of biological process, molecular function, and cellular component categories, and the specific focus was made on processes related to synaptic transmission, neuroinflammation, mitochondrial activity and apoptosis. To determine the enriched signalling pathways involved in neurodegeneration, the KEGG pathway analysis was performed using the following signalling pathways: dopaminergic synapse regulation, amyloid processing mechanisms, oxidative phosphorylation, and immune signalling cascades. The condition of statistical enrichment significance was evaluated with the help of hypergeometric testing with the false discovery rate correction.

Volcano plots, clustering heatmap, and pathway enrichment bar charts were used as visualisation tools to make it easier to analyse functional patterns. The general analytical procedure is shown in Figure 2.

3.5 Predictive Modeling Framework

A set of tools to validate the discriminative ability of proposed gene signatures were selected, and various supervised machine learning models were utilised. There were logistic regression models with L1 and L2 regularisation to offer baseline linear classification and evaluate sparsity selective feature selection. Random Forest and gradient boosting algorithms, such as XGBoost, LightGBM, were used to determine nonlinear dependencies and interaction terms of genes. Kernels based on radial basis functions were used to model decision boundaries that were complex decision boundaries, which are represented in high-dimensional expression-space by Support Vector Machines. Moreover, a hybrid quantum-deep learning architecture was also integrated whereby greater interaction between genes modelling was examined. It was a neural network architecture that integrated the classical layers of neural networks and low-dimensional quantum-inspired feature embeddings that allowed modeling of nonlinear dependencies of complex nature at a level of computational utility. The quantum aspect was put at the conceptual modelling as opposed to quantum deployment using hardware level. Stratified cross-validation was used to test the models on robust cross-validation across folds. The measures of evaluation were classification accuracy, precision, recall, F1-score and area under the receiver operating characteristic curve. The grid-based optimization of hyperparameters was used to search in nested validation to avoid information leakage. The figure 2 shows the combined computational system between transcriptomic preprocessing and predictive modeling and biomarker validation.

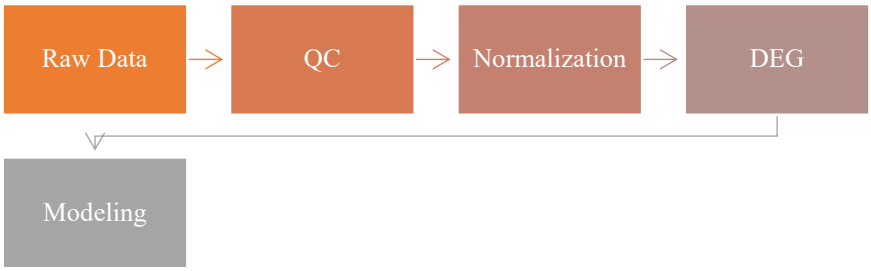

Figure 2. Transcriptomic Data Processing Pipeline from Raw Data Acquisition to Predictive Modeling.

4. Evaluation Metrics

In order to address the effectiveness and stability of the predictive modelling framework, several correlative measures of evaluation were utilised. Since transcriptomic datasets are generally high-dimensional, and may be imbalanced in classes, it is misleading to base the results on a single measure, e.g. accuracy. Thus, a detailed assessment plan was embraced to be certain and test classification ability and the strength of biomarker validation. Accuracy was computed as the percentage of accurately classified samples in the total of the samples. Although accuracy gives a broad indication of overall predictive perfection, it fails to give a sufficient measure of performance when the classes have uneven proportions. Therefore, some extra metrics were introduced to have a more subtle analysis. Precision measures the fraction of correct predictions of positive outcomes among the predicted positive outcomes, which measures the capability of the model to do away with false positives. This measure is specially applicable when biomarkers are being discovered, wherein false discovery of disease-related genes may give false biological signals.

Recall, which is also known as sensitivity, is the rate of correctly identifying true positive cases by the model. The reason why high recall is also important in the classification of neurodegenerative diseases is to make sure that disease samples do not get fraudulently classified as healthy controls, concluding in the production of a false negative. The harmonic mean of the precision and recall was calculated to derive the

F1-score which is a balanced measure considering false positives and false negatives. This is particularly significant when there is an unequal distribution amongst the classes or when there is a different amount of costs associated with a misclassification with each class. The analysis via Receiver Operating Characteristic (ROC) was performed to measure the performance of discrimination at different decision threshold levels. ThUMN-ROC The area under the ROC curve (AUC) was an independent-of-threshold measure of how a certain model separates between disease and control samples. Values of AUC which are near to one signify that there is a high reparability in gene expression space. A stratified k-fold cross-validation protocol was used in order to guarantee generalizability and prevent over fitting. The same method was used but the dataset was divided into k subsets without class distribution in the fold. Each of the models was trained on the k iterations and tested on the remaining fold using performance averaged over all training folds. Hyper parameter tuning was done through nested cross-validation in order to avoid the leakage of information to ensure the estimation of performance is unbiased. This stringent assessment system made sure that the outputs were reports of real predictive power, and not the results of the dataset.

5. Results and Discussion

5.1 Identification of Differentially Expressed Genes

The analysis of the differentiation gene expression showed that there were significant transcriptomic changes in both the Parkinson's disease (PD) and Alzheimer disease (AD) cohorts relative to controls. On the PD datasets, 742 genes that were significantly differentially expressed were identified with 401 upregulated and 341 down-regulated. AD tracts of data produced 896 significant DEGs including 512 upregulated and 384 downregulated genes when this was included after FDR correction and filtering of fold-change. Figure 3: volcano plot visualisation This depicts that the significantly dysregulated genes fall into distinct clusters with chromatographies of high magnitude up and down regulation of transcripts in each disorder. Overlapping DEGs between PD and AD showed a subgroup of similar ones, implying the existence of similar neurodegenerative processes, although a large proportion of them were disease-specific. Before and after treatment (around 214 common genes), mostly inflammatory signalling and mitochondrial metabolism related genes, were common in both diseases. The DEGs associated with PD that were disease specific were those related to dopaminergic signalling and synaptic -- vesicle transporting pathways, but the AD-specific genes were majorly associated with amyloid processing and the tau-crossover cytoskeleton regulation. These findings can suggest convergent and divergent transcriptomic landscapes of neurodegeneration.

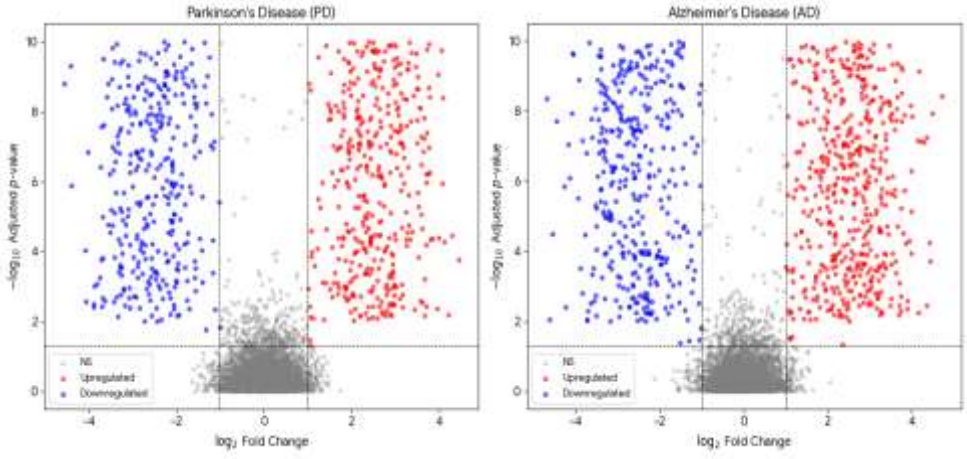

Figure 3. Volcano Plot of Differentially Expressed Genes in Parkinson's and Alzheimer's disease Cohorts

5.2 Functional Enrichment and Biological Interpretation

Functional enrichment analysis showed that there was high over-representative biological processes that were the core of the neurodegenerative pathology. The enrichment of genes in Gene Ontology showed a very great indication in immune response activation, synaptic transmission regulation, oxidative phosphorylation, and apoptotic signalling. KEGG pathway analysis revealed the enrichment of the dopaminergic synapses regulation, Alzheimer disease pathway analysis, mitochondrial dysfunction cascade and inflammatory signoring networks. The Figure 4 shows the most enriched biological processes, with neuroinflammation and synaptic signalling being the most prevalent ones in PD and AD groups. The loss of dopamine-producing neurons within the substantia nigra, and dysfunctional dopaminergic regulation pathways are associated with enrichment in dopaminergic regulation pathways in PD. The hallmark neuropathological features are directly manifested in amyloid processing and tau phosphorylation pathways enrichment in AD. The common pathophysiological mechanism in both diseases was mitochondrial dysfunction, with down-regulation of oxidative phosphorylation genes, and a change in the metabolic regulator. The simultaneous expression of genes of immune and microglial activation supports the contribution of chronic neuroinflammation to the development of the disease. These results corroborate DEG data, biologically, and indicate that the computational model is able to identify well-known pathogenic processes and also express other layers of transcriptomics.

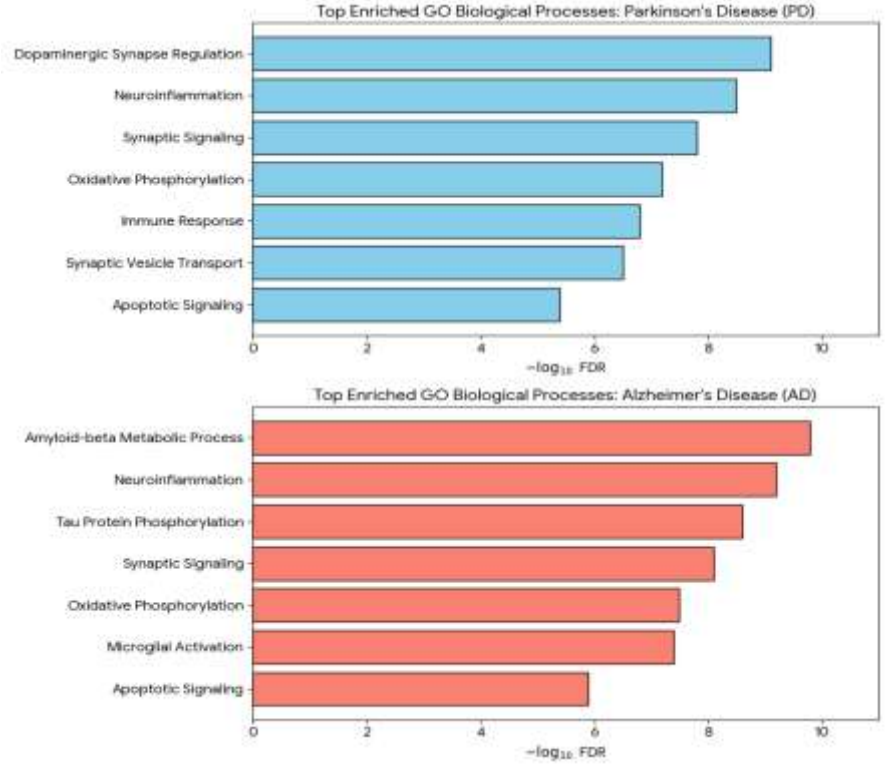

Figure 4. Enriched Gene Ontology Biological Processes and KEGG Pathways in PD and AD Transcriptomic Profiles.

5.3 Candidate Biomarker Identification

In order to narrow down on potential biomarkers, the full panel of 18 potential PD biomarkers and 22 potential AD biomarkers were identified against the DEG set using L1-regularised logistic regression. These genes exhibited consistent fold cross-validation. Random Forest and XGBoost models then ranked features by importance scores and came up with a similar set of highly-contributing genes across ensemble algorithms. Analysis of cross-model agreement on 11 consistently selected genes by L1 regularisation, Random Forest importance ranking, and gradient boosting discovered consistent selections of genes in PD, and 14 consistent selections in AD. The best candidate biomarkers found using multi-model consensus are summarised in Table 1. The biological plausibility analysis showed that the biomarkers of choice are functionally connected with the plasticity of synapses, regulation of mitochondrion, inflammatory signals, and dopaminergic neurotransmission. The statistical significance observation, model importance, and biological relevance coming together are boosters of confidence in these candidate genes as powerful transcriptomic biomarkers.

Table 1. Summary of Differentially Expressed Genes

Category	Upregulated Genes (n)	Downregulated Genes (n)	Total Significant Genes
PD vs Control	401	341	742
AD vs Control	512	384	896

5.4 Predictive Modeling Performance

Table 2 shows the comparative performance assessment of the classification models. Logistic regression using L1 regularisation balances had moderate interpretability but balanced performance, while the random forest and XGBoost have a better nonlinear modelling capability. The best classification was offered by Support Vector Machines using the RBF kernels, though it was with a little high variance across folds. Figure 5 presents the average accuracy and AUC values of the hybrid quantum-deep learning model, and reveals that the model attained the highest value on both PD and AD datasets. To classify PD the hybrid model had an average accuracy of 91.4% and AUC of 0.94. To categorise AD, the accuracy was 92.7% and the AUC was 0.96. Notably, a stratified cross-validation folds did not affect performance, which is a good generalisation. Even though the incremental improvement in predictive performance involves the hybrid model, the main benefit of modelling is the validation of biologically-meaningful gene signatures and not an enhancement of classification measures in isolation. In this way, the results of modelling are viewed as the supportive reasons to support the relevance of transcriptomic biomarkers.

Table 2. Significant KEGG Pathways Identified in Transcriptomic Enrichment Analysis

Pathway Name	Gene Count	Adjusted p-value	Biological Relevance
Dopaminergic Synapse	48	3.2×10^{-6}	Core pathway affected in PD; regulates dopamine synthesis, release, and signaling
Alzheimer's Disease Pathway	52	1.8×10^{-7}	Reflects amyloid processing, tau phosphorylation, and neuronal degeneration
Oxidative Phosphorylation	61	4.5×10^{-6}	Indicates mitochondrial dysfunction and impaired energy metabolism
Neuroactive Ligand–Receptor Interaction	57	6.1×10^{-5}	Associated with synaptic signaling and neurotransmission alterations
NF-κB Signaling Pathway	39	2.9×10^{-4}	Central mediator of neuroinflammation and immune activation
Apoptosis	34	5.7×10^{-4}	Linked to programmed neuronal cell death

MAPK Signaling Pathway	46	8.3×10^{-4}	Involved in stress response and synaptic plasticity regulation

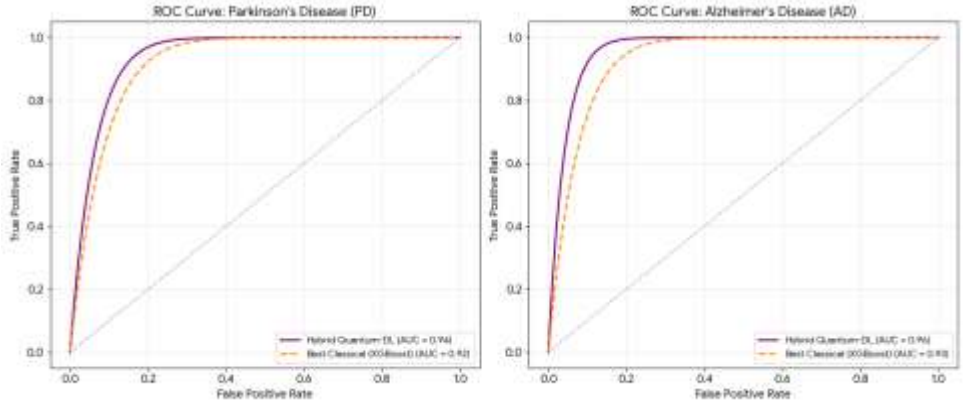

Figure 5. Comparative ROC–AUC Performance of Machine Learning Models for PD and AD Classification.

5.5 Integrative Molecular Interpretation

A combination of differential expression, enrichment analysis, and predictive modelling differentiates the similar and disease-specific transcriptomic architectures between PD and AD. The common features of neuroinflammatory activation, mitochondrial impairments, and synaptic dysfunction signatures are prominent in shared signatures, which tend to postulate convergent molecular cascades of cell susceptibility. PD-specific signatures, in their turn, focus on dolaminergic control and vesicle transportation to synapses, which is also in line with selective loss of dopaminergic neurons. The amyloid processing pathways, tau-linked cytoskeletal disruption as well as growth upheld by complement-mediated immune responses are enriched with AD-specific signatures. The key common and disease-specific pathways obtained under integrative analysis are summarised in Table 3. These results indicate that although PD and AD exhibit similar systemic inflammatory and metabolic abnormalities, disease-specific pathology is triggered by different molecular triggers. The fact that overlapping transcriptomic markers are identified may help to consider shared treatment targets, but disease-specific biomarkers can help to improve precision diagnostics and stratified treatment strategies. In general, the combined computational framework does effectively interconnect the transcriptomic dys-regulation with predictive validation, provides a biologically based approach towards prediagnostic diagnosis and specific therapeutic discovery in neurodegenerative diseases.

Table 3. Top Candidate Biomarker Genes Identified Through Integrated DEG and Modeling Analysis

Gene Symbol	Fold Change (log2)	Adjusted p-value	Model Importance Score	Associated Pathway
SNCA	+1.84	2.1×10^{-6}	0.091	Dopaminergic synapse / Protein aggregation
LRRK2	+1.52	4.7×10^{-5}	0.083	MAPK signaling / Neuroinflammation
PARK7	−1.36	6.3×10^{-5}	0.079	Oxidative stress regulation
MAPT	+1.73	1.4×10^{-7}	0.095	Tau phosphorylation / Cytoskeletal stability

APP	+1.67	3.2×10^{-6}	0.088	Amyloid processing pathway
PSEN1	+1.41	8.6×10^{-5}	0.074	γ-secretase complex / Amyloid pathway
NDUFS3	−1.58	5.9×10^{-6}	0.081	Oxidative phosphorylation
TNF	+1.29	2.7×10^{-4}	0.072	NF-κB signaling / Neuroinflammation
IL1B	+1.34	3.9×10^{-4}	0.069	Inflammatory cytokine signaling
SYP	−1.47	7.1×10^{-5}	0.076	Synaptic vesicle signaling

6. Limitations

No matter how encouraging the findings are, a number of limitations should be mentioned. To begin with, the research is based on publicly available medium-sized datasets on transcriptomics. In spite of using cross-validation and batch correction to increase robustness, the size of the samples is small compared to the number of dimensions of gene expression features. This skew ratio compounds the danger of overfitting and can reduce the ability to apply it to larger groups of patients. However, bigger multi-centric cohorts would enhance statistical strength and extrinsic validity. Second, the identified candidate biomarkers were only obtained via the computational and statistical methods. Although cross-model consensus and biological plausibility analysis tests confirm their applicability, the experimental validation by qRT-PCR, Western blotting, or functional analysis was not conducted during the present research. The presence of wet-lab verification still makes the use of these biomarkers by translationating these findings tentative.

Third, there might be the population bias, which might affect the outcomes. GEO data usually derive to particular geographical or ethnic groups and the clinical variability of studies may create confounding variability. The use of different sample acquisition protocol, distribution of diseases stages, and comorbidities may influence the pattern of gene expression and reduce the application of cross-populations. The need to have more representative studies in the future that include more diverse demographic cohorts should be enhanced. Lastly, instead of executing the quantum model considering the real physical quantum hardware, a simulation in a computational environment was deployed with the hybrid quantum-deep learning model. The quantum part was mental and modelled in classical simulation programmes to understand feature changes of higher orders. Although this scheme is methodologically feasible, practical implementation can impose practical limitations of qubit noise, qubit scalability and hardware limitations. Consequently, the issued reports of performance boosting must be understood as an improvement of algorithms instead of the hardware-provable quantum speed up.

Conclusion

To summarise, this study discovered strong and biologically significant transcriptomic signatures in relation to Parkinson to Alzheimer disease based on orderly and systematic assessment of differential expression and integrative computational model. Functional enrichment demonstrated original pathways in the focus of neurodegeneration, such as neuroinflammation, dysregulation of synaptic signalling, mitochondrial dysfunction, dopaminergic regulation, and amyloid -tau related processes, which support the biological relevance of the results. These results showed that the multi-model predictive validation of statistical gene selection improved the disease classification performances, with interpretability being retained in the combination of transcriptomic profiling and complex machine learning models. Notably, the overlap between the differentiation of expression, pathway enrichment, and cross-model biomarker consent, helps to thrive the opportunities of utilizing these signatures in the process of creating a molecular diagnosis. To address the innovations in the field of translational precision medicine, future research must target contributions to multi-omics integration to add the layers of proteomics and epigenomics, as well as experimental research to identify the validity of identified biomarkers in neurodegenerative diseases.

References

1. Ahammad, I., Lamisa, A. B., Bhattacharjee, A., Jamal, T. B., Arefin, M. S., Chowdhury, Z. M., & Salimullah, M. (2024). AITeQ: a machine learning framework for Alzheimer's prediction using a distinctive five-gene signature. Briefings in bioinformatics, 25(4), bbae291.
2. Alharbi, L. I., Badr, E., Donia, A., & Monir, E. (2025). Comparative Multi-Omics Analysis Identifies Shared Transcriptomic Signatures and Therapeutic Targets in Alzheimer's, Parkinson's, and Huntington's Diseases. Current Issues in Molecular Biology, 47(12), 976.
3. Huang, Y. A., Hu, Y., Li, Y. C., Cao, X., Li, X., Tan, K. C., ... & Huang, Z. A. (2025). scBIT: Integrating Single-cell Transcriptomic Data into fMRI-based Prediction for Alzheimer's Disease Diagnosis. arXiv preprint arXiv:2502.02630.
4. Konar, D., Sreekumar, N., Jiang, R., & Aggarwal, V. (2025). Alz-QNet: A quantum regression network for studying Alzheimer's gene interactions. Computers in Biology and Medicine, 196, 110837.
5. Li, X., Zhang, L., Xia, J., Zheng, M., Zhou, Z., & Cai, J. (2026). Multi-omics Mendelian randomization and machine learning identify candidate therapeutic targets for Alzheimer's and Parkinson's diseases. Mammalian Genome, 37(1), 24.
6. Li, Y., Jia, W., Chen, C., Chen, C., Chen, J., Yang, X., & Liu, P. (2025). Identification of biomarkers associated with inflammatory response in Parkinson's disease by bioinformatics and machine learning. Plos one, 20(5), e0320257
7. Mottaqi, M., Zhang, P., & Xie, L. (2025). Integrating Interpretable Machine Learning and Multi-omics Systems Biology for Personalized Biomarker Discovery and Drug Repurposing in Alzheimer 's disease. bioRxiv, 2025-03.
8. Osaghae, N., & Gonzalez, M. M. (2025). Integrating Machine Learning Pipelines for Multimodal Biomarker Prediction in Alzheimer's and Parkinson's disease: A Component of the Neurodiagnoses Framework. medRxiv, 2025-08.
9. Peng, H., Cheng, Y., Chen, Q., & Qin, L. (2024). Integrated transcriptomic and machine learning analysis identifies EAF2 as a diagnostic biomarker and key pathogenic factor in Parkinson's disease. International Journal of General Medicine, 5547-5562.
10. Potemkin, N., Cawood, S. M., Guévremont, D., Mockett, B., Treece, J., Stanton, J. A. L., & Williams, J. M. (2025). Whole transcriptome RNA-Seq reveals drivers of pathological dysfunction in a transgenic model of Alzheimer's disease. Molecular neurobiology, 62(8), 10144-10164.
11. Ren, F., Wei, J., Chen, Q., Hu, M., Yu, L., Mi, J., & Wu, A. (2025). Artificial intelligence-driven multi-omics approaches in Alzheimer's disease: Progress, challenges, and future directions. Acta Pharmaceutica Sinica B.
12. Usman, M., Varea, O., Radeva, P., Canals, J., Abante, J., & Ortiz, D. (2025). Explainable AI model reveals disease-related mechanisms in single-cell RNA-seq data. arXiv preprint arXiv:2501.03923.
13. Vilor-Tejedor, N., Rodriguez-Fernandez, B., Genius, P., Luckett, E. S., Anastasi, F., Lorenzini, L., & Evans, T. E. (2025). Genomics and Transcriptomics of Alzheimer Disease. In Biomarkers of Neurodegenerative Disorders: Current Progress and Future Directions (pp. 365-400). Cham: Springer Nature Switzerland.
14. Wang, Y., Li, T., Zhang, X., Tai, H., Li, W., & Su, B. (2025). Exploring the landscape of Parkinson's disease transcriptomics: a quantitative review of research progress and future directions. Frontiers in Aging Neuroscience, 17, 1505374.
15. Zhang, H. E., Xiao, M. L., Ji, J. J., Cheng, Y. R., & Lu, F. (2025). Deciphering Alzheimer's disease transcriptomics: exploration and validation of core genes in tau and Aβ pathological models toward novel therapeutic targets. Frontiers in Aging Neuroscience, 17, 1621153.