GMR
The Original

ARTIFICIAL INTELLIGENCE-ASSISTED GENOMIC PREDICTION MODELS FOR EARLY DIAGNOSIS OF COMPLEX HUMAN DISEASES

Dr Amit Kumar[1*], Dr.Manojkumar S.Langote[2], Amal Thomas K[3], A. Sivanesh Kumar[4], Ms.Anamika Upadhyay[5]

[1*]Associate Professor, Department of Mathematics, Government Degree College Modinagar, Ghaziabad UP. Email Id : akumarmaths@gmail.com ORCID ID: 0009-0003-5231-1521
[2]Associate Professor, MCA, Computer Applications, Dr.Moonje Institute of Management, Nashik, Maharashtra. 4122013. ORCID ID:0009-0000-4424-1090, Email ID: manojlangote@gmail.com
[3]Associate Professor, Artificial intelligence, Artificial intelligence and data science department, Royal College of Engineering and Technology Thrissur, Thrissur. 680602. amal.thomas@royalcet.ac.in, 0000-0001-5103-7776
[4]Associate Professor, Computer Science and Business Systems, Ramco Institute of Technology, Rajapalayam. Tamil Nadu., 626117, India. E-MAIL: kas.sivanesh@gmail.com, Orcid: 0000-0002-6558-1657
[5]Assistant Professor, MCA, Specialization: Computer Application, University -Savitribai Phule Pune University, Dr.Moonje Institute of Management and Computer Studies, Nashik 422005, ORCID ID : 0009-0006-6468-2815, Email ID: anamikaupadhyay100@gmail.com

ABSTRACT

Complex neurodegenerative diseases are difficult to diagnose early on because of the heterogeneity of molecular alterations and overlapping clinical stages. Transcriptomic profiling based on blood is a less invasive approach to detecting disease-related genomic activity. This study developed and evaluated artificial intelligence-assisted genomic prediction models for classifying Alzheimer's disease, mild cognitive impairment, and cognitively normal controls using peripheral blood gene expression data. Normalized microarray expression data from 329 samples were preprocessed, and missing values were handled using median imputation. Analysis of variance was applied to select the top 1000 discriminative features. Logistic regression, support vector machine, and random forest models were trained using a stratified 80:20 split, and performance was evaluated using accuracy, macro-averaged precision, recall, F1-score, confusion matrix, and multiclass ROC analysis. Functional enrichment analysis was conducted to assess biological relevance. Class-balanced logistic regression achieved the best performance, with an accuracy of 69.7%, macro precision of 0.6917, macro recall of 0.6966, and macro F1-score of 0.6873. ROC analysis showed strong discrimination, with AUC values of 0.89 for controls, 0.83 for Alzheimer's disease, and 0.76 for mild cognitive impairment. Enrichment results highlighted pathways related to ribosomal activity, oxidative phosphorylation, mitochondrial function, and neurodegeneration. These findings indicate that transcriptomic signatures can support interpretable artificial intelligence-assisted prediction of early disease states.

KEYWORDS: Artificial intelligence, genomic prediction, Alzheimer's disease, transcriptomics, mild cognitive impairment

INTRODUCTION

Multifactorial etiologies of genetic, environmental, and molecular interactions have characterized complex human diseases, and hence their diagnosis and management have been especially difficult. The development of human disease genetics has greatly enhanced the comprehension of human diseases, showing complex regulatory networks and heterogeneous biological mechanisms underlying disease onset and progression (Claussnitzer et al., 2020). Of these, neurodegenerative disorders have been a significant international health issue because of their progressive nature and rising prevalence as the population ages. These conditions are characterized by progressive neuronal loss, which results in cognitive and functional impairment, and cannot be cured with a high level of research conducted (Ricci, 2024). The recent events have also demonstrated the complexity of the pathways of neurodegenerative diseases and the necessity to adopt an integrative approach that would combine molecular understanding with the application of advanced analytical methods (Van Schependom & D'haeseleer, 2023). Recent discoveries of valid biomarkers have been at the forefront of enhancing the accuracy of diagnosis, especially at early stages of the disease. Biomarkers, which are based on molecular and cellular processes, can be used to gain insight into disease mechanisms and to offer the potential of early detection and monitoring (Hansson, 2021). Transcriptomic profiling based on blood has also shown promise as a minimally invasive technique to identify molecular signatures related to cognitive impairment and disease progression (Sood et al., 2015). The genetic architecture of Alzheimer's disease, which is one of the most common neurodegenerative disorders, is highly complex and involves a number of interacting loci and regulatory elements (Andrews et al., 2023). This is further complicated by the non-homogeneous clinical presentation of the disease, thus making it difficult to accurately and early diagnose the disease. The changing conceptualization of Alzheimer's disease highlights its conceptualization as a complex biological puzzle, one that requires complex analytical models to unravel its underlying mechanisms (Ferrari & Sorbi, 2021).

It is common knowledge that early diagnosis is a highly important element in enhancing clinical outcomes of neurodegenerative diseases. Mild cognitive impairment (MCI) as a transition phase represents a prime intervention window, as it occurs prior to the development of irreversible neurodegeneration and provides the opportunities of

therapeutical interventions aimed at postponing the disease progression (Mekulu et al., 2025). Nevertheless, conventional diagnostic methods usually do not address subtle molecular changes at this stage, which leads to delayed diagnosis and low treatment efficacy (Rasmussen and Langerman, 2019). As a result, there has been an increased interest in developing biomarker-based diagnostic systems that are capable of detecting disease-related changes prior to the appearance of overt clinical symptoms (Ausó et al., 2020). Parallel to these developments, artificial intelligence (AI) has become a formidable power in medical diagnostics, specifically in terms of analyzing high-dimensional biological data. AI practices, such as machine learning algorithms, have shown the capacity to identify complicated patterns and relationships that cannot be easily identified using non-AI statistical techniques (Ghaffar Nia et al., 2023). The methods have been widely used in the prediction and classification of disease, providing a more accurate and efficient methodology for the diagnostic processes. The AIs that are used to develop diagnostic systems can integrate various types of data and generate predictive models that will help in the clinical decision-making process, and thus improve the overall diagnostic procedure (Kaur et al., 2020).

Although these developments have occurred, the usage of AI in genomic and transcriptomic data to diagnose diseases at an early stage is a developing field. Although a number of recent studies have examined machine learning methods to classify Alzheimer's disease, many have isolated binary classification problems or have resorted to small feature selection strategies. Moreover, the biological interpretability of these models is usually inadequate and little effort is made to verify whether the identified features can be attributed to meaningful biological processes.

Current literature indicates that there is still a gap between discovering biomarkers and using AI to develop diagnostics; nevertheless, the literature indicates that there is still a gap between the discovery of biomarkers and the use of AI to develop diagnostics. The difficulties of dealing with high-dimensional transcriptomic data, such as redundancy of features and overfitting, are not adequately addressed by many studies. Also, the intermediate disease states, including MCI, have not been sufficiently examined, even though they are important in early diagnosis. It is also deficient in putting an emphasis on the connection between predictive models and underlying biological pathways that would limit the translational relevance of such studies. These constraints imply that methods must be developed that will provide robust feature selection, balanced classification schemes, and biological validation to guarantee both predictive and interpretable.

The current research will design and test artificial intelligence-based genomic prediction models to diagnose complex human diseases early in life using transcriptomic data. Using the expression profiles of genes based on blood, the study is aimed at classifying the individuals as being in Alzheimer's disease, mild cognitive impairment, and control, and in that way addressing the critical need for early detection. The methodology combines feature selection and class-balanced machine learning tools, which are supervised, to enhance predictive performance with minimal bias. Also, the study aims to determine the biological significance of the identified gene signatures based on the functional enrichment analysis, thus closing the gap between the computational prediction and molecular interpretation.

2. METHODOLOGY

2.1 Research Design

To build and test machine learning models to predict the early onset of complex human diseases using transcriptomic data, this study adopted a quantitative research design that is data-driven. The learning framework was carried out under supervision to categorize the individuals into the Alzheimer's disease (AD), mild cognitive impairment (MCI) and the control groups. The following workflow: data preprocessing, feature selection, model development, and performance evaluation. It focused on reproducibility, preventing data leakage, and interpretable modeling methods that could be applied to high-dimensional biological data.

2.2 Data Source

The data required in this research were acquired in a previously published study that was aimed at gathering transcriptomic signatures linked to cognitive health (Sood et al., 2015). The data consisted of the peripheral blood gene expression profiles produced with the help of the Illumina HumanHT-12 v3 microarray platform. It contained samples of persons with the diagnosis of Alzheimer's disease, mild cognitive impairment, and cognitively normal controls, which allowed for investigating the early-stage disease progression.

2.3 Data Preprocessing

The data on the gene expression was presented in a normalized format and was first represented with the probe identifiers as rows and samples as columns. The data was swapped to put samples in the form of observations and probes in the form of features. The identifiers of the samples were aligned with similar clinical labels obtained in the metadata.

The processes involved in data cleaning were the deletion of malformed or incomplete probe records and the processing of missing data. Features with values that are completely missing were dropped, whereas other missing values were imputed using the median of the values of the respective feature. All the feature values were transformed into numeric values to enable the ability to fit the machine learning algorithms.

2.4 Feature Selection

Since the dataset is highly dimensional, feature selection was conducted to eliminate noise and enhance the model's generalizability. An analysis of variance (ANOVA)-based and supervised method was used to detect features significantly linked to the class labels. The most ranked features were then picked based on the statistical scores, and a subset of 1000 genes was retained to be used in training the models. This method was used to make sure that the selected features were able to capture the class-discriminative information that was related to the disease status.

2.5 Model Development

Three machine learning models were created and compared: logistic regression, support vector machine (SVM), and random forest. The data were modeled using the logistic regression with a class weight to overcome moderate class

imbalance. SVM and random forest models were incorporated to evaluate non-linear and ensemble-based learning models, respectively.

In models that are sensitive to feature scaling, standardization was done by the use of z-score normalization. A uniform training pipeline was used to carry out the modeling process so that the methods could be fairly compared.

2.6 Training and Validation Strategy

A stratified split was used to divide the dataset into training and testing subsets in a way that maintained the proportion of classes in the dataset. The 80:20 split was taken as 80 % of the data was used to train the model, and 20 % were kept as independent data. This methodology made it possible to objectively assess the model's performance on unknown data.

2.7 Performance Evaluation

Multiple evaluation metrics, such as accuracy, precision, recall, and F1-score, were used to evaluate model performance. The metrics were reported as macro-average in order to give all classes equal weight. Moreover, the confusion matrices were employed to study the patterns of prediction by classes, and receiver operating characteristic (ROC) curves were plotted to analyze the discriminative power across classes.

2.8 Functional Enrichment Analysis

Functional enrichment analysis was used to examine the top-ranked genes to investigate their biological relevance. The gene identifiers were mapped to functional annotations, and enrichment analysis was carried out to identify the overrepresented biological processes and pathways. Gene Ontology (GO) biological processes, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were investigated. Significant terms were determined using p-values, and the most pertinent processes were reported to provide biological insight into the predictive model.

3. RESULTS

3.1 Dataset Characteristics

The study sample was a cohort of 329 peripheral blood samples representing three clinically diverse groups: Alzheimer's disease (AD), mild cognitive impairment (MCI), and cognitively normal controls. The distribution of samples across classes is summarized in Table 1.

Table 1. Distribution of samples across diagnostic classes

Class	Number of Samples	Percentage (%)
AD	145	44.07
Control	104	31.61
MCI	80	24.32

The data set has a moderate level of class imbalance, with AD having the highest percentage and MCI the lowest percentage. Although there is this imbalance, every class is well represented, such that multiclass modeling can be conducted. The comparatively lower representation of MCI is a result of its intermediate clinical status and may be a contributor to higher classification difficulty in subsequent analyses.

3.2 Model Performance Evaluation

The effectiveness of three machine learning models, namely, logistic regression, support vector machine (SVM), and random forest, was tested with a stratified train-test split. Table 2 shows performance metrics, such as accuracy and macro-averaged precision, recall, and F1-score.

Table 2. Comparative performance of machine learning models

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-score (Macro)
Logistic Regression	0.6970	0.6917	0.6966	0.6873
SVM	0.5606	0.4697	0.4647	0.4158
Random Forest	0.5758	0.5707	0.5080	0.5025

With an overall performance of 76.8%, class-balanced logistic regression had the highest overall performance, with an accuracy of 69.7% and a macro F1-score of 0.687. Conversely, SVM and random forest models showed less predictive performance, especially in terms of macro-averaged metrics, which indicates poorer ability to deal with class-level variability. The higher accuracy of the logistic regression indicates that the chosen transcriptomic features can be separated sufficiently well in a linear feature space when subjected to proper regularization and class balancing.

3.3 Classification Performance Analysis

An evaluation of the results of the class-wise predictions of the model with the highest performance (logistic regression) is demonstrated with the help of the confusion matrix in Figure 1. The confusion matrix shows that the control samples were most accurately classified, followed by AD and MCI. Particularly, 17 of 21 control samples were correctly identified, whereas 19 of 29 AD samples were correctly classified. In the case of the MCI group, 10 out of 16 samples were correctly predicted.

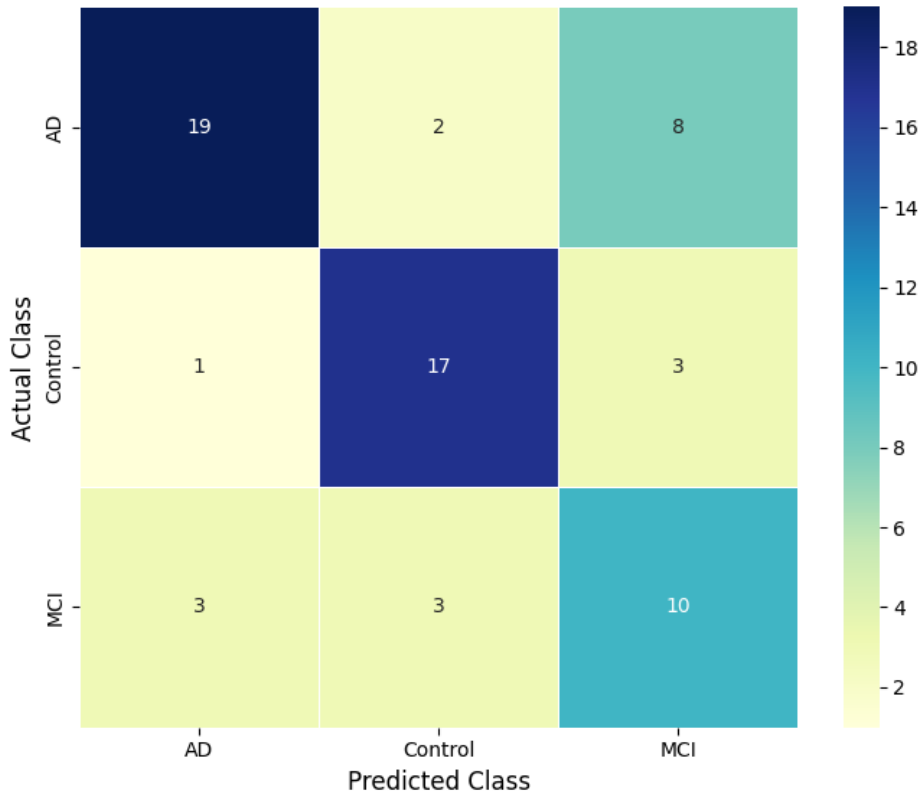


Figure 1. Confusion Matrix of the Best Prediction Model

The patterns of misclassification show that AD samples were sometimes predicted as MCI, and vice versa, which depicts the biological and clinical overlap of these two conditions. Notably, the misclassification between control and disease states (AD or MCI) was small, and this showed that the model was able to perform well in differentiating healthy individuals and those with cognitive impairment.

3.4 Receiver Operating Characteristic (ROC) Analysis

To further test the capability of model discrimination, multiclass ROC curves were constructed for each class as in Figure 2. The ROC analysis shows good classification performance across all classes, with the highest area under the curve (AUC = 0.89) being observed with the control samples (AUC = 0.89), followed by AD (AUC = 0.83) and MCI (AUC = 0.76).

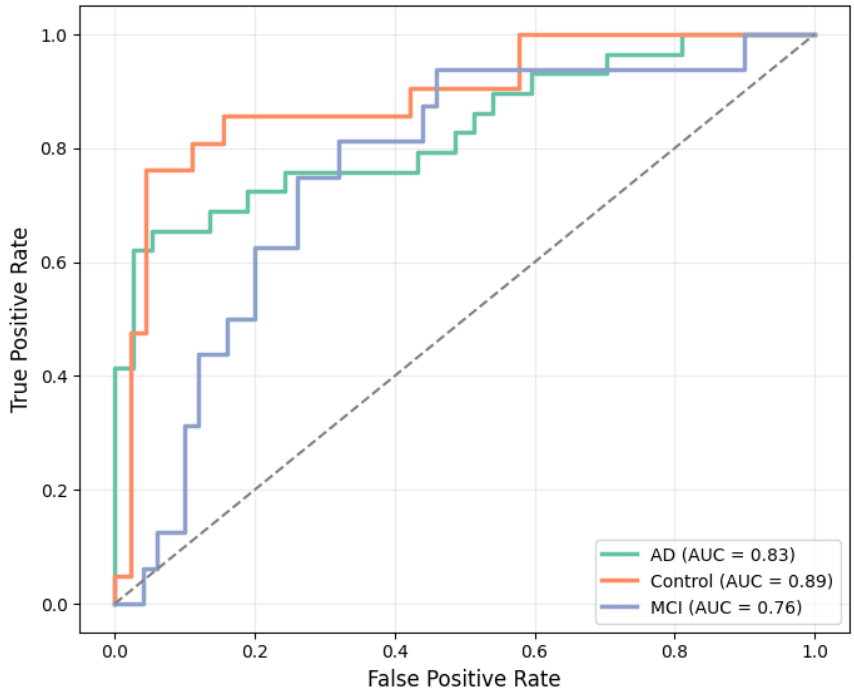


Figure 2. Multiclass ROC Curves

The relatively low AUC of MCI is an indication of its intermediate biological state and greater heterogeneity, making it difficult to clearly separate it from the AD and the control group. On the whole, the ROC curves reveal that the model has good discriminative capability, especially when it comes to distinguishing between healthy and diseased conditions.

3.5 Functional Enrichment Analysis

To determine the biological significance of the identified gene characteristics, functional enrichment analysis of the highest-ranking genes was conducted. Table 3 gives the most significantly enriched biological processes and pathways, and Figure 3 presents their corresponding visualization.

Table 3. Top enriched biological processes and pathways

Source	Term	P-value
KEGG	Ribosome	2.20e-09
KEGG	Oxidative phosphorylation	2.95e-09
KEGG	Parkinson disease	4.11e-08
GO:BP	Translation	5.18e-08
GO:BP	Protein biosynthetic process	5.18e-08
GO:BP	Proton transmembrane transport	1.69e-07
GO:BP	Oxidative phosphorylation	1.87e-06
KEGG	Huntington disease	2.54e-06
KEGG	Prion disease	3.81e-06
KEGG	Amyotrophic lateral sclerosis	4.81e-06

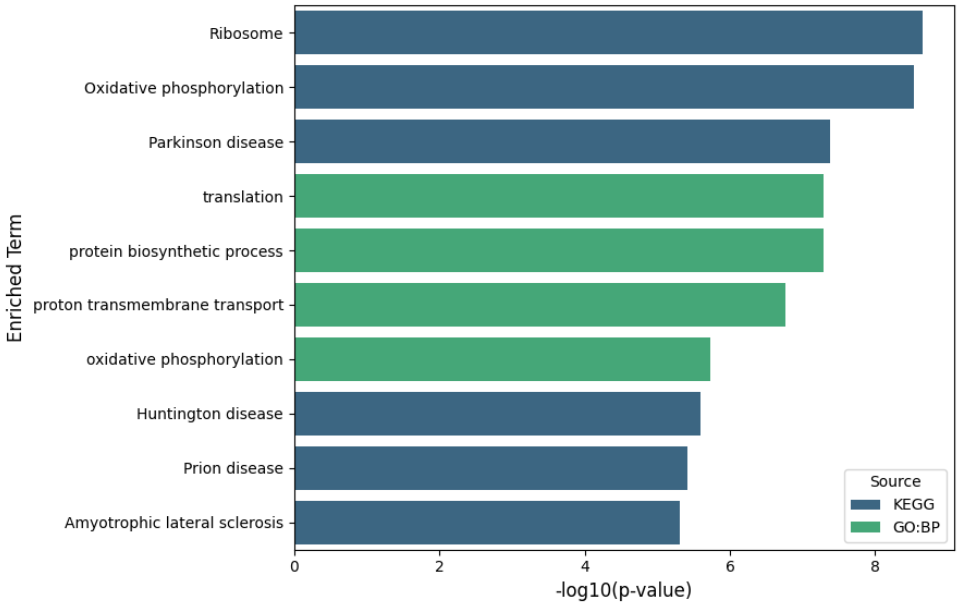


Figure 3. Top Enriched Biological Pathways and Processes

The results of Enrichment indicate a strong overrepresentation of pathways related to protein synthesis, mitochondrial functioning, and neurodegenerative diseases. Ribosomal activity and translation-related processes proved to be the most important biological functions, implying dysregulation of the protein production mechanisms. Also, the activation of the oxidative phosphorylation and proton transport signaling through enrichment is another well-understood characteristic of neurodegeneration. The fact that several neurodegenerative disease pathways are shared by Parkinson's, Huntington, and prions further indicates the presence of shared molecular mechanisms. Together, these results indicate that the predictive model represents biologically significant signals pertinent to the pathophysiology of the complex neurodegenerative disorders.

4. DISCUASSION

The classification outcomes show that transcriptomic profiles of peripheral blood can provide measurable disease-related signals that can be used to distinguish among Alzheimer's disease (AD), mild cognitive impairment (MCI), and cognitively normal controls. The most successful model was the class-balanced logistic regression, which was able to achieve an accuracy of 69.7% and a macro F1-score of 0.687. This performance is significant to a three-class biomedical prediction task, specifically because MCI is an intermediate and heterogeneous stage, but not a disease state. The model demonstrated the best results in recognizing control samples, followed by AD and MCI. This pattern was also supported by the ROC analysis, with the highest AUC being in the controls, followed by AD and MCI. The relatively worse MCI performance is clinically plausible due to the molecular similarities between MCI and normal aging, as well as MCI and early AD. This overlap was also reflected in the confusion matrix, where most of the errors were between AD and MCI and not between the control and disease groups. This implies that the model was able to capture variation related to disease,

but could struggle to distinguish between adjacent stages of cognitive decline. The strengthened biological plausibility of the predictive model was a result of the functional enrichment. The enrichment of ribosomal activity, translation, oxidative phosphorylation, proton transport, and mitochondrial translation indicates that the genes of interest can be associated with energy metabolism within the cell and protein synthesis. These mechanisms are also very pertinent to the neurodegenerative processes, in which dysfunctions of mitochondria, disturbed proteostasis, and maladjusted cellular homeostasis are often involved. The further enrichment of the Parkinson disease, Huntington disease, prion disease, and amyotrophic lateral sclerosis pathways further indicates that the identified transcriptomic signature reflects common mechanisms across neurodegenerative diseases and not random statistical variation.

The results are consistent with the previous literature, which demonstrates that artificial intelligence could be used to aid in the diagnosis of early AD by detecting subtle patterns of the disease and disease-related factors in complex biomedical data. Deep learning-based methods have already shown to hold potential in early AD prediction, albeit at the cost of larger datasets and potentially reduced interpretability as compared to simpler models (Ji et al., 2019). Recent research with combined imaging and genetic data has also identified an enhanced diagnostic potential with multimodal learning, emphasizing the importance of integrating a source of molecular and clinical data (Madanan et al., 2024). On the same note, neuroimaging-genomics models have proved useful in predicting future conversion to AD, further supporting the usefulness of biologically informed predictive models in assessing the disease at an early stage (Mirabnahrazam et al., 2022). In comparison with such multimodal models, the existing transcriptomic-only model is less data-heavy and simpler, but it might not be able to reflect the full biological complexity of the disease progression.

The preceding benchmarking research on predicting late-onset AD through genomic data also highlighted that no one algorithm consistently performs better than the others across all datasets, which is also consistent with the fact that logistic regression performed better than SVM and random forest in the present study (Oriol et al., 2019). The value of the tree-based models in detecting AD has been reported to have been achieved using single-nucleotide polymorphism-based gradient boosting, though the weaker performance of the XGBoost in the present workflow implies that transcriptomic microarray data may not necessarily benefit from the more complex non-linear models when the sample size is limited (Ahmed et al., 2022). The biological meaning is also aligned with the studies that combine the use of plasma biomarkers and genetics, where combined molecular markers have been demonstrated to enhance diagnosis and prediction of AD (Stevenson-Hoare et al., 2023). Genomic diagnostic studies based on learning further support the feasibility of using genetic and molecular features to classify AD, and also highlight the importance of careful model selection and validation (Jayasingh & Mallick, 2025). Similar outcomes were reported in other machine learning methods of early AD diagnosis (Uddin et al., 2023). It has also been suggested to use embedded feature selection frameworks to enhance early AD detection, which is in line with the current application of ANOVA-based feature selection to reduce high-dimensional noise before modeling (Mahendran & PM, 2022).

These findings indicate that transcriptomic profiles using blood can be utilized to support AI-assisted genomic prediction to enable early diagnosis of a complex neurodegenerative disease. Even though the dataset is not a readout of the variation in the DNA sequence, but rather a readout of the functional interaction of genomic processes in a disease condition, transcriptomic signatures do give a biologically significant readout of the molecular process interactions in a disease state. The interpretable and reproducible framework that is provided by the use of class-balanced logistic regression is valuable in the biomedical setting where transparency is paramount. The capacity to differentiate between MCI and AD and controls has significant implications for clinical intervention. Although the MCI classification was not as accurate as the AD classification and control classification, the fact that it was moderately detected indicates that the peripheral blood expression profiles could potentially hold early molecular signals of disease transition. The biological role of enrichment of mitochondrial and neurodegenerative pathways supports the significance of the chosen characteristics and enhances the connection between predictive modelling and disease mechanism.

There are some restrictions that are to be noted. To begin with, the analysis was founded on one transcriptomic dataset, which restricts external generalization. Second, the coverage of transcripts may be limited by the use of microarray data in comparison to RNA sequencing. Third, despite the fact that class balancing enhanced predictions of MCI, the intermediate nature of MCI still becomes a challenge to robust classification. Future research must test the model using independent cohorts and whether the model can be enhanced by adding additional layers of omics data, clinical variables, or neuroimaging markers. Particularly important would be external validation to confirm the stability of the picked gene signatures. Additional biological studies with the best-ranked genes can also be used to determine the candidate biomarkers that have translational potential.

5. CONCLUSION

The analysis of blood-based transcriptomic profiles by artificial intelligence demonstrated meaningful potential for early genomic prediction of complex neurodegenerative disease. Class-balanced logistic regression was the most effective as it offered balanced discrimination between the Alzheimer's disease, mild cognitive impairment and control group. The fact that mild cognitive impairment has moderate performance is consistent with the fact that it is in a transitional biological state, and thus is difficult to detect early disease states using molecular data alone. The functional enrichment analysis also indicated the biological relevance of the selected gene signatures, and pathways were identified that are related to ribosomal activity, oxidative phosphorylation, mitochondrial function, and common neurodegenerative mechanisms. These results suggest that the predictive features were not only statistically informative but also biologically relevant. The suggested framework offers a non-inferential and reproducible methodology of integrating artificial intelligence with functional genomics data to classify diseases at an early stage. It will be necessary to validate future use using independent cohorts and multi-omics datasets to strengthen generalizability and to facilitate clinical translation.

REFERENCES

1. Ahmed, H., Soliman, H., & Elmogy, M. (2022). Early detection of Alzheimer's disease using single-nucleotide polymorphism analysis based on gradient boosting tree. Computers in biology and medicine, 146, 105622.
2. Andrews, S. J., Renton, A. E., Fulton-Howard, B., Podlesny-Drabiniok, A., Marcora, E., & Goate, A. M. (2023). The complex genetic architecture of Alzheimer's disease: novel insights and future directions. EBioMedicine, 90.
3. Ausó, E., Gómez-Vicente, V., & Esquiva, G. (2020). Biomarkers for Alzheimer's disease early diagnosis. Journal of Personalized Medicine, 10(3), 114.
4. Claussnitzer, M., Cho, J. H., Collins, R., Cox, N. J., Dermitzakis, E. T., Hurles, M. E., ... & McCarthy, M. I. (2020). A brief history of human disease genetics. Nature, 577(7789), 179-189.
5. Ferrari, C., & Sorbi, S. (2021). The complexity of Alzheimer's disease: an evolving puzzle. Physiological reviews, 101(3), 1047-1081.
6. Ghaffar Nia, N., Kaplanoglu, E., & Nasab, A. (2023). Evaluation of artificial intelligence techniques in disease diagnosis and prediction. Discover Artificial Intelligence, 3(1), 5.
7. Hansson, O. (2021). Biomarkers for neurodegenerative diseases. Nature medicine, 27(6), 954-963.
8. Jayasingh, B. B., & Mallick, S. (2025, June). Learning Based Diagnosis of Alzheimer's Disease Using Genomic Data. In 2025 Second International Conference on Networks and Soft Computing (ICNSoC) (pp. 528-533). IEEE.
9. Ji, H., Liu, Z., Yan, W. Q., & Klette, R. (2019, June). Early diagnosis of Alzheimer's disease using deep learning. In Proceedings of the 2nd international conference on control and computer vision (pp. 87-91).
10. Kaur, S., Singla, J., Nkenyereye, L., Jha, S., Prashar, D., Joshi, G. P., ... & Islam, S. R. (2020). Medical diagnostic systems using artificial intelligence (AI) algorithms: principles and perspectives. Ieee Access, 8, 228049-228069.
11. Madanan, M., Kumar, R., Akhare, Y. D., Makanyadevi, K., Gamini, S., & Tiwari, M. (2024). Deep Learning-Based Diagnostic Models for Early Detection of Alzheimer's Disease Using MRI and Genetic Data. Frontiers in Health Informatics, 13(3).
12. Mahendran, N., & PM, D. R. V. (2022). A deep learning framework with an embedded-based feature selection approach for the early detection of the Alzheimer's disease. Computers in Biology and Medicine, 141, 105056.
13. Mekulu, K., Aqlan, F., & Yang, H. (2025). The mild cognitive impairment window for optimal Alzheimer's disease intervention. Journal of Alzheimer's Disease Reports, 9, 25424823251370768.
14. Mirabnahrazam, G., Ma, D., Lee, S., Popuri, K., Lee, H., Cao, J., ... & Alzheimer's Disease Neuroimaging Initiative. (2022). Machine learning based multimodal neuroimaging genomics dementia score for predicting future conversion to alzheimer's disease. Journal of Alzheimer's Disease, 87(3), 1345-1365.
15. Oriol, J. D. V., Vallejo, E. E., Estrada, K., Peña, J. G. T., & Alzheimer's Disease Neuroimaging Initiative. (2019). Benchmarking machine learning models for late-onset alzheimer's disease prediction from genomic data. BMC bioinformatics, 20, 709.
16. Rasmussen, J., & Langerman, H. (2019). Alzheimer's disease–why we need early diagnosis. Degenerative neurological and neuromuscular disease, 123-130.
17. Ricci, C. (2024). Neurodegenerative disease: from molecular basis to therapy. International Journal of Molecular Sciences, 25(2), 967.
18. Sood, S., Gallagher, I. J., Lunnon, K., Rullman, E., Keohane, A., Crossland, H., ... & Timmons, J. A. (2015). A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. Genome biology, 16(1), 185.
19. Stevenson-Hoare, J., Heslegrave, A., Leonenko, G., Fathalla, D., Bellou, E., Luckcuck, L., ... & Escott-Price, V. (2023). Plasma biomarkers and genetics in the diagnosis and prediction of Alzheimer's disease. Brain, 146(2), 690-699.
20. Uddin, K. M. M., Alam, M. J., Uddin, M. A., & Aryal, S. (2023). A novel approach utilizing machine learning for the early diagnosis of Alzheimer's disease. Biomedical Materials & Devices, 1(2), 882-898.
21. Van Schependom, J., & D'haeseleer, M. (2023). Advances in neurodegenerative diseases. Journal of clinical medicine, 12(5), 1709.