Journal of Integrative Neuroscience

J. Integr. Neurosci. **2024**; 23(4): 81
https://doi.org/10.31083/j.jin2304081

Original Research

Exploration of Imaging Genetic Biomarkers of Alzheimer's Disease Based on a Machine Learning Method

Yuanfei Wang[1], Xitao Wang[2,*]

[1]College of Electronic and Information Engineering, Harbin Vocational & Technical College, 150040 Harbin, Heilongjiang, China
[2]Information Center, The Second Affiliated Hospital of Harbin Medical University, 150086 Harbin, Heilongjiang, China
*Correspondence: wangxitao@hrbmu.edu.cn (Xitao Wang)
Academic Editor: Gernot Riedel
Submitted: 13 September 2023 Revised: 26 October 2023 Accepted: 6 November 2023 Published: 17 April 2024

Abstract

Background: Alzheimer's disease (AD) is an irreversible primary brain disease with insidious onset. The rise of imaging genetics research has led numerous researchers to examine the complex association between genes and brain phenotypes from the perspective of computational biology. **Methods**: Given that most previous studies have assumed that imaging data and genetic data are linearly related and are therefore unable to explore their nonlinear relationship, our study applied a joint depth semi-supervised nonnegative matrix decomposition (JDSNMF) algorithm to solve this problem. The JDSNMF algorithm jointly decomposed multimodal imaging genetics data into both a standard basis matrix and multiple feature matrices. During the decomposition process, the coefficient matrix A multilayer nonlinear transformation was performed using a neural network to capture nonlinear features. **Results**: The results using a real dataset demonstrated that the algorithm can fully exploit the association between strongly correlated image genetics data and effectively detect biomarkers of AD. Our results might provide a reference for identifying biologically significant imaging genetic correlations, and help to elucidate disease-related mechanisms. **Conclusions**: The diagnostic model constructed by the top features of the three modality data sets mined by the algorithm has high accuracy, and these features are expected to become new therapeutic targets for AD.

Keywords: Alzheimer's disease; image genetic; joint depth semi-supervised nonnegative matrix decomposition; machine learning; biomarker

1. Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disease with an incidence that is increasing yearly. The major clinical finding of AD is the accumulation of amyloid β (Aβ) and hyperphosphorylated tau tangles in the brain. Due to its complex pathogenesis, no effective treatment is available. The study of abnormal brain regions and genetic variants in AD is therefore necessary. With the rise and development of machine learning, many researchers have applied it to biomarker mining and early diagnosis of AD [1]. Imaging genetics seeks to associate genetic variants with structural and functional changes in the human brain through the joint analysis of imaging data and genetic data [2].

Machine learning methods have been widely used in AD data analysis, although there is a need for more partial modal data in AD analysis. Hu *et al*. [3] proposed an effective data augmentation method using generative adversarial networks to reconstruct missing positron emission tomography (PET) images in order to address the class imbalance challenge. Yu *et al*. [4] proposed a new multi-directional perceptual generative adversarial network (MP-GAN). This method delineates subtle lesions through magnetic resonance (MR) image transformation between source and predefined target domains. It is used to visualize morphological features indicative of AD severity in patients at different stages of the disease [4]. Based on the existing sliding window correlation test, Jo T *et al*. [5] proposed a cyclic sliding window correlation test method using a three-step approach (feature correlation analysis, feature selection, and classification) in order to improve the prediction accuracy of AD using serum-based metabolomics classification. Lee *et al*. [6] proposed a novel convolutional neural network model for interpolating tau PET images from more widely available cross-modal imaging inputs. This model can effectively improve the accuracy of AD classification [6].

A previous study has shown that joint nonnegative matrix decomposition is a robust algorithm for association analysis. Wang and others proposed a group sparse joint nonnegative matrix decomposition (GSJNMF) algorithm integrating single nucleotide polymorphism (SNP), functional magnetic resonance imaging (fMRI), and DNA methylation data for schizophrenia (SZ) [7]. The method incorporated the structural information of the three integrative findings based on the joint nonnegative matrix decomposition. The genetic data in the modules obtained by the algorithm were significantly correlated with the activity of several at-risk brain regions (including the insula, lingual gyrus, fusiform gyrus, postcentral gyrus, supramarginal gyrus, superior temporal gyrus, superior temporal pole, and lobule VI of the cerebellar hemisphere). Peng

et al. [8] added the introduction of orthogonal constraints on the basis matrix to discard insignificant features in the rows of the coefficient matrix to the GSJNMF algorithm, resulting in improved results. Wei *et al.* [9] proposed a joint connectivity-based nonnegative matrix decomposition (JCB-SNMF) algorithm and applied it to AD imaging genetic data. The algorithm added connectivity constraints on the coefficient matrix based on joint nonnegative matrix decomposition (JNMF) to incorporate the connectivity information between brain regions and genetic data of the brain. With this algorithm, some essential pairs of imaging genetic relations in AD were found. All AD samples were used as input to the proposed algorithm during the experiment. The Pearson correlation coefficient between the original matrix and the reconstructed matrix was used as an indicator to measure the algorithm's performance for parameter selection. Specifically, all parameters were selected within the range of [0.0001, 0.001, 0.01, 0.1, 1, 10] to evaluate the changes in the Pearson correlation coefficient of the algorithm under different parameter combinations. Finally, the parameter combination that maximized the Pearson correlation coefficient was selected as the final parameter.

Although the above algorithms incorporate a variety of prior information, they only consider the feature matrix's linear features and cannot capture its nonlinear features. To this regard, our study applied a joint depth semi-supervised non-negative matrix decomposition (JDSNMF) algorithm to integrate structural magnetic resonance imaging (sMRI), gene expression, and SNP data of AD. The top biomarkers mined by the algorithm are expected to provide a reference for the diagnosis and treatment of AD.

2. Method

2.1 Joint Depth Semi-Supervised Nonnegative Matrix Decomposition

We used $\mathrm{X_i} \in \mathrm{R^{n\times p_i}} (i = 1, 2, 3)$ to represent the region of interest (ROI), SNP, and gene expression data, respectively. n is the total number of samples and $\mathrm{p_i} (i = 1, 2, 3)$ represents the number of features of the three data, respectively. $\mathrm{X_i}$ can be decomposed into a basis matrix $\mathrm{W} \in \mathrm{R^{n\times k}}$ and three feature matrices $\mathrm{H_i} \in \mathrm{R^{k\times p_i}} (i = 1, 2, 3)$. k is the number of dimensionality reduction. JDSNMF uses $\mathrm{H_i}$ multilayer neural networks to capture nonlinear features based on the JDSNMF algorithm. The JDSNMF algorithm also controls the growth of the decomposed matrices by the F-parameter with the objective function shown below.

$$\begin{aligned} &\min \sum_{i=1}^{I} \|X_i - W H_{i_0}\|_F^2 + \lambda \|M\|'_F \\ &\text{s.t. } H_{i_0} = s\left(Z_{i_1} s\left(Z_{i_2} \ldots s\left(Z_{i_N} H_{i_N}\right)\right)\right) \\ &H_{i_{n-1}} = s\left(Z_{i_n} H_{i_n}\right), H_{i_0} \ldots H_{i_{N-1}} \geq 0, \\ &M \in \left\{U, Z_{i_1}, \ldots, Z_{i_N}, H_{i_N}\right\}, n = 1, \ldots, N \end{aligned} \tag{1}$$

$\mathrm{W} \in \mathbb{R}^{\mathrm{n}\times \mathrm{k_0}}$ is known as the sample potential matrix. $\mathrm{H_{i_0}} \in \mathbb{R}^{\mathrm{k_0}\times \mathrm{p_i}}$ is referred to as the feature potential matrix, which incorporates the first layer of the network. $\mathrm{H_{i_n}} \in \mathbb{R}^{\mathrm{k_n}\times \mathrm{p_i}}$ is called the characteristic potential feature of the $\mathrm{n}+1$ layer. $s(x)$ represents the activation function. The expression is as follows if it is a sigmoid activation function.

$$s(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

If $s(x)$ is a tanh activation function, the expression is as follows.

$$\tanh(x) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{3}$$

If $s(x)$ is the Rectified Linear Unit (ReLU) activation function, the expression can be defined as Eqn. 4.

$$s(x) = \begin{cases} x & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0 \end{cases} \tag{4}$$

According to the value of k, the algorithm will generate k co-expression modules by using H_{i_0} $(i = 1, 2, 3)$. We used z score to standardize each line of H_{i_0}, and then compared it with the artificially set threshold value T. If the normalized value was greater than the T value, it was considered that the index feature corresponding to this value was qualified to enter the co-expression module. In this study, we set the value of T to 1.

2.2 Feature Selection

In the feature selection, we used the scikit-learn package for Python (v3.7, Python Software Foundation, Portland, OR, USA) to achieve the weight assignment to ROI, SNP, and genes in the co-expression module using the random forest (RF) algorithm. The parameters were explicitly set: 'n_estimators' was selected between 100 and 600, and 'criterion' was identified between 'gini' and 'entropy'.

2.3 Diagnostic Model Construction

We performed a five-fold cross-validation on the training set to construct a diagnostic model using the GridSearchCV function. Finally, the optimal parameters were 'entropy' for 'criterion' and 500 for 'n_estimators'. In addition, we constructed diagnostic models for ROI, SNP, and genes based on the logistic regression (LR) algorithm utilizing IBM SPSS Statistics 26 software (IBM SPSS Statistics, Chicago, IL, USA).

2.4 Visualization of Top Brain Regions

The BrainNet Viewer package of Matlab 2018a software (https://www.nitrc.org/projects/bnv/) was used to visualize the important brain regions selected by the JDSNMF algorithm.

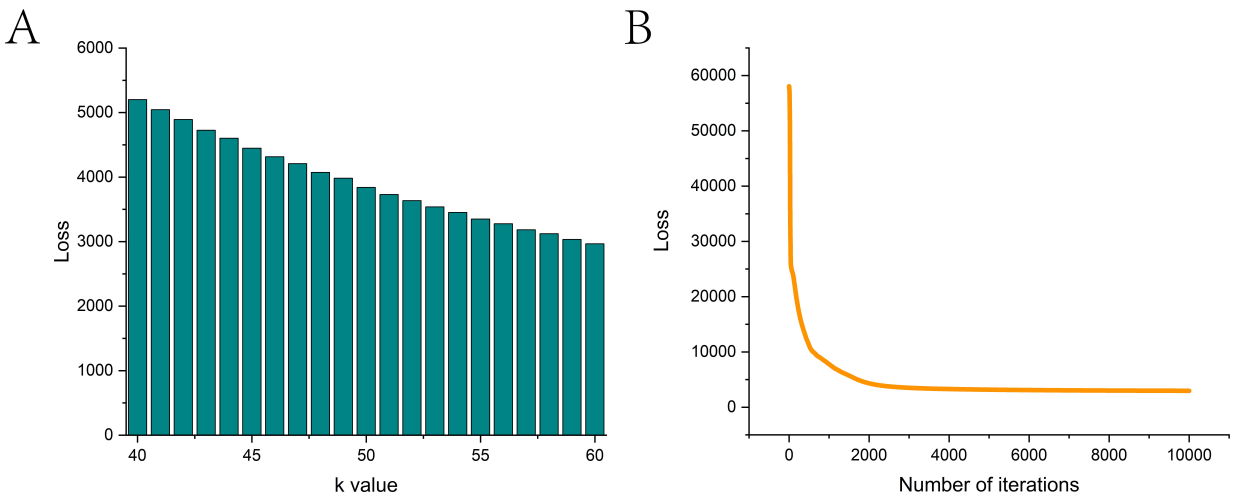

Fig. 1. The process of parameter selection. (A) Histogram showing the loss at values between 40 and 60. (B) Line graph showing the variation of the loss for the best combination of parameters.

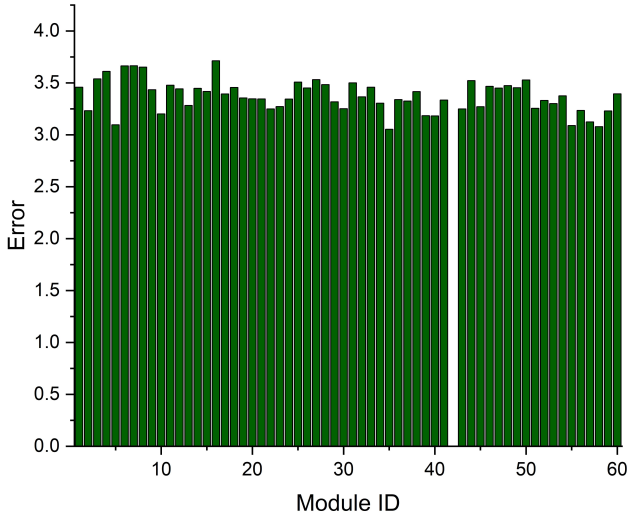

Fig. 2. Histogram showing the reconstruction error of co-expression modules.

Table 1. Sample set information.

Category	Quantity	Age (mean ± SD)	Sex (M/F)
AD	30	70.80 ± 5.34	17/13
MCI	100	71.66 ± 3.22	40/60
HC	50	70.96 ± 6.21	30/20

SD, Standard Deviation; M, male; F, female; AD, Alzheimer's disease; MCI, mild cognitive impairment; HC, healthy control.

Table 2. *p* values for age and sex for different group samples.

Category	AD vs MCI	AD vs HC	MCI vs HC
Age	0.032	0.034	0.23
Sex	0.107	0.769	0.021

2.5 Enrichment Analysis

The R package "clusterProfiler" (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) was applied to perform the Kyoto Encyclopedia of Genomes (KEGG) and Gene Ontology (GO) enrichment analysis of the important genes in the module. Bubble plots were then visualized using the R package "ggplot2" (v3.4.4, https://cran.r-project.org/web/packages/ggplot2/index.html).

3. Results

3.1 Data Collection and Pre-Processing

AD, mild cognitive impairment (MCI), and healthy control (HC) samples were downloaded from the The Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The information on their sample set is shown in Table 1. For sMRI, we aligned the data to the Montreal Neurological Institute (MNI) standard space using the statistical parametric mapping (SPM) toolkit of Matlab software. Correction, segmentation, and alignment were then performed and, finally, the gray matter density of 90 brain regions with the cerebellar regions removed were extracted as the sMRI ROIs. Regarding the SNP data, we used the PLINK tool (https://www.cog-genomics.org/plink2/) to remove SNPs that did not meet the criteria for sex detection, Hardy-Weinberg equilibrium, and minor allele frequencies less than 0.05. The SNPs were then genetically annotated using ANNOVAR (http://annovar.openbioinformatics.org/), and 2378 SNPs within ±5000 base pairs of multiple AD risk genes were extracted as input genetic data for the model. For the gene expression data, we obtained 414 genes differentially expressed in the HC and diseased groups using the limma algorithm. In addition, the differences in age and sex among the samples were explored based on the Wilcoxon test and Chi-squared test, respectively. The statistical results are presented in Table 2.

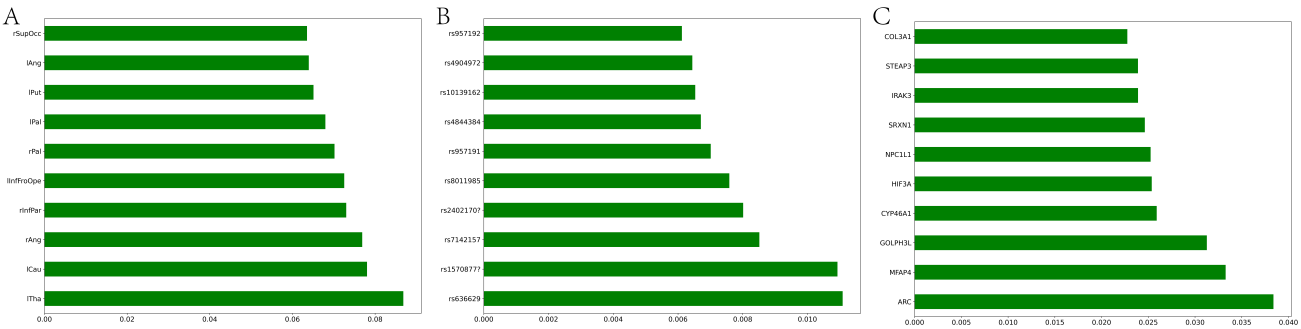

Fig. 3. Heat maps showing weights of top features. (A–C) Weighted histograms of the top 10 regions of interest, top 10 single nucleotide polymorphisms, and top 10 genes, respectively.

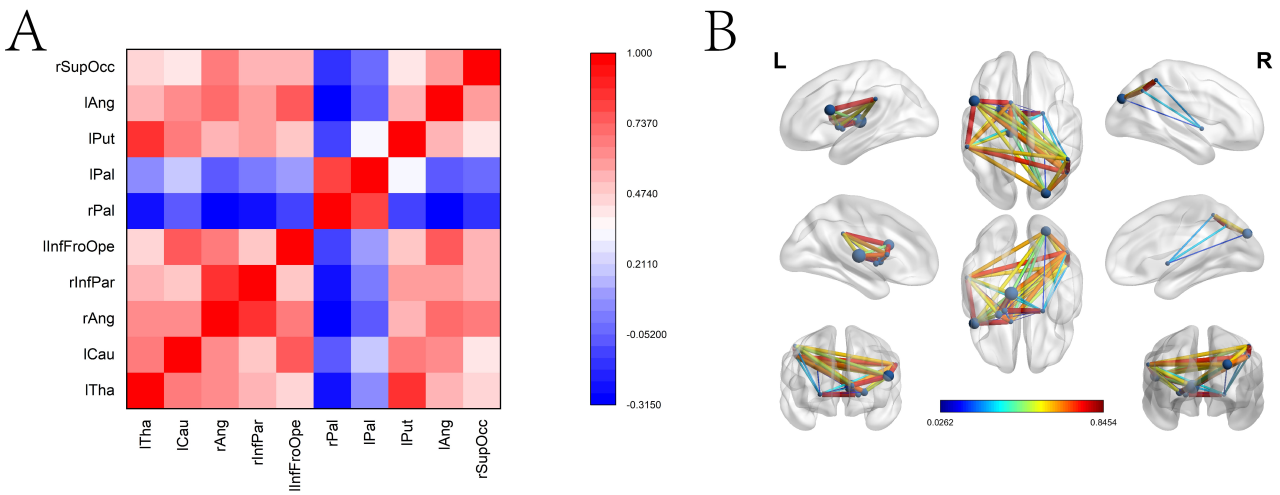

Fig. 4. Correlation analysis and visualization of the top 10 ROIs. (A) Heat map showing the correlation between the top 10 ROIs. (B) Visualization of the top 10 ROIs and the relationship between them on the brain template. ROIs, region of interests.

3.2 Parameter Selection

Since the JDSNMF algorithm was unsupervised, only the AD and MCI samples were used as input to the algorithm. The hyperparameters to be selected for the JDSNMF algorithm included the activation function, learning rate, and the number of dimensionality reduction. We set the number of iterations to 10,000. After fixing the other parameters, we first selected the activation function. The losses using the tanh, sigmoid, and rectified linear unit (ReLU) functions were 21,679.373, 4781.157, and 5563.1367, respectively. Therefore, we chose the sigmoid function for the subsequent analysis. Next, we selected the learning rate from the range of [0.1, 0.01, 0.001], with losses of 30,049, 4835, and 4781, respectively. Consequently, we set the learning rate to 0.001. Finally, we selected k. We selected it to be from 40 to 60. The changing loss with value trend is shown in Fig. 1A. This shows that loss takes the minimum value when it equals 60. Fig. 1B shows the line plot of the loss variation for the best combination of parameters. It can be seen that the loss tends to stabilize after iterating to 2000 for loss variation.

3.3 Co-Expression Module Selection

Since we set k to 60, we obtained 60 co-expression modules. Module 42 did not contain any ROIs and was excluded. We calculated the reconstruction error for the other modules (Fig. 2). Among them, module 35 had the smallest reconstruction error. This module had 15 ROIs, 62 genes, and 490 SNPs.

3.4 Significant Feature Screening and Diagnostic Model Construction

In this section, we ranked the feature importance of ROIs, genes, and SNPs in module 35 based on the RF algorithm. The histograms of the top 10 feature weights are given in Fig. 3A–C, respectively. The diagnostic performance of these top markers was then explored, and diagnostic models were constructed using these ROIs, genes, and SNPs, respectively.

3.5 Visualization and Biological Significance Analysis of Top Features

We analyzed the correlation of the top 10 ROIs as shown in Fig. 4. Fig. 4A is a heat map drawn from the Pearson correlation coefficients among the 10 ROIs. Fig. 4B

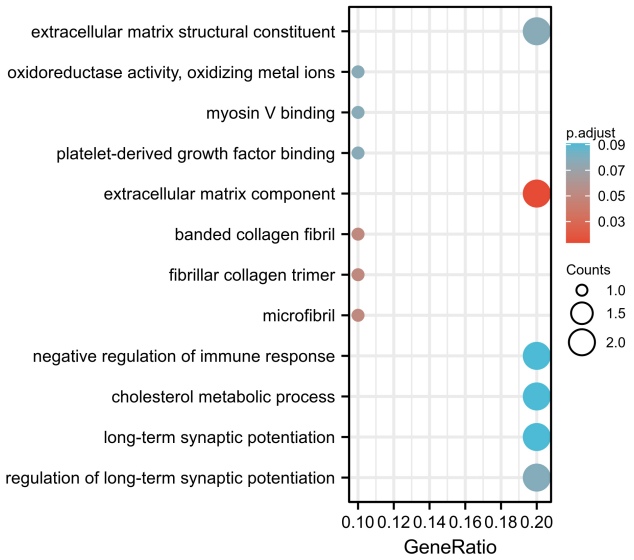

Fig. 5. Results of Gene Ontology and Kyoto Encyclopedia of Genomes enrichment analysis of the top 10 genes.

shows the connectivity of these ROIs in the brain template. There was a maximum positive correlation between rInfPar and rAng (corr = 0.9533), except for the correlation between each brain region and itself, which is 1. The maximum negative correlation can be found between rPal and rPal (corr = –0.9679). For the top 10 genes, Fig. 5 displays the results of their GO and KEGG enrichment analysis. We aimed to explore the biological significance of the top 10 ROIs, the top 10 genes, and the biological pathways in which they are involved. In addition, to confirm the algorithm's performance for association analysis, we plotted the correlation heat map of the top ROIs and top SNPs (Fig. 6A) and the correlation heat map of the top ROIs and top genes (Fig. 6B). For Fig. 6A, rs4844384 and lPal had the highest positive correlation (corr = 0.5754). rs957191 and lPut had the highest negative correlation (corr = –0.4794). As shown in Fig. 6B, sulfiredoxin 1 (SRXN1) and lTha had the highest positive correlation (corr = 0.4962). collagen type III alpha 1 chain (COL3A1) and Iput had the highest negative correlation.

3.6 Construction of Diagnostic Models Based on Top Markers

In order to construct a diagnostic model on top markers, we constructed a diagnostic model for AD using the top 10 ROIs, top 10 SNPs, and top 10 genes based on the LR algorithm, respectively. Fig. 7A–C shows the receiver operating characteristic (ROC) curves of the three diagnostic models. The top 10 genes in the test set area under the curve (AUC) can reach the maximum (AUC = 0.947). The respective ROC curves of the top 10 ROIs, SNPs, and genes are presented in Figs. 8,9,10. As can be seen from the figures, the majority of top markers have AUCs greater than 0.5.

Table 3. Algorithm performance comparison.

Algorithm	$Corr\,(\mathbf{X_1},\ \mathbf{WH_1})$	$Corr\,(\mathbf{X_2},\ \mathbf{WH_2})$	$Corr\,(\mathbf{X_3},\ \mathbf{WH_3})$
JNMF	0.7729	0.8387	0.7227
JCB-SNMF	0.7740	0.8385	0.7218
JDSNMF	0.9012	0.8391	0.9676

JNMF, joint nonnegative matrix decomposition; JCB-SNMF, joint connectivity-based nonnegative matrix decomposition; JDSNMF, joint depth semi-supervised nonnegative matrix decomposition.

3.7 Comparison with Other Algorithms

We compared the performance of the JDSNMF algorithm with the JNMF algorithm and the JCB-SNMF algorithm for correlation analysis (Table 3). Specifically, we introduced the Pearson correlation coefficients of the original and reconstructed matrices for comparison to achieve similarity between the original matrix and the reconstructed matrix after decomposition.

$Corr\,(X_i,\ WH_i)\ (i = 1, 2, 3)$ in the table represents the Pearson correlation coefficients between the original expression matrix and the reconstructed matrix for ROIs, genes, and SNPs, respectively. We found that the JDSNMF algorithm had the best performance.

4. Discussion

AD is a severe neurodegenerative disease that imposes a heavy burden on both families and society. Biomarker mining of AD can assist in relevant drug development and therapeutic target discovery. To this end, this study explored imaging genetic biomarkers of AD using the JDSNMF algorithm. Specifically, we integrated sMRI, SNP, and genetic data of AD using the JDSNMF algorithm. The algorithm adequately captured the non-linear features of the three sets of data. This module contained 15 ROIs, 42 SNP loci, and 49 genes. We ranked the feature importance of each of the three sets of data based on the RF algorithm and finally retained the top 10 ROIs, SNPs, and genes, respectively.

This study determined the top 10 brain regions (lCau, Right Angular gyrus (RANG), Right Inferior parietal lobule (RINFPAR), Left Inferior Frontal Gyrus (Linffroope), Right Pars orbitalis (RPAL), Left Pars orbitalis (LPAL), Left Putamen (LPUT), Language area (Lang), Right Superior occipital gyrus (RSUPOCC) using the algorithm JDSNMF. Indifference is a common neuropsychiatric symptom in AD patients. David *et al.* [10] indicated that dopaminergic dysfunction in the left caudate nucleus was related to atrophy of the left caudate nucleus. Udo *et al.* [11] found that the blood pressure in the left caudate nucleus was negatively correlated with the Indifference Assessment Scale-Japanese Version (AES-I-J) score in a study exploring whether dopaminergic activity was related to the development of AD apathy. The angular gyrus is the visual language center (reading center) and its activities are related to memory retrieval and formation, perceptual at-

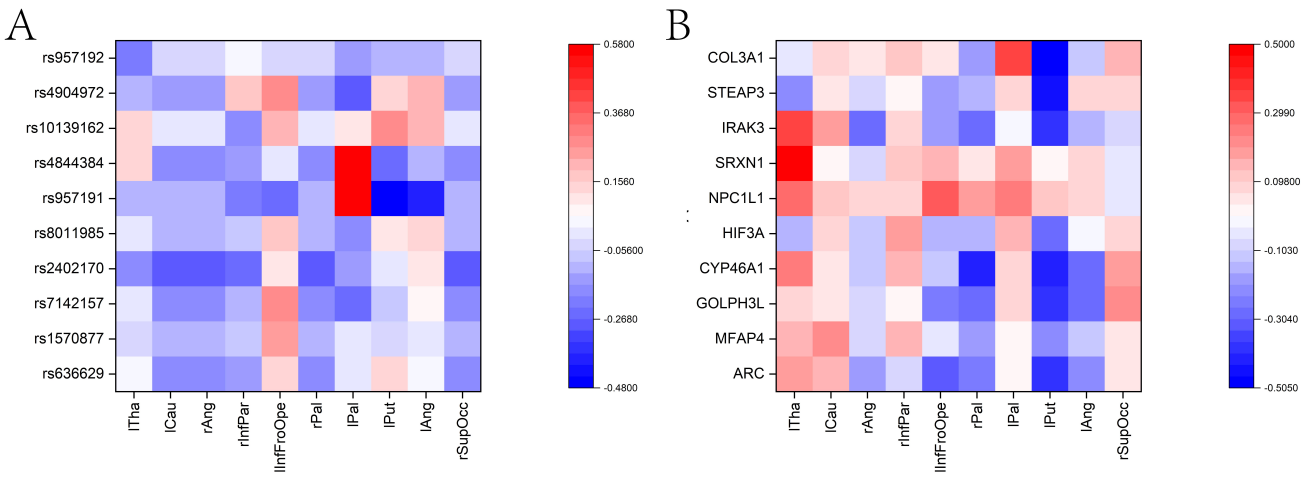

Fig. 6. Heat maps showing the correlation between the top 10 ROIs and the top 10 SNPs and top 10 genes. (A) Heat map showing the correlation between the top 10 ROIs and the top 10 SNPs. (B) Heat map showing the correlation between the top 10 ROIs and the top 10 genes. SNPs, single nucleotide polymorphism.

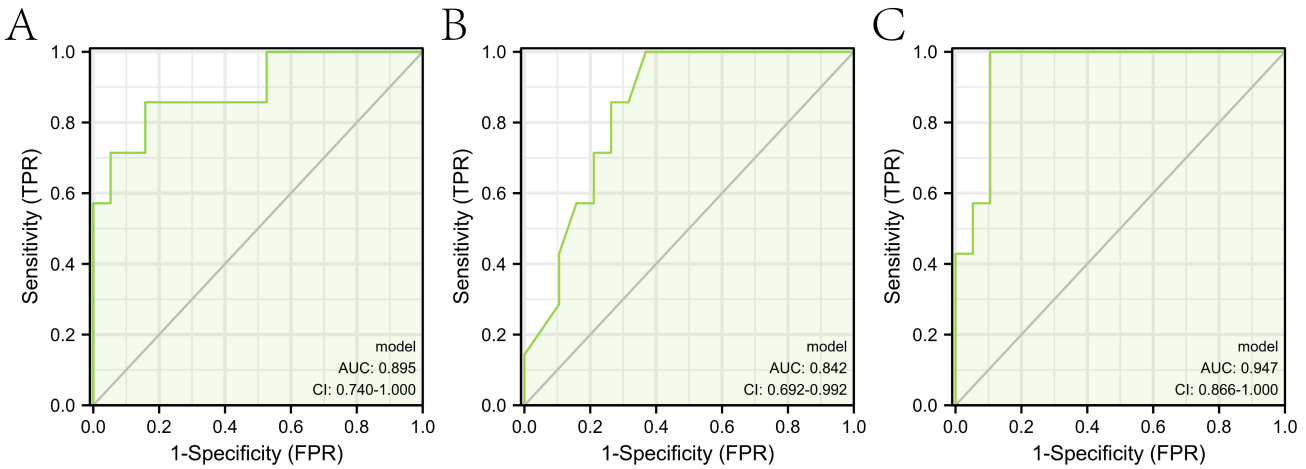

Fig. 7. Diagnostic model construction based on top markers. (A–C) ROC curves of the top 10 ROIs, top 10 SNPs, and top 10 genes, respectively. ROC, receiver operating characteristic; TPR, true positive rate; FPR, false positive rate.

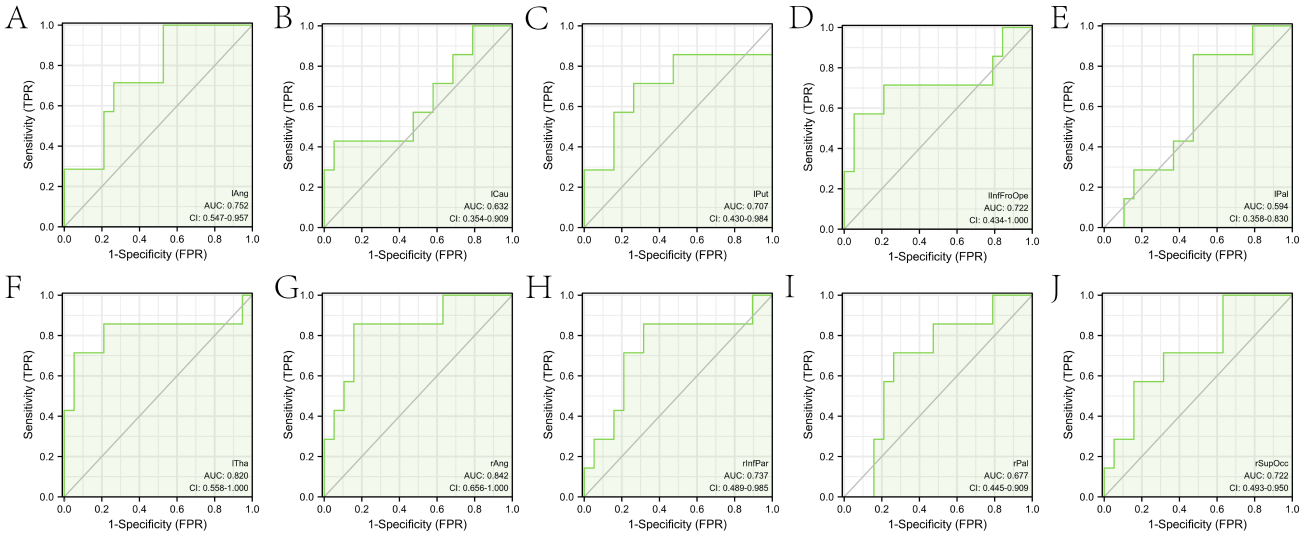

Fig. 8. Diagnostic performance validation of the top 10 ROIs. (A–J) ROC curves of the top 10 ROIs, respectively.

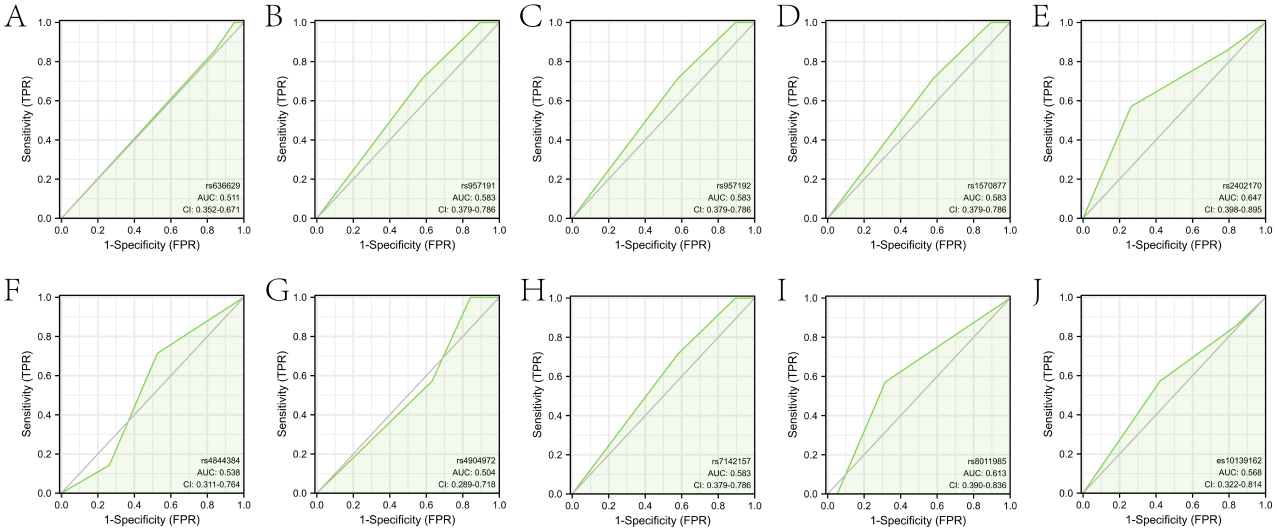

Fig. 9. Diagnostic performance validation of the top 10 SNPs. (A–J) ROC curves of the top 10 SNPs, respectively.

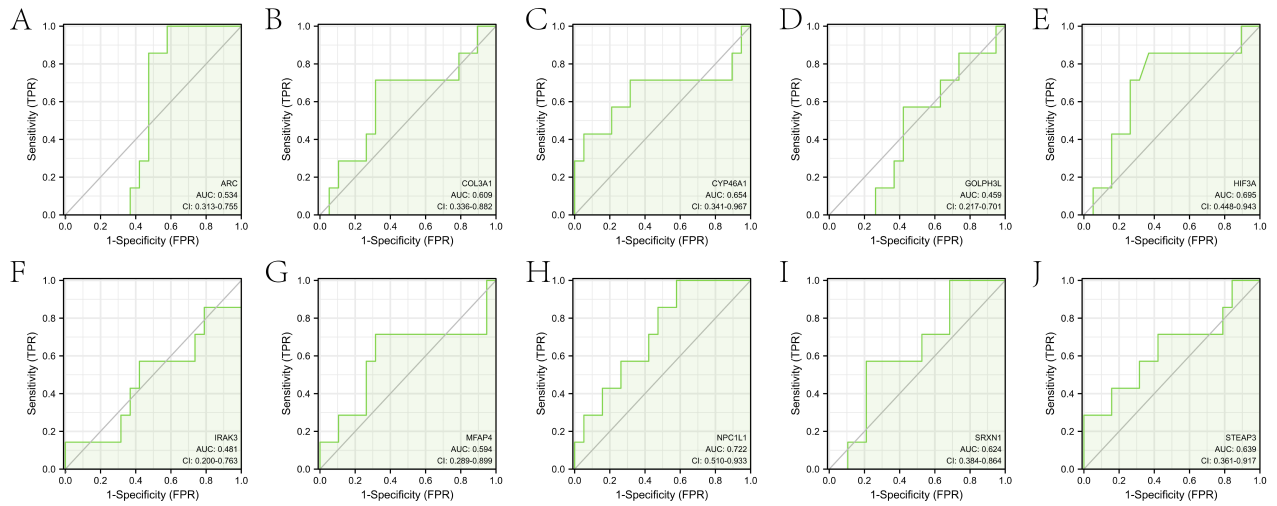

Fig. 10. Diagnostic performance validation of the top 10 genes. (A–J) ROC curves of the top 10 genes, respectively.

tention, decision-making, and manipulation [12]. Gaubert *et al*. [13] stated that the angular gyrus showed a significant metabolic decline in AD patients, and its dysfunction was related to cognitive impairment. The relationship between the globus pallidus and motor symptoms is closer than that of cognitive impairment. In the research on AD and normal aging over the last 20 years, Pini *et al*. [14] found that only one study mentioned slight morphological changes in the globus pallidus in AD.

We also determined that activity regulated cytoskeleton associated protein (*ARC*), golgi phosphoprotein 3 like (*GOLPH3L*), cytochrome P450 family 46 subfamily A member 1 (*CYP46A1*), NPC1 like intracellular cholesterol transporter 1 (*NPC1L1*), sulfiredoxin 1 (*SRXN1*), and interleukin 1 receptor associated kinase 3 (*IRAK3*) of the top 10 genes directly or indirectly participate in the pathological process of AD. *ARC* (Activity Regulatory Cytoskeleton Related Protein) is a protein-coding gene that plays a vital role in synaptic plasticity, learning, memory, and Aβ production. Landgren *et al*. [15] proposed that a polymorphism in the *ARC* gene is related to a reduced risk of developing AD. Bi *et al*. [16] also confirmed the genetic association between the *ARC* gene and AD through their research. The Golgi apparatus is the central organelle of the secretory pathway and it is also an essential organelle for post-translational modification, sorting, and transportation of membrane and secretory proteins. GOLPH3L may play a regulatory role in Golgi transport. Studies have shown that Golgi defects may lead to impaired neuronal function, and there is a particular relationship between AD and Golgi defects [17,18]. CYP46A1 is a specific enzyme of the central nervous system which is closely related to AD. Polymorphisms in the *CYP46A1* and apolipoprotein E (*APOE*) genes are a risk factor for AD [19]. In addition, an imbalance in cholesterol metabolism in the brain can affect AD development. Lerner *et al*. [20] discovered that low-dose

efavirenz can enhance the brain cholesterol metabolism of early-stage AD patients [20,21]. *NPC1L1* also plays a significant role in cholesterol homeostasis [22]. Oxidative stress leads to protein oxidation related to brain diseases such as AD. Li *et al*. [23] proposed that SRXN1 can be used as an intervention target for oxidative stress neuroprotection. Nho *et al*. [24] found that IRAK3, an interleukin-1 receptor-related kinase, was significantly dysregulated in late-onset AD. Another study has confirmed that some interleukin-1 receptor-related kinases are related to the pro-inflammatory process in AD, and these molecules are significant in AD [25].

Among the pathways enriched by the top 10 genes, some genes have been confirmed to be risk genes for AD. Corsi *et al*. [26] identified and characterized the functional characterization and performed pathway analysis of two fAD mutations in the presenilin-79 (*PSEN150*) gene, revealing profound expression changes in extracellular matrix components that are useful to help elucidate the affected cellular mechanism in AD neurons. Studies have shown that AD may be the pathological consequence of an aging immune system [27]. In addition, inflammation is a significant physiological immune response, and some essential proteins can promote the clearance of inflammatory mediators to participate in the immune response and play a role in the release of Aβ [28]. A review by Khan *et al*. [29] reported that neuraminidase 1 (NEU1) regulates elastic fiber assembly and is related to the pathology of AD. We identified regulatory pathways for long-term synaptic potentiation. Wang *et al*. [30] identified PYGM as a synaptic plasticity regulator involved in AD, which plays a crucial role in protrusion enhancement. We also identified lipid metabolism-related pathways. There is a strong correlation between hypercholesterolemia and AD, although cholesterol cannot cross the blood–brain barrier (BBB) and enter the brain. Unlike cholesterol, oxidized cholesterol metabolites known as oxysterols can cross the BBB from the circulatory system. The major oxysterols present in circulation are 24S-hydroxycholesterol and 27-hydroxycholesterol. Loera-Valencia *et al*. [31] reviewed the relevant evidence on their impact on AD progression. The β-amyloid fibrils constituting the core of AD brain senile plaques are microfiber-like structures, the critical formation pathway of which has been identified [32]. Platelet-derived growth factor binding has not been proven to play a role in AD, although Li *et al*. [33] explored its impact on the pathogenesis and treatment of Parkinson's disease. The antioxidant enzyme NQO1 plays a vital role in controlling cellular redox status. Du *et al*. [34] found that *NQO1* regulates the expression and alternative splicing of AD-related apoptosis genes in PC12 cells. In addition, the KEGG pathway (Ferroptosis) that we identified has also been confirmed to play a role in AD development. Bao *et al*. [35] found that the loss of ferroportin induces memory impairment by promoting ferroptosis in AD.

There are no data that support the correlation between the top 10 SNPs that were identified in this study and AD. What role these SNPs play in the pathological process of AD requires exploration in future studies. Finally, we built a diagnostic model based on these top biomarkers and explored their diagnostic performance. These biomarkers might be useful for the future diagnosis and treatment of AD.

5. Conclusions

In this study we analyzed the imaging genetic data of AD in detail using the JDSNMF algorithm and mined several biologically significant pathways for diagnosing AD. Additionally, multiple strongly correlated ROI-SNP pairs, as well as ROI-gene pairs, were identified. However, due to the imbalance of AD samples, in future studies we will introduce sample adoption strategies to mitigate the estimation bias caused by sample imbalance. The potential AD-related markers and association patterns identified remain to be validated by further experimental work.

Abbreviations

lCau, Left Caudate nucleus; RANG, Right Angular gyrus; RINFPAR, Right Inferior parietal lobule; Iinffroope, Left Inferior Frontal Gyrus; RPAL, Right Pars orbitalis; LPAL, Left Pars orbitalis; LPUT, Left Putamen; Lang, Language area; RSUPOCC, Right Superior occipital gyrus.

Availability of Data and Materials

The data used in this paper came from the ADNI database (https://adni.loni.usc.edu/).

Author Contributions

YW: Conceptualization, Data curation, Software, Visualization, Writing—original draft. XW: Methodology, Writing—review & editing. Both authors contributed to editorial changes in the manuscript. Both authors read and approved the final manuscript. Both authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

We thank the anonymous reviewers for their constructive comments which have helped improve the manuscript.

Funding

This research received no external funding.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Tan CC, Yu JT, Tan L. Biomarkers for preclinical Alzheimer's disease. Journal of Alzheimer's Disease. 2014; 42: 1051–1069.

[2] Wu L, Gilyazova N, Ervin JF, Wang SHJ, Xu B. Site-Specific Phospho-Tau Aggregation-Based Biomarker Discovery for AD Diagnosis and Differentiation. ACS Chemical Neuroscience. 2022; 13: 3281–3290.

[3] Hu S, Yu W, Chen Z, Wang S. 2020 IEEE 6th International Conference on Computer and Communications (ICCC) (pp. 1323–1327). IEEE: Chengdu, China. 2020.

[4] Yu W, Lei B, Wang S, Liu Y, Feng Z, Hu Y, *et al*. Morphological Feature Visualization of Alzheimer's Disease via Multidirectional Perception GAN. IEEE Transactions on Neural Networks and Learning Systems. 2023; 34: 4401–4415.

[5] Jo T, Kim J, Bice P, Huynh K, Wang T, Arnold M, *et al*. Circular-SWAT for deep learning based diagnostic classification of Alzheimer's disease: application to metabolome data. eBioMedicine. 2023; 97: 104820.

[6] Lee J, Burkett BJ, Min HK, Senjem ML, Dicks E, Corriveau-Lecavalier N, *et al*. Synthesizing images of tau pathology from cross-modal neuroimaging using deep learning. Brain. 2023; awad346.

[7] Wang M, Huang TZ, Fang J, Calhoun VD, Wang YP. Integration of Imaging (epi)Genomics Data for the Study of Schizophrenia Using Group Sparse Joint Nonnegative Matrix Factorization. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2020; 17: 1671–1681.

[8] Peng P, Zhang Y, Ju Y, Wang K, Li G, Calhoun VD, *et al*. Group Sparse Joint Non-Negative Matrix Factorization on Orthogonal Subspace for Multi-Modal Imaging Genetics Data Analysis. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2022; 19: 479–490.

[9] Wei K, Kong W, Wang S. Integration of Imaging Genomics Data for the Study of Alzheimer's Disease Using Joint-Connectivity-Based Sparse Nonnegative Matrix Factorization. Journal of Molecular Neuroscience. 2022; 72: 255–272.

[10] David R, Koulibaly M, Benoit M, Garcia R, Caci H, Darcourt J, *et al*. Striatal dopamine transporter levels correlate with apathy in neurodegenerative diseases A SPECT study with partial volume effect correction. Clinical Neurology and Neurosurgery. 2008; 110: 19–24.

[11] Udo N, Hashimoto N, Toyonaga T, Isoyama T, Oyanagi Y, Narita H, *et al*. Apathy in Alzheimer's Disease Correlates with the Dopamine Transporter Level in the Caudate Nuclei. Dementia and Geriatric Cognitive Disorders Extra. 2020; 10: 86–93.

[12] Wang JX, Rogers LM, Gross EZ, Ryals AJ, Dokucu ME, Brandstatt KL, *et al*. Targeted enhancement of cortical-hippocampal brain networks and associative memory. Science. 2014; 345: 1054–1057.

[13] Gaubert M, Villain N, Landeau B, Mézenge F, Egret S, Perrotin A, et al. Neural Correlates of Self-Reference Effect in Early Alzheimer's Disease. Journal of Alzheimer's Disease. 2017; 56: 717–731.

[14] Pini L, Pievani M, Bocchetta M, Altomare D, Bosco P, Cavedo E, *et al*. Brain atrophy in Alzheimer's Disease and aging. Ageing Research Reviews. 2016; 30: 25–48.

[15] Landgren S, von Otter M, Palmér MS, Zetterström C, Nilsson S, Skoog I, *et al*. A novel ARC gene polymorphism is associated with reduced risk of Alzheimer's disease. Journal of Neural Transmission. 2012; 119: 833–842.

[16] Bi R, Kong LL, Xu M, Li GD, Zhang DF, Alzheimer's Disease Neuroimaging Initiative, *et al*. The Arc Gene Confers Genetic Susceptibility to Alzheimer's Disease in Han Chinese. Molecular Neurobiology. 2018; 55: 1217–1226.

[17] Joshi G, Wang Y. Golgi defects enhance APP amyloidogenic processing in Alzheimer's disease. BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology. 2015; 37: 240–247.

[18] Joshi G, Bekier ME, 2nd, Wang Y. Golgi fragmentation in Alzheimer's disease. Frontiers in Neuroscience. 2015; 9: 340.

[19] Li L, Zeng F, Liu YH, Li HY, Dong SY, Peng ZY, *et al*. CYP46A1 and the APOEε4 Allele Polymorphisms Correlate with the Risk of Alzheimer's Disease. Molecular Neurobiology. 2018; 55: 8179–8187.

[20] Lerner AJ, Arnold SE, Maxfield E, Koenig A, Toth ME, Fortin B, *et al*. CYP46A1 activation by low-dose efavirenz enhances brain cholesterol metabolism in subjects with early Alzheimer's disease. Alzheimer's Research & Therapy. 2022; 14: 198.

[21] Gamba P, Giannelli S, Staurenghi E, Testa G, Sottero B, Biasi F, *et al*. The Controversial Role of 24-S-Hydroxycholesterol in Alzheimer's Disease. Antioxidants. 2021; 10: 740.

[22] Yamanashi Y, Takada T, Shoda JI, Suzuki H. Novel function of Niemann-Pick C1-like 1 as a negative regulator of Niemann-Pick C2 protein. Hepatology. 2012; 55: 953–964.

[23] Li Q, Yu S, Wu J, Zou Y, Zhao Y. Sulfiredoxin-1 protects PC12 cells against oxidative stress induced by hydrogen peroxide. Journal of Neuroscience Research. 2013; 91: 861–870.

[24] Nho K, Nudelman K, Allen M, Hodges A, Kim S, Risacher SL, *et al*. Genome-wide transcriptome analysis identifies novel dysregulated genes implicated in Alzheimer's pathology. Alzheimer's & Dementia. 2020; 16: 1213–1223.

[25] Cui JG, Li YY, Zhao Y, Bhattacharjee S, Lukiw WJ. Differential regulation of interleukin-1 receptor-associated kinase-1 (IRAK-1) and IRAK-2 by microRNA-146a and NF-kappaB in stressed human astroglial cells and in Alzheimer disease. The Journal of Biological Chemistry. 2010; 285: 38951–38960.

[26] Corsi GI, Gadekar VP, Haukedal H, Doncheva NT, Anthon C, Ambardar S, *et al*. The transcriptomic landscape of neurons carrying PSEN1 mutations reveals changes in extracellular matrix components and non-coding gene expression. Neurobiology of Disease. 2023; 178: 105980.

[27] Chee SEJ, Solito E. The Impact of Ageing on the CNS Immune Response in Alzheimer's Disease. Frontiers in Immunology. 2021; 12: 738511.

[28] Karasinska JM, de Haan W, Franciosi S, Ruddle P, Fan J, Kruit JK, *et al*. ABCA1 influences neuroinflammation and neuronal death. Neurobiology of Disease. 2013; 54: 445–455.

[29] Khan A, Sergi CM. NEU1-A Unique Therapeutic Target for Alzheimer's Disease. Frontiers in Pharmacology. 2022; 13: 902259.

[30] Wang T, Zhou YQ, Wang Y, Zhang L, Zhu X, Wang XY, *et al*. Long-term potentiation-based screening identifies neuronal PYGM as a synaptic plasticity regulator participating in Alzheimer's disease. Zoological Research. 2023; 44: 867–881.

[31] Loera-Valencia R, Goikolea J, Parrado-Fernandez C, Merino-Serrais P, Maioli S. Alterations in cholesterol metabolism as a risk factor for developing Alzheimer's disease: Potential novel targets for treatment. The Journal of Steroid Biochemistry and Molecular Biology. 2019; 190: 104–114.

[32] Inoue S, Kuroiwa M, Kisilevsky R. Basement membranes, microfibrils and beta amyloid fibrillogenesis in Alzheimer's disease: high resolution ultrastructural findings. Brain Research. Brain Research Reviews. 1999; 29: 218–231.

[33] Li D, Huang LT, Zhang CP, Li Q, Wang JH. Insights Into the Role of Platelet-Derived Growth Factors: Implications for Parkinson's Disease Pathogenesis and Treatment. Frontiers in Aging Neuroscience. 2022; 14: 890509.

[34] Du Y, Liu G, Chen D, Yang J, Wang J, Sun Y, *et al*. NQO1 regulates expression and alternative splicing of apoptotic genes associated with Alzheimer's disease in PC12 cells. Brain and Behavior. 2023; 13: e2917.

[35] Bao WD, Pang P, Zhou XT, Hu F, Xiong W, Chen K, *et al*. Loss of ferroportin induces memory impairment by promoting ferroptosis in Alzheimer's disease. Cell Death and Differentiation. 2021; 28: 1548–1562.