Research article

Classification of Alzheimer's disease using robust TabNet neural networks on genetic data

Yu Jin[1,2,†], Zhe Ren[1,2,†], Wenjie Wang[1,2], Yulei Zhang[1,2], Liang Zhou[1], Xufeng Yao[1,*] and Tao Wu[1,*]

[1] College of Medical Imaging, Jiading District Central Hospital affiliated Shanghai University of Medicine and Health Sciences, Shanghai 201318, China

[2] School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

* **Correspondence:** Email: yao6636329@hotmail.com, mazuiwu88@vip.163.com.

† These authors equally contributed to this study.

Abstract: Alzheimer's disease (AD) is one of the most common neurodegenerative diseases and its onset is significantly associated with genetic factors. Being the capabilities of high specificity and accuracy, genetic testing has been considered as an important technique for AD diagnosis. In this paper, we presented an improved deep learning (DL) algorithm, namely differential genes screening TabNet (DGS-TabNet) for AD binary and multi-class classifications. For performance evaluation, our proposed approach was compared with three novel DLs of multi-layer perceptron (MLP), neural oblivious decision ensembles (NODE), TabNet as well as five classical machine learnings (MLs) including decision tree (DT), random forests (RF), gradient boosting decision tree (GBDT), light gradient boosting machine (LGBM) and support vector machine (SVM) on the public data set of gene expression omnibus (GEO). Moreover, the biological interpretability of global important genetic features implemented for AD classification was revealed by the Kyoto encyclopedia of genes and genomes (KEGG) and gene ontology (GO). The results demonstrated that our proposed DGS-TabNet achieved the best performance with an accuracy of 93.80% for binary classification, and with an accuracy of 88.27% for multi-class classification. Meanwhile, the gene pathway analyses demonstrated that there existed two most important global genetic features of AVIL and NDUFS4 and those obtained 22 feature genes were partially correlated with AD pathogenesis. It was concluded that the proposed DGS-TabNet could be used to detect AD-susceptible genes and the biological interpretability of susceptible genes also revealed the potential possibility of being AD biomarkers.

1. Introduction

Alzheimer's disease (AD) is one common progressive neurodegenerative disease that is accompanied by typical clinical symptoms including memory loss and impairment in daily communications and activities [1]. It accounts for approximately 60–80% of all dementia and the etiology of AD is still unknown. The occurrence of AD may be caused by the accumulation of specific proteins in and around neurons [2]. And it was thought to be having an estimated 70% of the risk attributable to genetic factors [3]. The AD population was expected to be 131.5 million by 2050 worldwide [4]. Once it was diagnosed, there were no effective treatments and intervened techniques. Therefore, the early diagnosis of AD has played a great role in delaying the progression of AD [5].

In clinical applications, the techniques of AD diagnosis mainly involved neuropsychological scores, cerebrospinal fluid (CSF) testing, neuroimaging examinations, and genetic tests [6]. Each approach demonstrated typical capabilities and limitations for AD diagnosis. With wide application in clinical practices, the neuropsychological scores were easily disturbed by subjective and objective factors, such as misrepresentation of family members, cross-cultural differences, etc. [7]. While CSF testing was a kind of invasive examination of damaged injury for clinical screening [8]. According to the neuroimaging examinations, positron emission tomography (PET) and magnetic resonance imaging (MRI) could present morphological, functional and metabolic imaging features. While the metabolic PET was greatly limited by availability and radiation doses [9]. MRI modalities preserved different limitations, such as low sensitivity of structural MRI (sMRI), poor stability of functional MRI (fMRI), and incomplete characterization of brain microstructure for diffusion tensor imaging (DTI) for early AD [10–12]. Especially, with the robustness of fast and accurate abilities, genetic testing has become the most potential technique for AD detection [13].

Conventional machine learning (ML) was applied to deduce the biological problems in genetic data for the AD intelligence diagnosis [14,15]. Ha [16] proposed a novel computational framework to predict miRNA-disease associations via matrix factorization with a disease similarity constraint, and got the area under curves (AUCs) of 0.9147 and 0.8905 for global and local leave-one-out cross-validation (LOOCV), respectively; then, a similarity-based matrix factorization framework was presented to identify miRNA-disease associations, yielding better AUCs of 0.9227 and 0.8952 for global and local LOOCV, respectively [17]. Once, Oriol et al. [18] used the MLs of least absolute shrinkage and selection operator (LASSO), k-nearest neighbor (KNN), support vector machine (SVM) on single nucleotide polymorphism (SNP) data to classify late-onset AD, and the SVM achieved the best performance with the AUC of 0.72. Xu et al. [19] attempted the application of SVM to analyze the protein sequences encoded by genes for AD classification with an accuracy (ACC) of 85.70%. Again, Castillo et al. [20] implemented the SVM with three genes (PSEN1, PSEN2 and APP) to classify AD and obtained the ACC around 80%. Moreover, Voyle et al. [21] used recursive feature elimination-random forest (RFE-RF) to distinguish normal controls (NC), mild cognitive impairment (MCI) and AD with an ACC of 62.70% for gene expression data. Besides, Moradi et al. [22] applied linear discriminant analysis (LDA) to analyze blood gene expression profiles and got an AUC of 0.84 for the classification of NC and AD, and 0.80 for the classification of NC and MCI. Up to now, classical ML algorithms including SVM, LDA and RF have been widely used and hold big promise for the early

detection of AD [23]. However, the tedious feature selection and limited performance of MLs still could not meet the clinical needs [24].

Recently, deep learning (DL) has been used for the intellectual diagnosis of AD [25]. Wang et al. [26] proposed an ensemble of 3D densely connected convolutional networks (3D-DenseNets) for AD and MCI diagnosis based on 3D MRI images, with the ACCs of 98.75% for AD/NC, 93.55% for AD/MCI, 98.36% for MCI/NC and 97.19% for AD/MCI/NC, respectively. Yu et al. [27] proposed a novel tensorizing generative adversarial network (GAN) with high-order pooling to assess MCI/AD based on the T1-weighted MR images, with the ACC of 89.29% for MCI/NC, 85.71% for AD/MCI, and 95.92% for AD/NC; then, a novel multidirectional perception generative adversarial network (MP-GAN) was presented to visualize the morphological features indicating the severity of AD for patients of different stages, and the MP-GAN achieved better classification performance than GAN in terms of AUC, ACC, specificity, and sensitivity [28]. Lee et al. [29] used deep neural networks (DNN) to distinguish NC and AD based on blood gene expression data, with an AUC of 0.86. Mahendran et al. [30] proposed an enhanced deep recurrent neural network (EDRNN) for DNA methylation data to classify AD and NC, and obtained an ACC of 89.40%. Park [31] proposed a deep neural network-based prediction model to assess AD and NC based on gene expression and DNA methylation data, with an average ACC of 82.30%. Due to the low ACC of existing DLs, the betterment of DL models is imperative to improve ACC for AD diagnosis [32,33].

In our study, an improved DL approach, namely differential genes screening-TabNet (DGS-TabNet), was proposed for AD binary and multi-class classifications. It was characterized by two modules of DGS and TabNet [34,35]. With the usage of DGS, the susceptible genetic features were effectively chosen to facilitate the learning rate and enhance the generalization ability for the proposed model. And the main contributions of our study are summarized as follows:

1) Via the comparison with three novel DLs of multi-layer perceptron (MLP), neural oblivious decision ensembles (NODE), TabNet as well as five classical MLs of decision trees (DT), RF, gradient boosting decision tree (GBDT), light gradient boosting machine (Light GBM) and SVM [36–43], our DGS-TabNet achieved the best performance with the ACC of 93.80% for binary classification, and with the ACC of 88.27% for multi-class classification.

2) The biological interpretability related to AD pathogenesis for outputted global important genetic features was validated by gene enrichment analyses of Kyoto encyclopedia of genes and genomes (KEGG) and gene ontology (GO) [44]. This would help to acknowledge the capability of genetic features in classification models and provide critical clues for the exploration of the biological interpretability of AD-susceptible genes.

2. Materials and methods

2.1. Subjects

Two batches of gene expression data were collected from the gene expression omnibus (GEO) datasets (http://www.ncbi.nlm.nih.gov/geo) from the AddNeuroMed consortium. GEO is an open international public repository that archives and freely distributes microarray, next-generation sequencing, and other forms of high-throughput functional genomics data. It also offers a number of web-based tools and strategies to assist users to query, analyze and visualize data. During the genetic testing, batch 1 (GSE63060) was prepared by Illumina Human HT-12 v3 Expression BeadChips, and

batch 2 (GSE63061) was processed by Illumina Human HT-12 v4 Expression BeadChips.

This study enrolled 145 patients with AD, 80 MCIs and 104 NCs for batch 1, 131 patients with AD, 101 MCIs and 126 NCs for batch 2. The detailed demographics of enrolled participants were listed in Table 1. The enrolled standard of the AD group met with the requirement of a diagnostic and statistical manual of mental disorders (DSM) [45]; MCI participants were included according to Petersen diagnostic criteria; NC cases were chosen by complying with the consortium to establish a registry for Alzheimer's disease (CERAD) [46].

**Table 1.** Demographic information of enrolled participants.

Demographic information	AD		MCI		NC	
	Batch 1	Batch 2	Batch 1	Batch 2	Batch 1	Batch 2
Number of subjects	145	131	80	101	104	126
Age (Mean ± SD)	75.40 ± 6.58	77.82 ± 6.64	74.45 ± 5.99	78.06 ± 7.40	72.37 ± 6.34	75.35 ± 6.04
Sex (Male/Female)	46/99	50/81	41/39	43/58	42/62	48/78

2.2. Preprocessing

The preprocessing of gene expression data was performed by R (version 4.0.2), with the following steps: 1) The GSE63060 and GSE63061 datasets were downloaded by the *getGEO* function in the GEOquery package [47]. 2) To exclude unannotated probes, available probes with clinical information were kept. Once a gene corresponded to multiple probes, the median of the probes was selected [48]. 3) Based on the probes with clinical information, the expression matrices were calculated by the *exprs* function on initial 16,789 gene features for batch 1 data and 24,899 gene features for batch 2 data, respectively. 4) With the alignment of the two batches, only 16,352 common probes were left in both datasets of batches 1 and 2 [49].

2.3. Differential gene screening TabNet (DGS-TabNet)

The pipeline of DGS-TabNet mainly had two modules: one is differential gene screening (DGS) and the other is TabNet, as shown in Figure 1. The first module of DGS was implemented to select the significant differential genes by the reduction of genetic expression data dimensionality; the second module of TabNet distinguished significant differential genes as susceptible genetic features for binary and multi-class classifications, and output the global important genetic features for the further biological interpretability of AD [50].
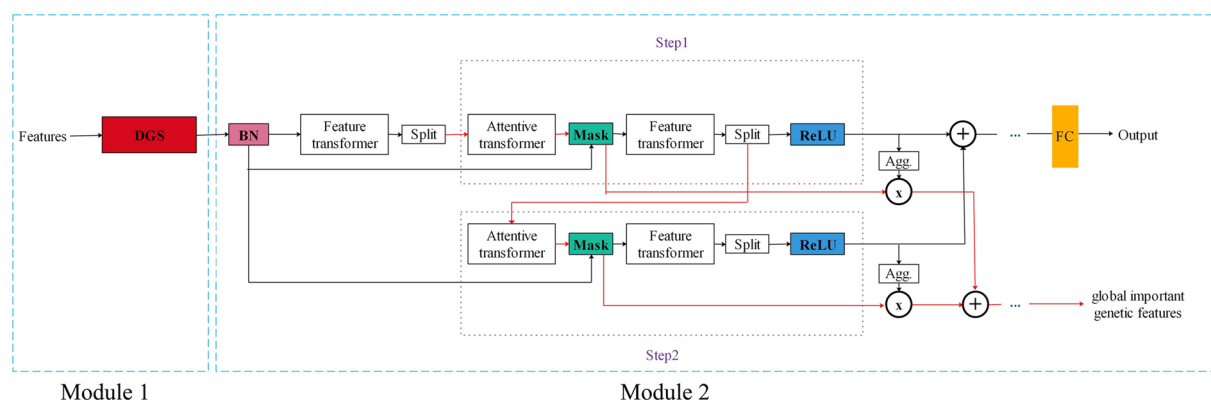
Figure 1. The pipeline of the DGS-TabNet.

2.3.1. Differential gene screening (DGS)

For expression genetics, DGS was recognized as one of the most popular methods to discover and understand the molecular pathways behind AD [51]. In our scheme, it was used to label differential genes and played a crucial role in AD classification.

It was performed by the package of limma [52] and the three main steps were listed as follows: 1) The least-squares fit was conducted by the *lmFit* function, and the comparison matrices were constructed by the *makeContrasts* function; 2) The log-fold change rate (logFC) and t-statistic were calculated by the *contrasts.fit* function; 3) The F-statistic was calculated by the *eBayes* function. The statistical information of the differential genes was summarized by the *Toptable* function for the parameters of logFC, AveExpr, P.Value, adj.P.value and the B statistic. Here, LogFC provided the logarithmic value; AveExpr provided the average log expression level of the gene; P.Value reflected the gene's false discovery rate (FDR) of the gene; adj. P.Value represented the FDR correction; the B statistic represented the logarithmic ratio of the differential expression gene. Since multiple T tests might lead to an increase of FDR, the adj.P.Value < 0.01 was chosen to maintain the false/true positive ratio.

2.3.2. TabNet

The TabNet is known as one kind of neural network for the processing of tabular data. It retains the representational automatic learning characteristics for deep neural networks and preserves the capability of providing predictive analyses. In our study, the TabNet consisted of multiple-step decision units and each step included four parts: attentive transformer, feature transformer, split and ReLU function. At the end of TabNet, the global important genetic features were outputted. The workflow of TabNet was summarized as follows:

1) The susceptible genetic features processed by DGS were inputted to the batch normalization (BN) layer for avoiding overfitting and reducing gradient explosion and gradient disappearance [53]. The feature calculation was conducted by the feature transformer layers [38] (Figure 2). It consisted of two parts: the first part was trained together on all steps, while the second part was trained separately on each step. The gated linear units (GLU) were used in the feature transformer to determine the information passed to the next layer [54]. Here, four GLU blocks (two shared and two independent blocks) were used for robust and parameter-efficient learning. Residual connections were used in the layers to maintain the network stability and to avoid the dramatic changes of variances by multiplying

$\sqrt{0.5}$ [55]. According to the scheme of the attentive transformer layer, it was calculated as:

$$M[i] = sparse\ max(P[i-1] \bullet h_i(a[i-1])) \tag{1}$$

where *a[i-1]* was the portion of the split layer divided from the previous step, $h_i(\bullet)$ represented the FC and BN layers, and *p[i]* denoted the prior scales and was formulated as:

$$P[i] = \prod_{j=1}^{i}(\gamma - M[j]) \tag{2}$$

here, the prior scales were used to indicate the application degree of one feature in the previous step.

2) The mask layer was calculated through the attentive transformer [33] (Figure 3). The final output Mask (*M*) was used to weight the input data by using the fully connected (FC) layer, the BN layer and Sparsemax layer. Where the Sparsemax layer assigned feature weights to each feature and the sum of all feature weights equaled to 1.0 [56]. To achieve better sparse results, the sparse regularization was calculated and the value of each step was factored into the overall loss, which was calculated as:

$$L_{sparse} = \sum_{i=1}^{N_{steps}} \sum_{b=1}^{B} \sum_{j=1}^{D} \frac{-M_{b,j}[i]}{N_{steps}.B} \log\left(M_{b,j}[i] + \varepsilon\right) \tag{3}$$

where ε was generally assigned a very small value. The sparser of *M* resulted in smaller *L*, and meant smaller of loss and vice versa.

3) The vectors outputted from the feature transformer layer were divided into two components by the split layer, one was used for computing the final output of the model by *ReLU* function, and another was implemented to calculate the Mask matrices for the next step.
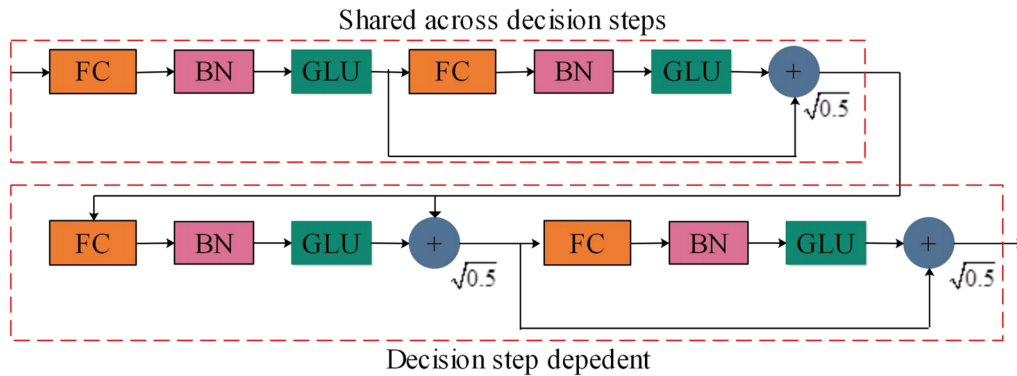
Figure 2. The scheme of feature transformer.
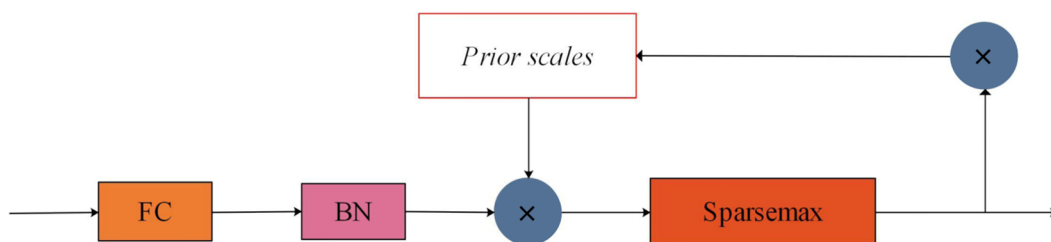
Figure 3. The scheme of attentive transformer.

2.4. *Experimental setup*

In order to validate the performance of our proposed algorithm, it was compared with three novel DLs of MLP, NODE and TabNet and five classical ML models of DT, RF, GBDT, Light GBM and SVM on the data set of GEO for binary classification (AD *vs* NC) and multi-class classification (AD *vs* MCI *vs* NC), respectively. During the performance evaluation, five indices of ACC, AUC, recall (REC), precision (PRE) and F1-Score were measured for quantitative evaluation [57]. And the interpretability of DGS-TabNet was also evaluated by enrichment analyses of KEGG and GO.

For classical MLs, it was noted that the differential genes were selected by the DGS and boruta algorithm [58]. Here, the experiments for classification were carried out on a PC with Windows 10, 64 bit, with 3.00 GHz Intel® core^{TM} i9-10980XE processor. The training and testing groups were randomly divided at a ratio of 1:1, and the repeated iterative training was carried out by 10-fold cross-validation. Here, the normalization was conducted by the *minmix* function, and automatic tuning of the model was conducted by the *tune_model* function with the iteration of 100.

Similar to MLs, the DGS was also used to screen differential genes for DLs comparison. During the stage of pre-training, the experimental hardware were 2 chips of NVIDIA RTX6000 GPU and the models of DLs were built on the pytorch framework. The training and test groups were randomly assigned at a 1:1 ratio, and the learning rate was set to 1e-3 with the maximum epoch of 200. Besides, the strategies of dropout and early stopping (patience = 5) were used to alleviate the overfitting of our models. The main hyperparameters were listed as: the width of the decision prediction layer (n_d): 32, 44, 52; the width of the attention embedding for each mask (n_a): 10, 14, 36, 38; coefficient for feature reusage in the masks (gamma): 1.0, 1.2, 1.3, 1.5; number of shared GLU at each step (n_shared): 1, 2, 5; sparsity loss coefficient (lambda_sparse): 0.0011-0.0081; optimizer: Adam; number of steps in the architecture (n_steps): 3, 4, 5; number of instances per batch (batch_size): 10. Here, the Bayesian optimization algorithm was used to tune the hyperparameters [59,60].

The biological interpretability of DGS-TabNet was conducted by the analyses of KEGG and GO via the package of clusterProfiler [61,62]. Here, the enrichment analysis of GO included three aspects of biological process (BP), cellular component (CC) and molecular function (MF). The globally important genetic features outputted by the DGS-TabNet encoder at the best classification were chosen for biological pathways on the two batches of datasets. Here, only the globally important gene features above the weight threshold of 0 were kept for biological interpretability.

3. Results

3.1. *Results of binary and multi-class classification*

The comparison of MLs and DLs for binary classification was presented in Table 2. It was clear that the DGS-TabNet exhibited the best performance with the ACC of 93.80%, AUC of 98.53%, REC of 93.80%, PRE of 93.96% and F1 of 93.79% for batch 2. Besides, the SVM proved to be the best model in the five traditional MLs. The convergence curves of training and testing were shown in Figure 4.

Table 2. Results of binary classification.

| Method	ACC		AUC		REC		PRE		F1	
	Batch 1	Batch 2	Batch 1	Batch 2	Batch 1	Batch 2	Batch 1	Batch 2	Batch 1	Batch 2
DT	73.60	87.60	76.29	86.14	**92.31**	76.19	62.34	**97.96**	74.42	85.71
RF	76.00	89.15	85.88	96.83	65.38	90.48	73.91	87.69	69.39	89.06
GBDT	75.20	90.70	86.75	97.98	76.92	90.48	67.80	90.48	72.07	90.48
LightGBM	78.40	90.70	85.91	97.47	76.92	87.30	72.73	93.22	74.77	90.16
SVM	88.00	**93.80**	94.02	98.46	86.54	93.65	84.91	93.65	85.71	93.65
MLP	76.00	70.54	82.90	74.07	76.00	70.54	76.63	71.15	76.14	70.22
NODE	75.20	76.74	87.88	84.60	75.20	76.74	77.84	76.75	73.37	76.73
TabNet	80.00	88.00	89.14	93.89	80.00	88.00	79.95	88.16	79.97	88.36
Proposed	**89.60**	**93.80**	**95.10**	**98.53**	**89.60**	**93.80**	**89.72**	**93.96**	**89.51**	**93.79**
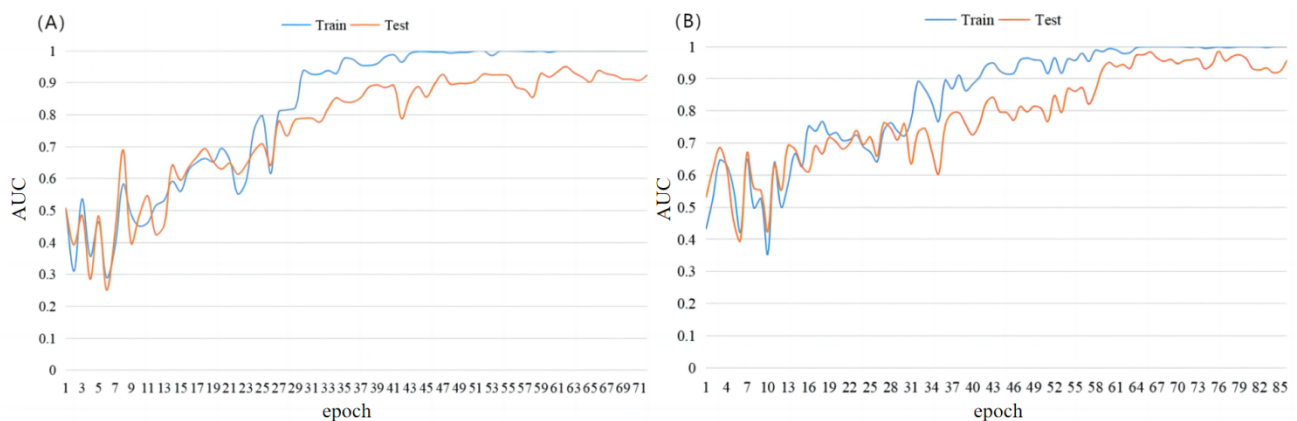
Figure 4. Convergence curves of training and testing for binary classification. (A) batch 1; (B) batch 2.

The comparison of transitional MLs and DLs for multi-class classification was displayed in Table 3. Similarly, the proposed approach of DGS-TabNet outperformed the best performance with the ACC of 88.27%, AUC of 94.97%, REC of 88.27%, PRE of 88.43% and F1 of 88.24% on batch 2. It was obvious that the performance of multi-class classification was lower than that of binary classification on both data sets of batches 1 and 2. The convergence curves of training and testing were shown in Figure 5.

Table 3. Results of multi-class classification.

Method	ACC		AUC		REC		PRE		F1	
	Batch 1	Batch 2	Batch 1	Batch 2	Batch 1	Batch 2	Batch 1	Batch 2	Batch 1	Batch 2
DT	53.33	65.92	63.85	78.44	50.23	65.04	52.09	68.55	52.43	66.47
RF	63.64	70.39	80.45	88.98	60.98	70.63	63.92	71.12	62.87	70.62
GBDT	61.21	85.47	79.42	94.53	58.05	84.97	61.54	85.72	60.59	85.53
LightGBM	65.45	81.01	82.32	94.17	63.10	80.40	64.71	81.29	64.94	81.10
SVM	72.73	78.77	87.21	92.21	72.53	78.30	72.91	79.59	72.79	79.00
MLP	55.76	46.37	74.89	60.17	55.76	46.37	58.31	33.89	55.73	33.85
NODE	56.97	54.19	78.27	70.65	56.97	54.19	61.88	52.92	54.70	53.09
TabNet	61.82	49.72	74.41	62.85	61.82	49.72	61.06	52.90	60.22	47.86
Proposed	**80.00**	**88.27**	**88.77**	**94.97**	**80.00**	**88.27**	**79.90**	**88.43**	**79.68**	**88.24**
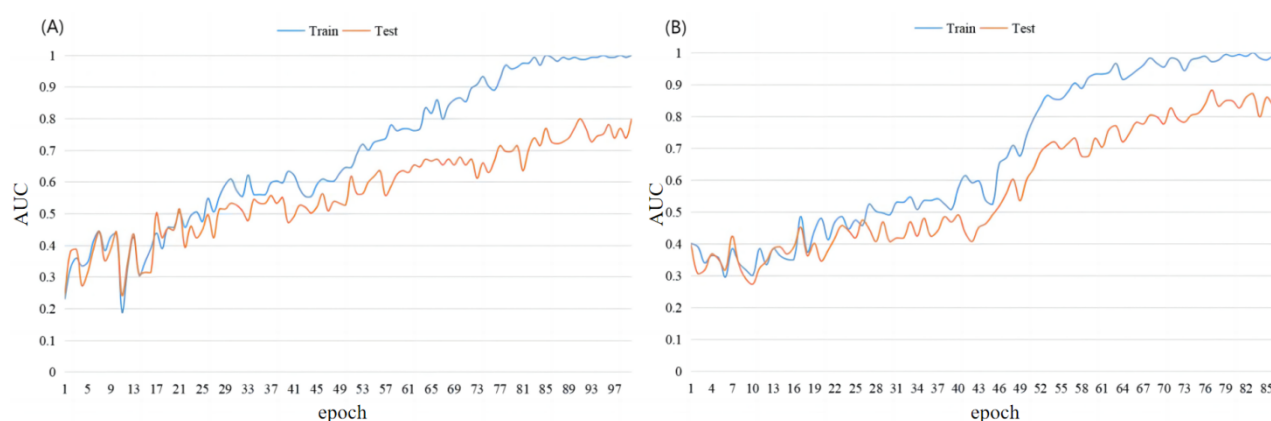
**Figure 5.** Convergence curves of training and testing for multi-class classification. (A) batch 1; (B) batch 2.

*3.2. Interpretability of the DGS-TabNet*

In Table 4, the binary classification of batch 2 showed the best performance. There were the 22 global important genetic features left, and the two largest global important feature weights of AVIL and NDUFS4 were 0.3701 and 0.3077, respectively.

The biological pathways enriched by KEGG were found to be related to 19 varied diseases including Ribosome, Coronavirus disease (COVID-19), Huntington disease, Parkinson disease, AD etc. (Figure 6). Especially, the rich factor of AD was about 0.0025 and showed weak correlation with the AD. While the GO analysis enriched a total of 48 BP, 24 CC and 17 MF for biological processes of AD (Figure 7).

Table 4. Global important genetic features.

Genetic features	Feature weights	Genetic features	Feature weights
AVIL	0.370100	LOC652864	0.005331
NDUFS4	0.307700	LOC649864	0.005196
REXO2	0.065700	CD79B	0.004798
RPS25	0.053190	WDFY2	0.004436
KCNG1	0.051161	CMTM2	0.003629
SNRPB2	0.029107	RPL3	0.003358
EGFL6	0.027134	RSL1D1	0.003124
SNORD33	0.026315	EIF3E	0.000668
MAP3K6	0.016846	TBC1D2B	0.000256
LOC653702	0.012487	KIAA1160	0.000132
HNRNPH2	0.009302	ASNSD1	0.000002
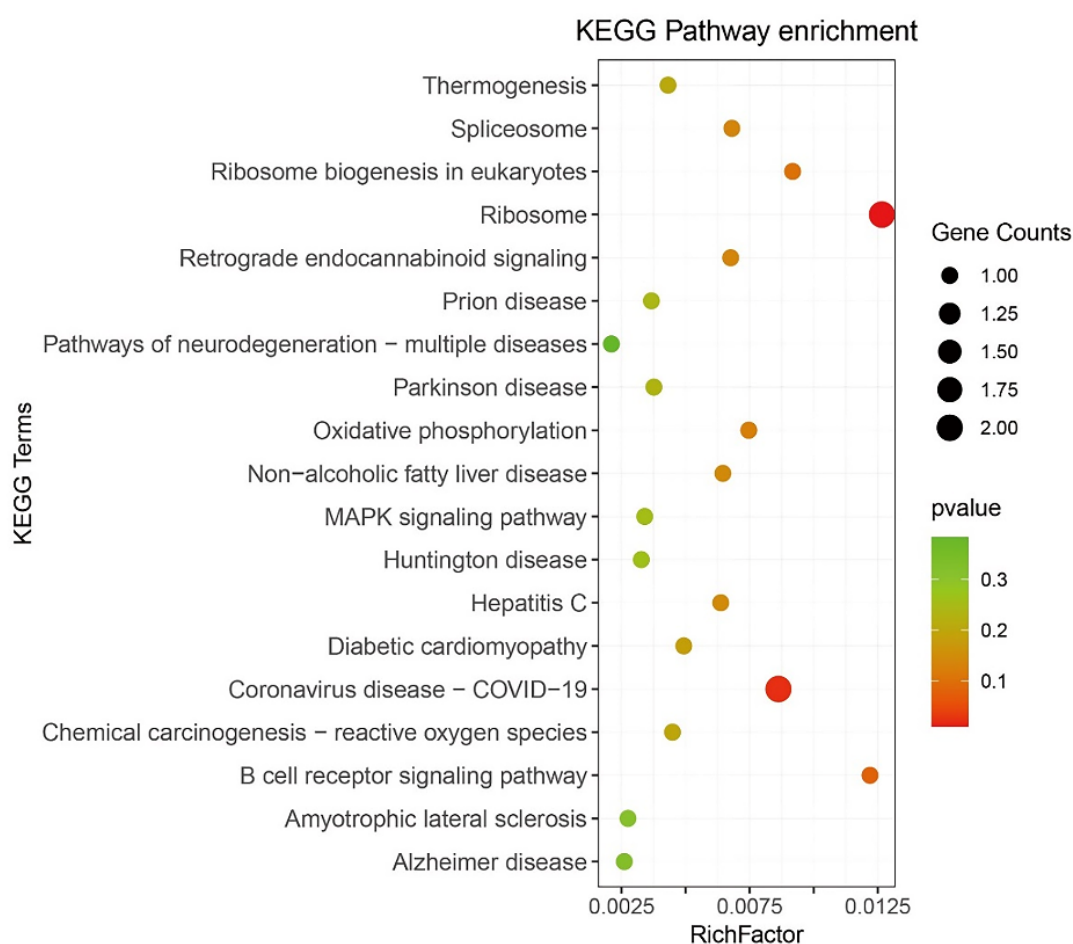
Figure 6. The KEGG enrichment analysis of global important genetic features.
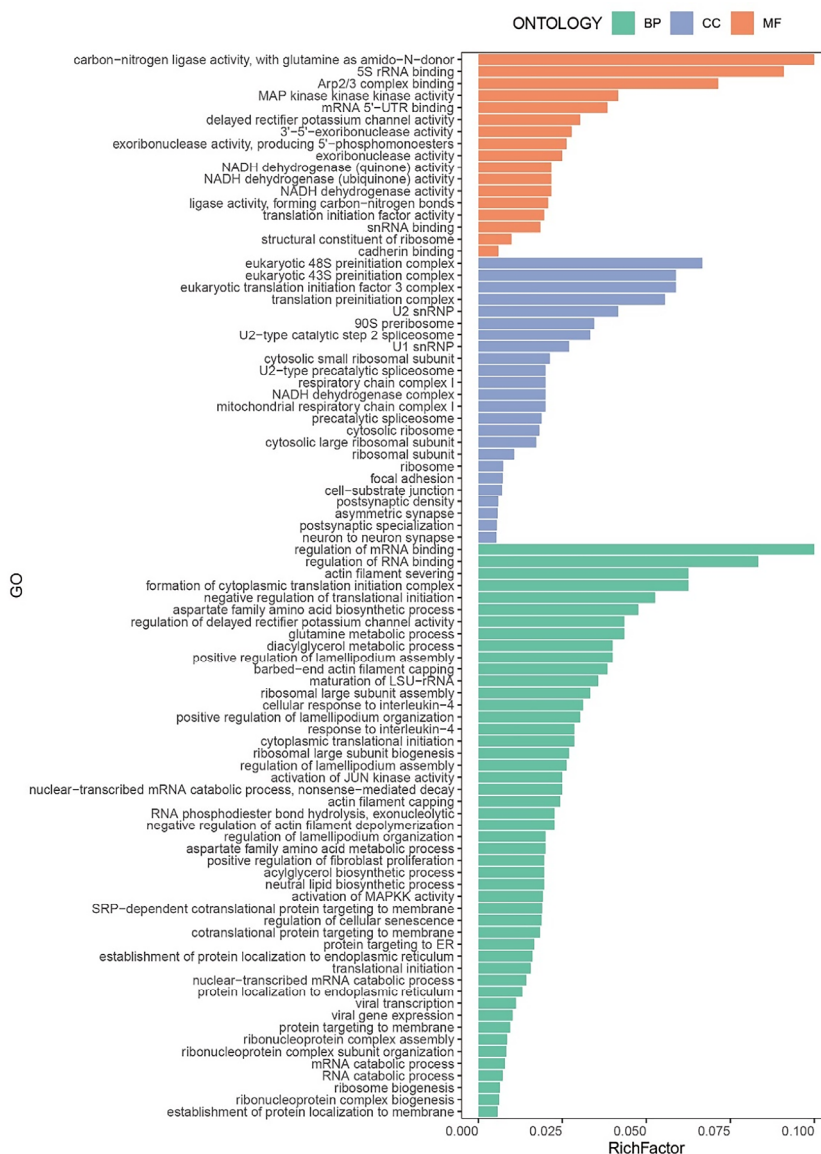
Figure 7. The GO enrichment analysis of global important genetic features. The x-axis and y-axis stand for RichFactor and GO items, respectively.

4. Discussion

Till now, a variety of classical MLs and novel DLs have been used to distinguish AD on gene expression data. The MLs generally relied on feature engineering to downscale high-dimensional gene expression features to achieve a more accurate classification of AD. While it could not be generalized by the limited performance in clinical practices. To this purpose, our proposed DL algorithm was implemented to improve the performance of AD diagnosis on tabular gene data, and even was used to exploit the intrinsic association of gene expression features with AD pathology. It was verified that our proposed DGS-TabNet presented the best performance for AD classification on public gene expression data.

Since the gene expression data usually contained redundant and irrelevant features, the purpose of DGS was generally used to select the most relevant subset of features from whole features. It was clear that the feature selection decided the performance of classifiers and the DGS applied to genetic

data had aroused the concerns of many studies [63]. During the performance evaluation, the traditional SVM performed better performance than four other MLs. This was because the SVM had the advantage of being robust against noise in genetic data and deducing the solution unique for linearly separable problems [64]. For DLs, the DGS-TabNet presented the best performance for AD classification. It was due to the fact that the DGS module could effectively select the deferential genes for biological interpretability between normal phenotypes and disease phenotypes. Without reliable feature extraction, the MLP and NODE preserved low accuracies for AD classification.

Learned from the KEGG and GO enrichment analyses, the globally important genetic features were used to explore the biological interpretability of the proposed DL model. It was proved that one of the global important genes of AVIL actively participated in the protein encoding of the gelsolin/villin family of proteins and it might play a great role in the development of neuronal cells [65]. Whereas the neuropathological features of AD included neurogenic fiber tangles with hyperphosphorylated tau proteins [66]. These processes included the proliferation, differentiation and maturation of neural stem cells and the regulation of their synaptic and neurotransmission-related processes through interactions. It was suggested that the AVIL gene might be potentially relevant to the formation of AD. According to another NDUFS4 gene, it was thought that the mutations of NDUFS4 would lead to mitochondrial complex I defects [67]. During the KEGG enrichment analysis, the gene NDUFS4 was reported to correlate with 12 biological neurodegenerative pathways, including Huntington's disease, amyotrophic lateral sclerosis, and AD, etc. The NDUFS4 enriched in 1 BP, 3 CC3 and 3 MF was proved to be correlated with mitochondrial NADH dehydrogenase and the mitochondrial integrity declined with age and affected a variety of brain functions, such as memory, learning and sensory processes [68]. Other 20 global important genes were also associated with molecular pathways involved in the development of AD such as neuroinflammation, oxidative stress, defects in mitochondrial dynamics and function, cholesterol and fatty acid metabolism, and impairment of glucose energy pathways in the brain [69]. These enrichment analyses suggested that the globally important genes were potentially highly relevant to the pathogenesis of AD, and it might serve as strong support for the biological interpretability of the proposed model.

Generally, there were several limitations in our study. First, there was no gold standard genetic data available for reliable method evaluation. Due to the data inconsistency, more large sample data sets should be used to verify the generalizability and robustness of our proposed DL algorithm. Secondly, for the betterment of robustness, more types of DL models should be investigated to assess the biological interpretability of AD. In the future, the AD prediction on longitudinal genetic data would be an important prospect for AD diagnosis.

5. Conclusions

It was concluded that the proposed DGS-TabNet could well deduce the susceptible genes, and the biological interpretability of susceptible genes also revealed the potential possibility of being AD biomarkers.

Acknowledgments

Building for Local Universities (23010502700).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. X. Liu, D. Hou, F. Lin, J. Luo, J. Xie, Y. Wang, et al., The role of neurovascular unit damage in the occurrence and development of Alzheimer's disease, *Rev. Neurosci.*, **30** (2019), 477–484. https://doi.org/10.1515/revneuro-2018-0056

2. F. Falahati, E. Westman, A. Simmons, Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging, *J. Alzheimers Dis.*, **41** (2014), 685–708. https://doi.org/10.3233/JAD-131928

3. A. B. Sallim, A. A. Sayampanathan, A. Cuttilan, R. Chun-Man Ho, Prevalence of mental health disorders among caregivers of patients with Alzheimer disease, *J. Am. Med. Dir. Assoc.*, **16** (2015), 1034–1041. https://doi.org/10.1016/j.jamda.2015.09.007

4. E. L. G. E. Koedam, V. Lauffer, A. E. van der Vlies, W. M. van der Flier, P. Scheltens, Y. A. L. Pijnenburg, Early-versus late-onset Alzheimer's disease: More than age alone, *J. Alzheimers Dis.*, **19** (2010), 1401–1408. https://doi.org/10.3233/JAD-2010-1337

5. Y. Freudenberg-Hua, W. Li, P. Davies, The role of genetics in advancing precision medicine for Alzheimer's disease—a narrative review, *Front. Med.*, **5** (2018), 108. https://doi.org/10.3389/fmed.2018.00108

6. E. Giacobini, G. Gold, Alzheimer disease therapy—moving from amyloid-β to tau, *Nat. Rev. Neurol.*, **9** (2013), 677–686. https://doi.org/10.1038/nrneurol.2013.223

7. R. J. Jutten, S. A. M. Sikkes, R. E. Amariglio, R. F. Buckley, M. J. Properzi, G. A. Marshall, et al., Identifying sensitive measures of cognitive decline at different clinical stages of Alzheimer's disease, *J. Int. Neuropsychol. Soc.*, **27** (2021), 426–438. https://doi.org/10.1017/S1355617720000934

8. D. A. McGrowder, F. Miller, K. Vaz, C. Nwokocha, C. Wilson-Clarke, M. Anderson-Cross, et al., Cerebrospinal fluid biomarkers of Alzheimer's disease: Current evidence and future perspectives, *Brain Sci.*, **11** (2021), 215. https://doi.org/10.3390/brainsci11020215

9. R. L. Cazzato, J. Garnon, B. Shaygi, G. Koch, G. Tsoumakidou, J. Caudrelier, et al., PET/CT-guided interventions: Indications, advantages, disadvantages and the state of the art, *Minimally Invasive Ther. Allied Technol.*, **27** (2018), 27–32. https://doi.org/10.1080/13645706.2017.1399280

10. M. Amini, M. M. Pedram, A. Moradi, M. Jamshidi, M. Ouchani, Single and combined neuroimaging techniques for Alzheimer's disease detection, *Comput. Intell. Neurosci.*, **2021** (2021), 9523039. https://doi.org/10.1155/2021/9523039

11. C. E. Wierenga, M. W. Bondi, Use of functional magnetic resonance imaging in the early identification of Alzheimer's disease, *Neuropsychol. Rev.*, **17** (2007), 127–143. https://doi.org/10.1007/s11065-007-9025-y

12. N. J. Gong, C. C. Chan, L. M. Leung, C. S. Wong, R. Dibb, C. Liu, Differential microstructural and morphological abnormalities in mild cognitive impairment and Alzheimer's disease: Evidence from cortical and deep gray matter, *Hum. Brain Mapp.*, **38** (2017), 2495–2508. https://doi.org/10.1002/hbm.23535

13. C. Van Cauwenberghe, C. Van Broeckhoven, K. Sleegers, The genetic landscape of Alzheimer disease: Clinical implications and perspectives, *Genet. Med.*, **18** (2016), 421–430. https://doi.org/10.1038/gim.2015.117

14. B. L. Romero-Rosales, J. G. Tamez-Pena, H. Nicolini, M. G. Moreno-Treviño, V. Trevino, Improving predictive models for Alzheimer's disease using GWAS data by incorporating misclassified samples modeling, *PLoS One*, **15** (2020). https://doi.org/10.1371/journal.pone.0232103

15. T. Jo, K. Nho, A. J. Saykin, Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data, *Front. Aging Neurosci.*, **11** (2019). https://doi.org/10.3389/fnagi.2019.00220

16. J. Ha, MDMF: predicting miRNA-disease association based on matrix factorization with disease similarity constraint, *J. Pers. Med.*, **12** (2022). https://doi.org/10.3390/jpm12060885

17. J. Ha, SMAP: Similarity-based matrix factorization framework for inferring miRNA-disease association, *Knowl-Based Syst.*, **263** (2023). https://doi.org/10.1016/j.knosys.2023.110295

18. J. De. Velasco Oriol, E. E. Vallejo, K. Estrada, Benchmarking machine learning models for late-onset Alzheimer's disease prediction from genomic data, *BMC Bioinf.*, **20** (2019), 1–17. https://doi.org/10.1186/s12859-019-3158-x

19. L. Xu, G. Liang, C. Liao, G. D. Chen, C. C. Chang, An efficient classifier for Alzheimer's disease genes identification, *Molecules*, **23** (2018), 3140. https://doi.org/10.3390/molecules23123140

20. D. Castillo-Barnes, L. Su, J. Ramírez, D. Salas-Gonzalez, F. J. Martinez-Murcia, I. A. Illan, et al., Autosomal dominantly inherited Alzheimer disease: Analysis of genetic subgroups by machine learning, *Inf. Fusion*, **58** (2020), 153–167. https://doi.org/10.1016/j.inffus.2020.01.001

21. N. Voyle, A. Keohane, S. Newhouse, K. Lunnon, C. Johnston, H. Soininen, et al., A pathway based classification method for analyzing gene expression for Alzheimer's disease diagnosis, *J. Alzheimers Dis.*, **49** (2016), 659–669. https://doi.org/10.3233/JAD-150440

22. E. Moradi, M. Marttinen, T. Häkkinen, M. Hiltunen, M. Nykter, Supervised pathway analysis of blood gene expression profiles in Alzheimer's disease, *Neurobiol. Aging*, **84** (2019), 98–108. https://doi.org/10.1016/j.neurobiolaging.2019.07.004

23. D. Cheng, M. Liu, Classification of Alzheimer's disease by cascaded convolutional neural networks using PET images, in *Machine Leaening in Medical Imaging*, Springer, (2017), 106–113. https://doi.org/10.1007/978-3-319-67389-9_13

24. M. Grassi, G. Perna, D. Caldirola, K. Schruers, R. Duara, D. A. Loewenstein, A clinically-translatable machine learning algorithm for the prediction of Alzheimer's disease conversion in individuals with mild and premild cognitive impairment, *J. Alzheimers Dis.*, **61** (2018), 1555–1573. https://doi.org/10.3233/JAD-170547

25. S. M. Plis, D. R. Hjelm, R. Salakhutdinov, E. A. Allen, H. J. Bockholt, J. D. Long, et al., Deep learning for neuroimaging: a validation study, *Front. Neurosci.*, **8** (2014). https://doi.org/10.3389/fnins.2014.00229

26. S. Wang, H. Wang, Y. Shen, X. Wang, Automatic recognition of mild cognitive impairment and Alzheimers disease using ensemble based 3D densely connected convolutional networks, in *17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, (2018), 517–523. https://doi.org/10.1109/icmla.2018.00083

27. W. Yu, B. Lei, M. K. Ng, A. C. Cheung, Y. Shen, S. Wang, Tensorizing GAN with high-order pooling for Alzheimer's disease assessment, *IEEE Trans. Neural Networks Learn. Syst.*, **33** (2022), 4945–4959. https://doi.org/10.1109/TNNLS.2021.3063516

28. W. Yu, B. Lei, S. Wang, Y. Liu, Z. Feng, Y. Hu, et al., Morphological feature visualization of Alzheimer's disease via multidirectional perception GAN, *IEEE Trans. Neural Networks Learn. Syst.*, **2022** (2022), 1–15. https://doi.org/10.1109/TNNLS.2021.3118369

29. T. Lee, H. Lee, Prediction of Alzheimer's disease using blood gene expression data, *Sci. Rep.*, **10** (2020), 3485. https://doi.org/10.1038/s41598-020-60595-1

30. N. Mahendran, P. Durai Raj Vincent, A deep learning framework with an embedded-based feature selection approach for the early detection of the Alzheimer's disease, *Comput. Biol. Med.*, **141** (2022), 105056. https://doi.org/10.1016/j.compbiomed.2021.105056

31. C. Park, J. Ha, S. Park, Prediction of Alzheimer's disease based on deep neural network by integrating gene expression and DNA methylation dataset, *Expert Syst. Appl.*, **140** (2020). https://doi.org/10.1016/j.eswa.2019.112873

32. Y. Liu, Z. Li, Q. Ge, N. Lin, M. Xiong, Deep feature selection and causal analysis of Alzheimer's disease, *Front. Neurosci.*, **13** (2019). https://doi.org/10.3389/fnins.2019.01198

33. S. Spasov, L. Passamonti, A. Duggento, P. Liò, N. Toschi, A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease, *NeuroImage*, **189** (2019), 276–287. https://doi.org/10.1016/j.neuroimage.2019.01.031

34. S. Gauthier, B. Reisberg, M. Zaudig, R. C. Petersen, K. Ritchie, K. Broich, et al., Mild cognitive impairment, *Lancet*, **367** (2006), 1262–1270. https://doi.org/10.1016/S0140-6736(06)68542-5

35. M. Grundman, R. C. Petersen, S. H. Ferris, R. G. Thomas, P. S. Aisen, D. A. Bennett, et al., Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials, *Arch. Neurol.*, **61** (2004), 59–66. https://doi.org/10.1001/archneur.61.1.59

36. A. Kadra, M. Lindauer, F. Hutter, J. Grabocka, Well-tuned simple nets excel on tabular datasets, 2021.

37. S. Popov, S. Morozov, A. Babenko, Neural oblivious decision ensembles for deep learning on tabular data, *arXiv preprint*, 2019, arXiv:1909.06312. https://doi.org/10.48550/arXiv.1909.06312

38. C. Shah, Q. Du, Y. Xu, Enhanced TabNet: attentive interpretable tabular learning for hyperspectral image classification, *Remote Sens.*, **14** (2022), 716. https://doi.org/10.3390/rs14030716

39. Y. Y. Song, Y. Lu, Decision tree methods: applications for classification and prediction, *Shanghai Arch Psychiatry*, **27** (2015), 130–135.

40. G. Biau, E. Scornet, A random forest guided tour, *TEST*, **25** (2016), 197–227. https://doi.org/10.1007/s11749-016-0481-7

41. C. Zhang, C. Liu, X. Zhang, G. Almpanidis, An up-to-date comparison of state-of-the-art classification algorithms, *Expert Syst. Appl.*, **82** (2017), 128–150. https://doi.org/10.1016/j.eswa.2017.04.003

42. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, et al., Lightgbm: A highly efficient gradient boosting decision tree, 2017.

43. M. Pirooznia, J. Y. Yang, M. Q. Yang, Y. Deng, A comparative study of different machine learning methods on microarray gene expression data, *BMC Genomics*, **9** (2008). https://doi.org/10.1186/1471-2164-9-S1-S13

44. Q. S. Zhang, S. C. Zhu, Visual interpretability for deep learning: a survey, *Front. Inf. Technol. Electron. Eng.*, **19** (2018), 27–39. https://doi.org/10.1631/FITEE.1700808

45. S. Lovestone, P. Francis, I. Kloszewska, P. Mecocci, A. Simmons, H. Soininen, et al., AddNeuroMed-the european collaboration for the discovery of novel biomarkers for Alzheimer's disease, *Ann. N. Y. Acad. Sci.*, **1180** (2009), 36–46. https://doi.org/10.1111/j.1749-6632.2009.05064.x

46. S. Sood, I. J. Gallagher, K. Lunnon, E. Rullman, A. Keohane, H. Crossland, et al., A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status, *Genome Biol.*, **16** (2015), 185. https://doi.org/10.1186/s13059-015-0750-x

47. S. Davis, P. S. Meltzer, GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor, *Bioinformatics*, **23** (2007), 1846–1847. https://doi.org/10.1093/bioinformatics/btm254

48. X. Li, H. Wang, J. Long, G. Pan, T. He, O. Anichtchik, et al., Systematic analysis and biomarker study for Alzheimer's disease, *Sci. Rep.*, **8** (2018), 17394. https://doi.org/10.1038/s41598-018-35789-3

49. A. Antonell, A. Llado, J. Altirriba, T. Botta-Orfila, M. Balasa, M. Fernandez, et al., A preliminary study of the whole-genome expression profile of sporadic and monogenic early-onset Alzheimer's disease, *Neurobiol. Aging*, **34** (2013), 1772–1778. https://doi.org/10.1016/j.neurobiolaging.2012.12.026

50. S. Arık, T. Pfister, TabNet: Attentive interpretable tabular learning, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **35** (2021), 6679–6687. https://doi.org/10.1609/aaai.v35i8.16826

51. N. N. Parikshak, M. J. Gandal, D. H. Geschwind, Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders, *Nat. Rev. Genet.*, **16** (2015), 441–458. https://doi.org/10.1038/nrg3934

52. G. K. Smyth, limma: Linear models for microarray data, in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Springer, (2005), 397–420. https://doi.org/10.1007/0-387-29362-0_23

53. C. Garbin, X. Zhu, O. Marques, Dropout vs. batch normalization: an empirical study of their impact to deep learning, *Multimedia Tools Appl.*, **79** (2020), 12777–12815. https://doi.org/10.1007/s11042-019-08453-9

54. Y. N. Dauphin, A. Fan, M. Auli, D. Grangier, Language modeling with gated convolutional networks, in *Conference on Machine Learning*, (2017), 933–941.

55. J. Gehring, M. Auli, D. Grangier, D. Yarats, Y. N. Dauphin, Convolutional sequence to sequence learning, in *Proceedings of the 34th International Conference on Machine Learning*, **70** (2017), 1243–1252.

56. A. Martins, R. Astudillo, From softmax to sparsemax: a sparse model of attention and multi-label classification, in *Proceedings of the 33rd International Conference on Machine Learning*, **48** (2016), 1614–1623.

57. N. Deepa, S. P. Chokkalingam, Optimization of VGG16 utilizing the Arithmetic Optimization Algorithm for early detection of Alzheimer's disease, *Biomed. Signal Process. Control*, **74** (2022), 103455. https://doi.org/10.1016/j.bspc.2021.103455

58. M. B. Kursa, W. R. Rudnicki, Feature selection with the Boruta package, *J. Stat. Software*, **36** (2010), 1–13. https://doi.org/10.18637/jss.v036.i11

59. A. Kulshrestha, O. Farooq, Seizure prediction using fybrid features, in *IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, (2020), 1–6. https://doi.org/10.1109/upcon50219.2020.9376552

60. R. Martinez-Cantin, Bayesian optimization with adaptive kernels for robot control, in *IEEE International Conference on Robotics and Automation (ICRA)*, (2017), 3350–3356. https://doi.org/10.1109/ICRA.2017.7989380

61. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, et al., ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data, *Innovation*, **2** (2021), 100141. https://doi.org/10.1016/j.xinn.2021.100141

62. The Gene Ontology Consortium, The gene ontology resource: 20 years and still GOing strong, *Nucleic Acids Res.*, **47** (2019), 330–338. https://doi.org/10.1093/nar/gky1055

63. J. Krawczuk, T. Łukaszuk, The feature selection bias problem in relation to high-dimensional gene data, *Artif. Intell. Med.*, **66** (2016), 63–71. https://doi.org/10.1016/j.artmed.2015.11.001

64. S. S. Mehta, N. S. Lingayat, Development of SVM based classification techniques for the delineation of wave components in 12-lead electrocardiogram, *Biomed. Signal Process. Control*, **3** (2008), 341–349. https://doi.org/10.1016/j.bspc.2008.04.002

65. Z. Tümer, P. J. P. Croucher, L. R. Jensen, J. Hampe, C. Hansen, V. Kalscheuer, et al., Genomic structure, chromosome mapping and expression analysis of the human AVIL gene, and its exclusion as a candidate for locus for inflammatory bowel disease at 12q13–14 (IBD2), *Gene*, **288** (2002), 179–185. https://doi.org/10.1016/S0378-1119(02)00478-X

66. S. Hong, V. F. Beja-Glasser, B. M. Nfonoyim, A. Frouin, S. Li, S. Ramakrishnan, et al., Complement and microglia mediate early synapse loss in Alzheimer mouse models, *Science*, **352** (2016), 712–716. https://doi.org/10.1126/science.aad8373

67. A. Quintana, S. E. Kruse, R. P. Kapur, E. Sanz, R. D. Palmiter, Complex I deficiency due to loss of Ndufs4 in the brain results in progressive encephalopathy resembling Leigh syndrome, *Proc. Natl. Acad. Sci.*, **107** (2010), 10996–11001. https://doi.org/10.1073/pnas.1006214107

68. D. F. F. Silva, A. R. Esteves, C. R. Oliveira, S. M. Cardoso, Mitochondria: the common upstream driver of amyloid-β and tau pathology in Alzheimer's disease, *Curr. Alzheimer Res.*, **8** (2011), 563–572. https://doi.org/10.2174/156720511796391872

69. M. Calabrò, C. Rinaldi, G. Santoro, C. Crisafulli, The biological pathways of Alzheimer disease: a review, *AIMS Neurosci.*, **8** (2021), 86–132. https://doi.org/10.3934/Neuroscience.2021005