



Identifying Key MicroRNA Signatures for Neurodegenerative Diseases With Machine Learning Methods

ZhanDong Li^{1†}, Wei Guo^{2†}, ShiJian Ding^{3†}, Lei Chen⁴, KaiYan Feng⁵, Tao Huang^{6,7*} and Yu-Dong Cai^{3*}

¹College of Food Engineering, Jilin Engineering Normal University, Changchun, China, ²Key Laboratory of Stem Cell Biology, Shanghai Jiao Tong University School of Medicine (SJTUSM) and Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS), Shanghai, China, ³School of Life Sciences, Shanghai University, Shanghai, China, ⁴College of Information Engineering, Shanghai Maritime University, Shanghai, China, ⁵Department of Computer Science, Guangdong AIB Polytechnic College, Guangzhou, China, ⁶Bio-Med Big Data Center, CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China, ⁷CAS Key Laboratory of Tissue Microenvironment and Tumor, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China

OPEN ACCESS

Edited by:

Liang Cheng,
Harbin Medical University, China

Reviewed by:

Jing Lu,
Yantai University, China
Yue Zhang,
Harbin Engineering University, China

*Correspondence:

Tao Huang
tohuangtao@126.com
Yu-Dong Cai
cai_yud@126.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Neurogenomics,
a section of the journal
Frontiers in Genetics

Received: 22 February 2022

Accepted: 30 March 2022

Published: 21 April 2022

Citation:

Li Z, Guo W, Ding S, Chen L, Feng K,
Huang T and Cai Y-D (2022) Identifying
Key MicroRNA Signatures for
Neurodegenerative Diseases With
Machine Learning Methods.
Front. Genet. 13:880997.
doi: 10.3389/fgene.2022.880997

Neurodegenerative diseases, including Alzheimer's disease (AD), Parkinson's disease, and many other disease types, cause cognitive dysfunctions such as dementia via the progressive loss of structure or function of the body's neurons. However, the etiology of these diseases remains unknown, and diagnosing less common cognitive disorders such as vascular dementia (VaD) remains a challenge. In this work, we developed a machine-learning-based technique to distinguish between normal control (NC), AD, VaD, dementia with Lewy bodies, and mild cognitive impairment at the microRNA (miRNA) expression level. First, unnecessary miRNA features in the miRNA expression profiles were removed using the Boruta feature selection method, and the retained feature sets were sorted using minimum redundancy maximum relevance and Monte Carlo feature selection to provide two ranking feature lists. The incremental feature selection method was used to construct a series of feature subsets from these feature lists, and the random forest and PART classifiers were trained on the sample data consisting of these feature subsets. On the basis of the model performance of these classifiers with different number of features, the best feature subsets and classifiers were identified, and the classification rules were retrieved from the optimal PART classifiers. Finally, the link between candidate miRNA features, including hsa-miR-3184-5p, has-miR-6088, and has-miR-4649, and neurodegenerative diseases was confirmed using recently published research, laying the groundwork for more research on miRNAs in neurodegenerative diseases for the diagnosis of cognitive impairment and the understanding of potential pathogenic mechanisms.

Keywords: neurodegenerative disease, microRNA, feature selection, expression pattern, classification algorithm

1 INTRODUCTION

Dementia is one kind of cognitive impairment that is characterized by difficulties in memory, language, and behavior. Of all chronic diseases, dementia has become one of the most important contributors to dependence and disability (Iliffe et al., 2009). With an increasing number of morbidity, dementia has become a great concern worldwide (Prince et al., 2016). Unfortunately, there is no cure for this disease at present, and earlier diagnosis and interventions to slow down the disease progress are needed (Iliffe et al., 2009). Therefore, researchers have focused on searching effective diagnostic methods, including the identification of new biomarkers for diagnosis, and interventions for dementia.

Although young-onset cases are increasingly recognized, dementia is typically a condition that affects older people. Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the most common cause of intellectual deficit in populations older than 65 years. More than 20% of individuals over 80 years of age are affected by AD, and epidemiological data predict that there will be over 35 million AD patients by 2050 (Danborg et al., 2014). Other less common causes of cognitive impairment include vascular dementia (VaD) whose definition and distinction remain controversial, mixed dementia, and dementia with Lewy bodies (DLB) (McKeith et al., 1996). Diagnosing dementia is markedly difficult due to its insidious onset and diversity of other presenting symptoms such as difficulty in making decisions (Kostopoulou et al., 2008). Recent studies have reported that certain protein biomarkers in cerebrospinal fluid (CSF) can be applied in the clinical diagnosis of AD with a high predictive accuracy (De Meyer et al., 2010). However, such biomarkers have their limitations in differentiating AD from other types of dementia. In addition, biomarkers in CSF require an invasive collection process; thus, new methods through less invasive procedures are needed. Considering that the diagnosis of dementia subtypes is important to manage different therapies, disease courses, and outcomes for different dementias (Robinson et al., 2015), development of better biomarkers for AD and other dementias will contribute to more accurate diagnosis for an early and specialized treatment.

For a better clinical care in disease prevention and treatment, several computational models have been developed to predict dementia risk or subtypes (Stephan et al., 2010). For example, Licher et al. (2019) reported a dementia risk model using optimism-corrected C-statistics, which can be used to identify individuals with high risk of dementia with an accuracy of 0.86. This model was based on comprehensive clinical information such as age, cognitive impairment, and lifestyle factors. Interestingly, a novel machine learning prediction model for dementia risk identification using the voice data from daily conversations was proposed by Shimoda et al. (2021). They applied three strategies including extreme gradient boosting, random forest (RF), and logistic regression methods in developing models, which had AUCs of 0.86, 0.88, and 0.89, respectively. Li et al. (2019) reported a deep learning model for the early prediction of AD using hippocampal magnetic

resonance imaging data, which achieved a concordance index of 0.762. In addition, genetic data were taken into account to improve the ability of the prediction model given that many genes were confirmed to be associated with AD (Seshadri et al., 2010). So far, models in dementia prediction lack molecular signatures such as transcriptional expression, which can reflect the underlying pathogenic mechanisms.

MicroRNAs (miRNAs) are small non-coding RNA molecules of approximately 22 nucleotides in length, which have been shown to regulate gene expression by binding to complementary regions of messenger transcripts (Lagos-Quintana et al., 2001). The detection of circulating miRNA levels has been proposed to be a potential diagnostic tool for a number of diseases (Gilad et al., 2008). MiRNAs play a crucial role in the control of neuronal cell development (Mistur et al., 2009). The alteration of the expression of some miRNAs has been shown to relate to various neurological diseases including AD. For example, miR-137, miR-181c, and miR-29a/b were reported to be involved in AD by modulating ceramide levels (Geekiyana and Chan, 2011). The downregulation of miR-16, miR-195, and miR-103 was observed in the brain of AD patients, and these miRNAs were shown to target the β -site amyloid precursor protein cleaving enzyme 1 (*BACE1*), which is involved in amyloid plaque formation (Bekris et al., 2013). Cogswell et al. found significantly decreased expression of miR-9, which regulates neuronal differentiation, in the human hippocampus of AD patients (Cogswell et al., 2008; Coolen et al., 2013). Different expression patterns of miRNAs have also been found between AD and other neurodegenerative diseases; for example, miR-15a is uniquely elevated in the plasma of AD patients (Bekris et al., 2013). Therefore, miRNAs in the blood or serum are easily accessible and noninvasive biomarkers for diagnosing dementia. In addition, some miRNAs can be used to distinguish different subtypes of dementia for more precise treatment.

In this study, on the basis of the miRNA expression profiles from 1601 serum samples (Shigemizu et al., 2019a), including AD cases, VaD cases, DLB cases, mild cognitive impairment (MCI) cases, and normal controls (NC), we computationally analyzed such expression data. The data was first analyzed by Boruta (Kursa and Rudnicki, 2010), irrelevant miRNA features were excluded. Remaining miRNA features were evaluated by minimum redundancy maximum relevance (mRMR) (Peng et al., 2005) and Monte Carlo feature selection (MCFS) (Dramiński et al., 2007), respectively. Two feature lists were generated, which were fed into incremental feature selection (IFS) (Liu and Setiono, 1998), incorporating random forest (RF) (Breiman, 2001) or PART (Frank and Witten, 1998). As a result, we identified the crucial miRNAs that show the most relevance to the distinction of four different types of dementia and NC, suggesting that these selected miRNAs may play crucial roles in neuronal development. Furthermore, we also identified interesting classification rules, which suggested different miRNA expression patterns on different dementia subtypes and NC. These results can guide further research about the interaction between miRNAs and neurodegenerative diseases. Finally, we constructed two optimal classifiers with high accuracy to group individuals into the corresponding categories (four dementia

TABLE 1 | Sample size for normal control and four neurodegenerative diseases.

Disease case	Sample size
Alzheimer's disease (AD)	1,021
Vascular dementia (VaD)	91
Dementia with lewy bodies (DLB)	169
Mild cognitive impairment (MCI)	32
Normal control (NC)	288

subtypes and NC). They can be useful tools for the precise diagnosis of dementia subtypes. Our study highlights the potential application of miRNAs in dementia subtype diagnosis, indicating that the prediction framework using serum miRNA expression data can provide feasible therapeutic and diagnostic targets for dementia.

2 MATERIALS AND METHODS

2.1 Dataset

In this study, the miRNA expression profiles were obtained from the Gene Expression Omnibus database under the accession code GSE120584 (Shigemizu et al., 2019a; Shigemizu et al., 2019b; Asanomi et al., 2021). These expression profiles include 1,601 samples, which are composed of AD cases, VaD cases, DLB cases, MCI cases, and NC. The sample sizes of different cases are provided in **Table 1**. A total of 2547 miRNAs were identified in the expression profiles. Subsequently, we performed a computational workflow to detect key miRNA features and expression patterns in the expression profiles.

2.2 Boruta Feature Filtering

Aside from the time and energy costs of dealing with a high number of features, most machine learning algorithms work better when the number of predicting features employed is kept as small as possible. We thus applied a Boruta analysis on the miRNA expression profiles to reduce feature dimension and retain important miRNA features (Kursa and Rudnicki, 2010). Boruta is a feature selection approach based on the RF model to access feature importance (Z-score) by comparing the relevance of real features with shadow features, which are randomly shuffled from original features. The python application from https://github.com/scikit-learn-contrib/boruta_py with default parameters was used for Boruta feature selection in this analysis.

2.3 Feature Ranking

2.3.1 Minimum Redundancy Maximum Relevance

The mRMR algorithm (Peng et al., 2005) is an entropy-based feature selection method that calculates the mutual information (MI) between a group of features and class variable. The MI is defined as follows:

$$I(X; Y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (1)$$

where $p(x, y)$ is the joint probability density function of X and Y , $p(x)$ and $p(y)$ are the marginal probability density functions of X and Y , respectively. In the mRMR method, the correlation (D) between features and target label and the redundancy (R) between features and other features are computed as follows:

$$D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c), \quad (2)$$

where S is the selected features and $I(x_i; c)$ is the MI between feature x_i and the target label c .

$$R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j), \quad (3)$$

where $I(x_i, x_j)$ is the MI between feature x_i and feature x_j . To repeatedly add a new feature to a feature subset S , the following objective function is optimized:

$$\max \Phi(D, R), \Phi = D - R, \quad (4)$$

In this study, we used the mRMR program acquired from <http://home.penglab.com/proj/mRMR/> to rank all the features obtained by Boruta analysis, resulting in an mRMR feature list.

2.3.2 Monte Carlo Feature Selection

The MCFS method (Dramiński et al., 2007) evaluates the feature importance by creating numerous decision trees. More specifically, for a dataset with M features, MCFS first randomly constructs s feature subsets with m features ($m \ll M$). For each feature subset, t decision trees are constructed using the bootstrap sampling method. Finally, $s \times t$ classification trees are constructed and evaluated. The RI score of feature g based on these classification trees is defined as follows:

$$RI_g = \sum_{\tau=1}^{st} (wAcc)^\mu \sum_{n_g(\tau)} IG(n_g(\tau)) \left(\frac{no.in n_g(\tau)}{no.in \tau} \right)^\nu \quad (5)$$

where $wAcc$ is the weight accuracy of the decision tree τ ; $IG(n_g(\tau))$ denotes the gain information of node $n_g(\tau)$; $(no.in n_g(\tau))$ and $(no.in \tau)$ represent the number of samples of node $n_g(\tau)$ and the number of samples in tree τ , respectively; and μ and ν are parameters that are recommended to be 1. After MCFS processing, all features are ranked in a feature list in descending order of RI values. In this study, we applied the MCFS program developed by Draminski et al., which can be accessed at <http://www.ipipan.eu/staff/m.draminski/mcfs.html>, for feature sorting, and the parameters were set to default values. The obtained feature list was called MCFS feature list.

2.4 Incremental Feature Selection

In the previous analysis, the mRMR and MCFS feature ranking lists were obtained, but it was not possible to determine the optimal feature subsets for classifying disease cases. Thus, the IFS method (Liu and Setiono, 1998) was used in this study to identify the best number of features in a feature list for a specific classification algorithm. IFS first generates a series of feature subsets on the basis of a step size. For example, if the step size

equals to 1, the first feature subset includes one top-ranked feature, the second feature subset is made up of two top-ranked features, and so on. Then, the sample datasets represented by these feature subsets are trained by one classification algorithm (RF or PART in this study). The classifiers are evaluated by using 10-fold cross-validation (Kohavi, 1995; Tang and Chen, 2022; Yang and Chen, 2022). The evaluation metrics (e.g., Matthews correlation coefficient [MCC]) for each classifier with different number of features are obtained and used to plot IFS curves, where the X-axis is the number of features and the Y-axis is the evaluation metrics. In the end, the optimal feature subsets that achieves the best classification results are identified, and the optimal classifiers are built.

2.5 Classification Algorithms

2.5.1 RF

The RF (Breiman, 2001) is an ensemble learning algorithm that takes decision trees as the base learner. It first produces a number of training sets from the original dataset using a bootstrapping method with randomized put-back sampling. These training sets are then used to train the decision tree model individually, and the generated decision trees are formed into a forest. Lastly, the final result is determined by aggregating the voting results of many tree classifiers. As RF is powerful, it is always an important candidate for constructing efficient classifiers (Chen et al., 2017; Zhao et al., 2018; Chen et al., 2021; Li X. et al., 2022; Li Z. et al., 2022; Chen et al., 2022; Ding et al., 2022). In this study, the RF program in Weka (Frank et al., 2004) was employed with default parameters.

2.5.2 PART

In contrast to black-box models, such as RF, rule learning models may learn rules from data to make discriminations on unknown data, and these rules are commonly expressed in an IF-THEN structure, which clearly expresses the patterns existing in the data. PART is a rule-generating method that combines the Ripper and C4.5 approaches without the need for global optimization (Frank and Witten, 1998). It uses a separate-and-conquer technique to develop several partial decision trees, in which a rule is constructed each time. Then, the instances it covers are eliminated, and rules are created recursively for the remaining instances until the end. The PART program in WEKA was used with the default parameters in this investigation.

2.6 SMOTE

The distribution of samples under five cases is uneven, which may lead to the poor performance of the established classifiers. To address this issue, we applied SMOTE methods to increase the sample size of the minority class, which is an oversampling technique presented by Chawla et al. (2002). SMOTE generates synthetic samples randomly between samples of a minority class and their neighbors on the basis of the k-nearest neighbor concept. The SMOTE algorithm in Weka software was used to process the miRNA expression profiles in this investigation, resulting in an equal number of samples in each class. It was necessary to pointed out that SMOTE was only used

in evaluating the performance of classifiers in the IFS method. Pseudo samples generated by SMOTE did not participate in the mRMR or MCFS methods as they can influence the feature selection results.

2.7 Performance Measurement

For the 10-fold cross-validation, we used the MCC as a predictive metric for the evaluation of classifiers. In this study, considering that the analyzed miRNA dataset includes multiple disease cases, the multi-categorical version of MCC (Gorodkin, 2004) was applied and calculated as follows:

$$MCC = \frac{cov(X, Y)}{\sqrt{cov(X, X)cov(Y, Y)}} \quad (6)$$

where the binary matrix X represents the prediction results, the binary matrix Y indicates the real class label, and $cov(X, Y)$ stands for the covariance of the two matrices. The MCC ranges from -1 to 1 , with a value closer to 1 indicating stronger model performance.

To fully display the performance of classification models, we also calculated other measurements, including individual accuracy on each class and overall accuracy (ACC). For one class, its individual accuracy was defined as the proportion of correctly predicted samples in this class. The ACC was defined as the proportion of correctly predicted samples.

3 RESULTS

3.1 Feature Selection Results on miRNA Expression Profiles

A flow chart of the present study is illustrated in **Figure 1**. We started by removing unnecessary features using the Boruta feature selection method, and the 108 retained features are listed in **Supplementary Table S1**.

Then, using mRMR and MCFS, remaining 108 features were ranked according to feature importance, yielding two ranked feature lists (mRMR feature list and MCFS feature list), as shown in **Supplementary Table S1**. Top ten miRNA features in these two lists were investigated, as shown in **Figure 2**. Four miRNAs, including hsa-miR-3184-5p, hsa-miR-1227-5p, hsa-miR-3181, and hsa-miR-6088, appeared in the top 10 features yielded by two methods, highlighting their visibility and importance. The biological roles of these miRNA features will be explored in **Section 4**.

3.2 IFS Results on the mRMR Feature List

Based on the mRMR feature list, it was fed into the IFS method with a step size of 1, returning 108 feature subsets. For example, the first feature subset includes the first feature, the second feature subset includes the first two features, and so on. The RF and PART classifiers were trained using the sample set consisting of these feature subsets, and the performance was assessed using 10-fold cross-validation. Obtained measurements are provided in **Supplementary Table S2**. To clearly display the performance of classifiers on different feature subsets, an IFS curve was plotted

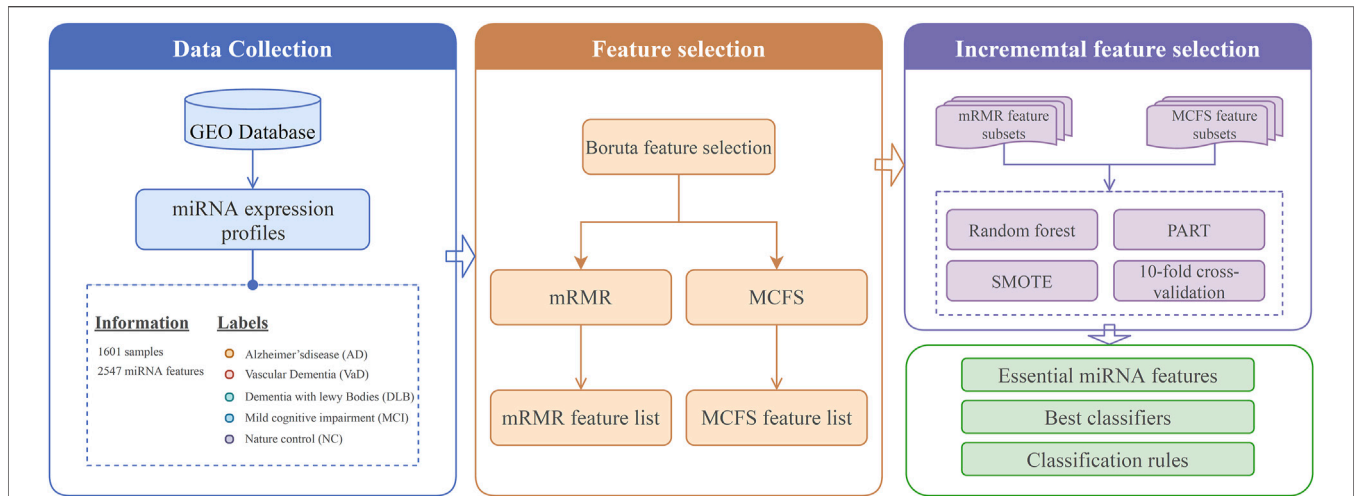


FIGURE 1 | Analysis flowchart for this study, which consists of three main steps: 1) miRNA dataset collection; 2) filtering and ranking of miRNA features in the dataset using Boruta, mRMR, and MCFS; 3) determining the essential miRNA features and building the best classifiers and classification rules using IFS method with RF and PART algorithms.

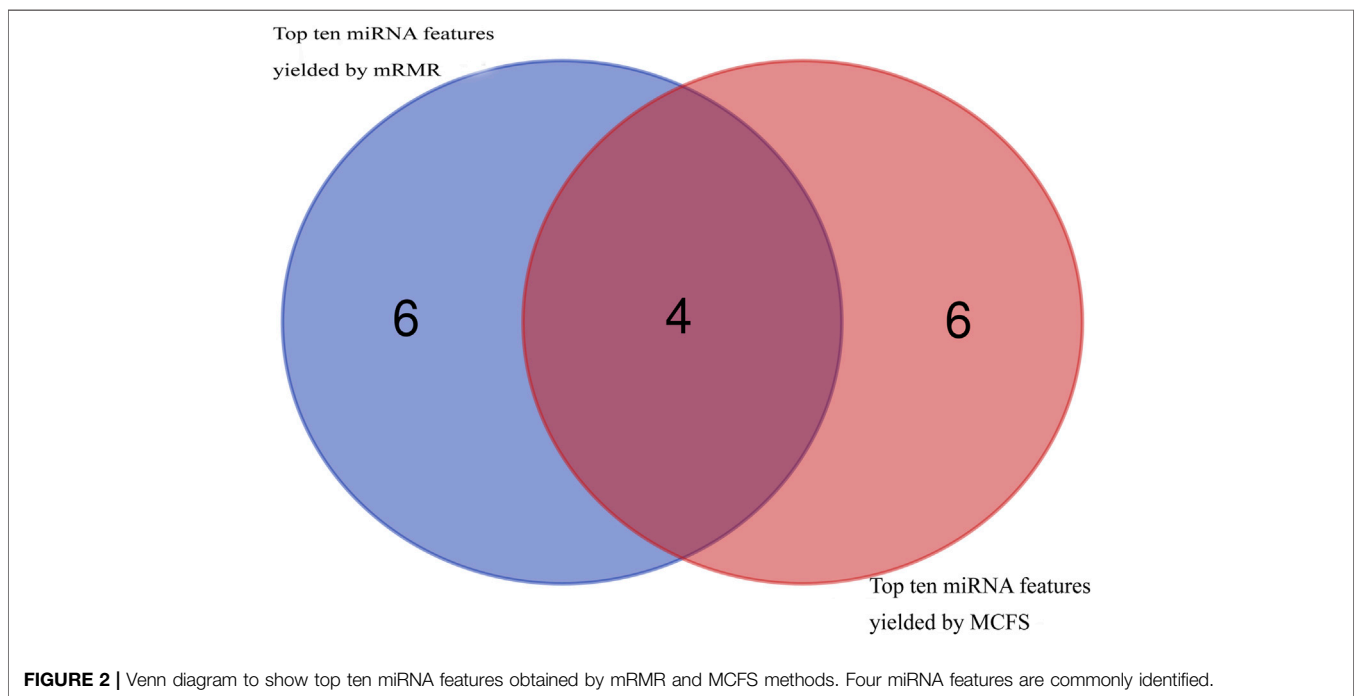


FIGURE 2 | Venn diagram to show top ten miRNA features obtained by mRMR and MCFS methods. Four miRNA features are commonly identified.

for each classification algorithm, which is shown in **Figure 3A**. When RF was selected as the classification algorithm in the IFS method, the highest MCC was 0.683, which was obtained by using top 106 features. Accordingly, the optimal RF classifier can be built with these features. The ACC of such classifier was 0.802, as listed in **Table 2**. As for PART, the highest MCC was 0.359. It was obtained by using top 72 features, with which the optimal PART classifier can be built. The ACC of such PART classifier was 0.570, as listed in **Table 2**. Clearly, the optimal PART classifier was

much inferior to the optimal RF classifier. As for their performance on five classes, individual accuracies are shown in **Figure 4A**. Evidently, the optimal RF classifier provided better performance than the optimal PART classifier on all classes. Both MCI and VaD have an individual accuracy of over 0.900 in the optimal RF classifier.

Although the optimal RF classifier gave good performance, it was not very proper to do large-scale tests because lots of miRNA features involved. In view of this, we carefully checked the IFS

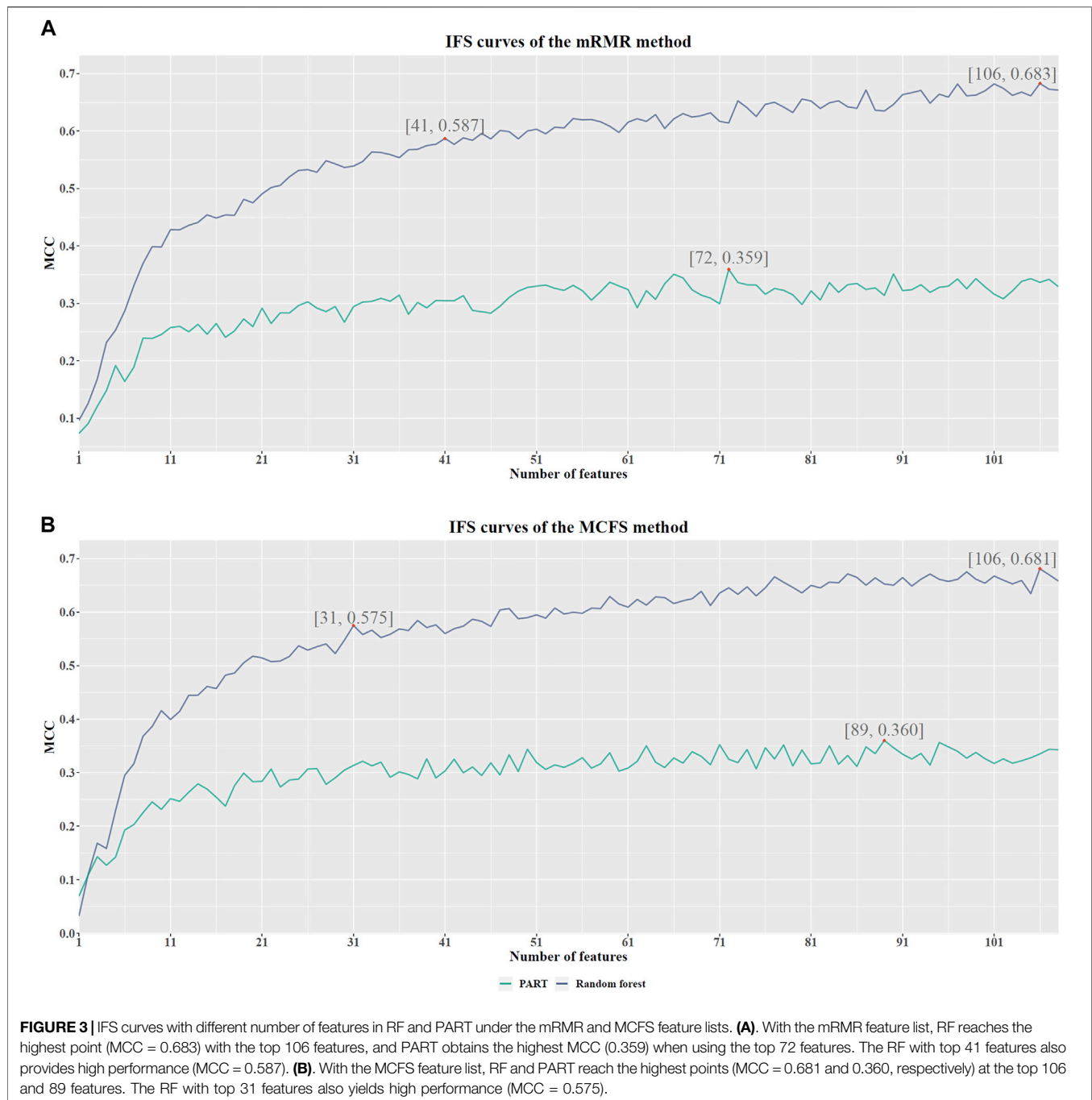
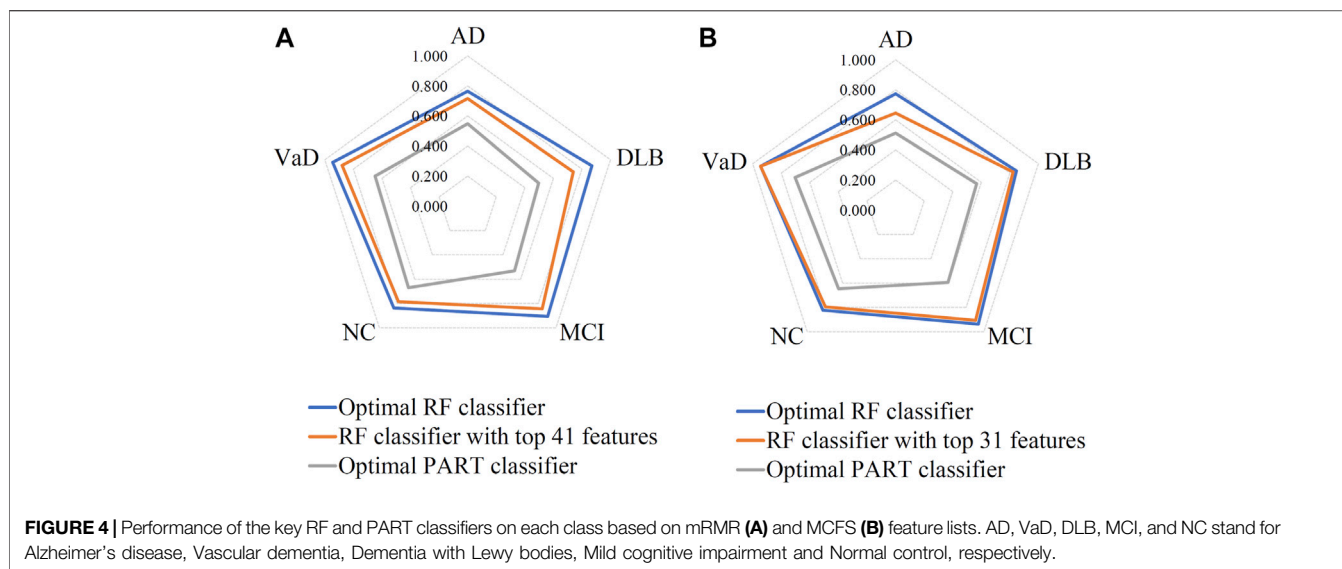


TABLE 2 | Performance of key classifiers with different algorithms based on the mRMR feature list.

Classification algorithm	Number of features	ACC	MCC
Random forest	106	0.802	0.683
Random forest	41	0.743	0.587
PART	72	0.570	0.359

results with RF and found that RF provided the MCC of 0.587 when top 41 features were used (**Figure 3A**). This classifier yielded the ACC of 0.743 (**Table 2**). Its performance on five classes is shown in **Figure 4A**. Although it provided lower performance than the optimal RF classifier, it was much faster as much less miRNA features were needed. This classifier can be an efficient tool to identify four dementia subtypes and NC.



3.3 IFS Results on the MCFS Feature List

For the MCFS feature list, the same procedures were conducted. Detailed performance of RF and PART on different number of features is listed in **Supplementary Table S3**. Likewise, an IFS curve was plotted for each classification algorithm to display the performance of them on different feature subsets, as illustrated in **Figure 3B**. It can be observed that the highest MCC for RF was 0.681, which was obtained by using top 106 features. Thus, we can build the optimal RF classifier with these features. The ACC of such classifier was 0.803, as listed in **Table 3**. Its performance on each class is shown in **Figure 4B**. Compared with the performance of the optimal RF classifier in **Section 3.2**, their performance was almost equal. As for PART, its highest MCC was 0.360. It was obtained by using top 89 miRNA features. Accordingly, the optimal PART classifier was built using these features. The ACC of this classifier was 0.555 (**Table 3**). The performance of this classifier on each class is shown in **Figure 4B**. Evidently, this PART classifier provided equal performance to the optimal PART classifier in **Section 3.2**. However, they were all inferior to the optimal RF classifiers.

Similar to the optimal RF classifier in **Section 3.2**, this optimal RF classifier also need several features. It was necessary to discover another RF classifier with a higher efficiency. After careful checking, we found that RF classifier with top 31 features can produce the MCC of 0.575 (**Figure 3B**) and ACC

of 0.713 (**Table 3**). Its performance on five classes is shown in **Figure 4B**. Clearly, it was inferior to the optimal RF classifier. However, it had a higher efficiency because it used much less features. Thus, it can be a useful tool to identify four dementia subtypes and NC. Furthermore, the performance of such RF classifier and RF classifier with top 41 features yielded by mRMR method was almost equal.

3.4 miRNA Expression Patterns Extracted From the Optimal PART Classifiers

Although the performance of two optimal PART classifier was much lower than two optimal RF classifiers, they can give interpretable rules, which can help us uncover the difference between four dementia subtypes and NC at miRNA level. For the mRMR feature list, the optimal PART classifier used top 72 features. With these features, PART was applied to all samples, resulting in 245 rules. These rules are provided in **Supplementary Table S4**. Likewise, for the MCFS feature list, top 89 features were adopted in the optimal PART classifier. 251 decision rules were obtained by applying PART on these features, which are also available in **Supplementary Table S4**. Accordingly, we accessed two groups of decision rules. For each group, each class received some rules. The number of rules for each class on each group is shown in **Figure 5**. With the exception of MCI, which has a relatively small number of rules, the numbers of rules of other classes were quite considerable. Some key expression rules are listed in **Tables 4, 5** and the relevance of these rules in differentiating neurological disorders will be reviewed in **Section 4.1**.

TABLE 3 | Performance of key classifiers with different algorithms based on the MCFS feature list.

Classification algorithm	Number of features	ACC	MCC
Random forest	106	0.803	0.681
Random forest	31	0.713	0.575
PART	89	0.555	0.360

3.5 Comparison of Optimal Classifiers Without SMOTE

In the IFS method, we employed SMOTE to reduce the influence of imbalanced problem. To elaborate the utility of SMOTE, the

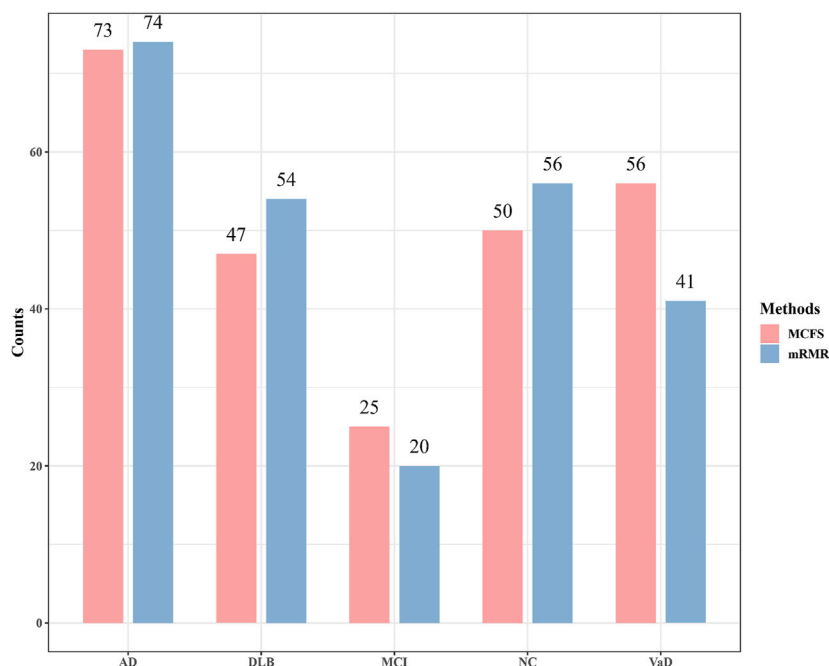


FIGURE 5 | Number of rules generated by the optimal PART classifiers based on mRMR and MCFS feature lists. AD, VaD, DLB, MCI, and NC stand for Alzheimer's disease, Vascular dementia, Dementia with Lewy bodies, Mild cognitive impairment and Normal control, respectively.

RF and PART classifiers mentioned in **Sections 3.2, 3.3** were tested when SMOTE was not adopted. All classifiers were assessed by 10-fold cross-validation. The ACCs and MCCs of these classifiers are listed in **Table 6**. Compared with the ACCs and MCCs listed in **Tables 2, 3**, MCC greatly decreased by at least 19%, even over 30% for the optimal RF classifiers. The ACC also decreased, but the degree was much smaller than that of the MCC. As the dataset was imbalanced, classifiers directly built on such dataset may be apt to the major classes (AD and NC in this study). Individual accuracies on these classes may be high, whereas individual accuracies on other classes may be low. The individual accuracies shown **Figure 6** confirmed this fact. The individual accuracies on AD were very high, followed by those on NC, whereas the individual accuracies on other three classes were very low, even zero. By employing SMOTE, the individual accuracies on AD decreased and those on other classes greatly increased, improving the entire performance of the classifiers. All these indicated the utility of the SMOTE.

4 DISCUSSION

The alteration of miRNA expression has been shown to relate with many pathological processes, including nervous system disorders. In this study, using the expression data of serum miRNAs, two optimal classifiers were constructed with high accuracy to identify the expression features of miRNAs through mRMR and MCFS method. We identified several putative miRNA biomarkers, which displayed strong relevance to the classification, suggesting that these miRNAs have specific

effect in different types of neurodegenerative diseases. Additionally, the optimal PART classifiers yielded by mRMR and MCFS feature lists were then applied to generate 245 and 251 decision rules, respectively, which can classify each sample into one of five categories, namely, AD, VaD, DLB, MCI, and NC. In this section, we mainly focused on several optimal and common features identified both by mRMR and MCFS methods, considering that common features are much more important in the classification. We examined the selected features and decision rules for the function and target genes of each miRNA using miRBase, an online database of miRNA sequences and annotation (Kozomara et al., 2018). For some miRNAs that have never been reported, we conducted bioinformatic analysis using miRDB for miRNA target prediction and functional annotation (Liu and Wang, 2019). Through literature review, several pieces of experimental evidence have been found to support the reliability of our prediction.

4.1 Analysis of Decision Rules Identified by mRMR and MCFS Methods

The most impactful feature in our computational analysis is miR-3184-5p, the mature miRNA product originating from the stem-loop precursor miRNA through cleavage by ribonuclease. As demonstrated by miRNA array experiment in multiple system atrophy disorders, a downregulated expression of miR-3184-5p was found in the FFPE sample of pons compared with controls, which indicates that this miRNA molecule plays an important role in normal brain development and may contribute

TABLE 4 | Some important rules extracted by the optimal PART classifier under the mRMR feature list.

Index	Decision Rules	Class
1	(hsa-miR-6088 \leq 10.1065) & (hsa-miR-520f-5 > 2.0854) & (hsa-miR-6836-3 \leq 8.6821) & (hsa-miR-6811-5 > 1.8782) & (hsa-miR-4667-5 > 6.1925) & (hsa-miR-6823-5 > 1.8811) & (hsa-miR-7851-3 > 5.1826) & (hsa-miR-4667-5 \leq 7.1244) & (hsa-miR-6756-5 \leq 8.7714)	Normal control
2	(hsa-miR-6088 \leq 9.9516) & (hsa-miR-4327 > 7.8591) & (hsa-miR-6861-5 > 6.5728) & (hsa-miR-4485-5 \leq 6.5037) & (hsa-miR-3622a-3 > 4.5067) & (hsa-miR-6875-5 \leq 10.0546) & (hsa-miR-7854-3 \leq 4.8701)	Alzheimer's disease
3	(hsa-miR-208a-5 > 5.8741) & (hsa-miR-548f-3 \leq 2.1097) & (hsa-miR-4667-5 > 6.7261) & (hsa-miR-6761-3 > 4.7880) & (hsa-miR-520f-5 \leq 1.8849)	Vascular dementia
4	(hsa-miR-208a-5 > 5.8741) & (hsa-miR-548f-3 \leq 2.1097) & (hsa-miR-4649-5 > 10.8160) & (hsa-miR-3622a-3 \leq 4.4907) & (hsa-miR-6070 > 1.8843) & (hsa-miR-663b \leq 8.7018)	Dementia with lewy bodies
5	(hsa-miR-520f-5 \leq 1.8945) & (hsa-miR-6840-3 \leq 7.6738) & (hsa-miR-185-5 \leq 2.9551)	Mild cognitive impairment

TABLE 5 | Some important rules extracted by the optimal PART classifier under the MCFS feature list.

Index	Decision rules	Class
1	(hsa-miR-6088 \leq 10.1065) & (hsa-miR-520f-5 > 2.0854) & (hsa-miR-6836-3 \leq 8.6821) & (hsa-miR-6811-5 > 1.8782) & (hsa-miR-4667-5 > 6.1925) & (hsa-miR-4746-3 \leq 7.4409) & (hsa-miR-3917 > 5.1453) & (hsa-miR-6070 \leq 2.9233) & (hsa-miR-6869-3 > 1.8805)	Normal control
2	(hsa-miR-6088 \leq 9.9516) & (hsa-miR-4327 > 7.8591) & (hsa-miR-1292-3 > 4.0332) & (hsa-miR-6861-5 > 6.5728) & (hsa-miR-125b-1-3 \leq 4.7145) & (hsa-miR-128-1-5 > 7.0405) & (hsa-miR-7854-3 \leq 4.8762) & (hsa-miR-6088 \leq 9.7663) & (hsa-miR-4506 \leq 3.6756)	Alzheimer's disease
3	(hsa-miR-520f-5 \leq 1.8945) & (hsa-miR-4485-3 > 1.8928) & (hsa-miR-3184-5 \leq 8.4938) & (hsa-miR-4496 > 1.8938) & (hsa-miR-6756-5 \leq 8.5013) & (hsa-miR-548f-3 \leq 1.8935) & (hsa-miR-6822-5 > 3.4091) & (hsa-miR-4472 \leq 6.3202) & (hsa-miR-1914-5 \leq 4.1342) & (hsa-miR-6776-3 > 4.0568) & (hsa-miR-548o-3 > 1.8798)	Vascular dementia
4	(hsa-miR-208a-5 > 5.8741) & (hsa-miR-548f-3 \leq 2.1097) & (hsa-miR-4667-5 \leq 6.7261) & (hsa-miR-4649-5 > 10.8290) & (hsa-miR-195-3 > 1.8967)	Dementia with lewy bodies
5	(hsa-miR-520f-5 \leq 1.8945) & (hsa-miR-4485-3 > 1.8928) & (hsa-miR-1254 > 6.9170) & (hsa-miR-197-5 > 7.3729)	Mild cognitive impairment

in the prevention of neurodegenerative disorders (Wakabayashi et al., 2016). In another research of spinocerebellar ataxia type 3 (SCA3), which is known as a highly heterogeneous neurodegenerative disorder, significantly downregulated expression of miR-3184 was observed in plasma from SCA3 patients compared with healthy controls (Hou et al., 2019). Therefore, we concluded that miR-3184-5p is necessary for the normal function of the brain, and the depletion of this molecule will lead to certain neurodegenerative disorders. Consistent with this finding, several decision rules in which miR-3184-5p is implicated show similar prediction that low expression levels of miR-3184-5p indicate AD and VaD categories, while relatively high expression levels indicate healthy controls.

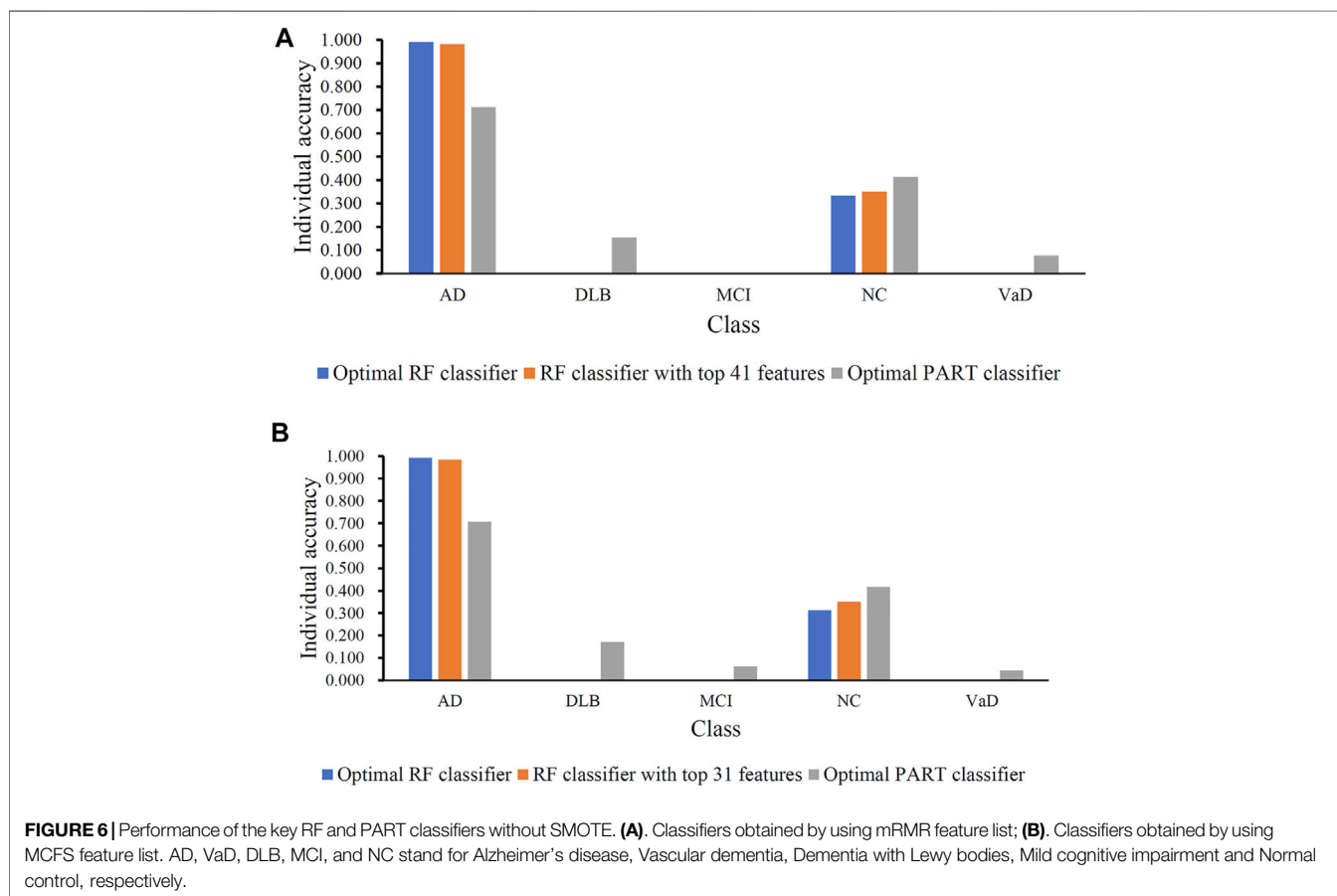
In many decision rules that indicate the AD category, a relatively high expression of miR-6088 is required for the classification. Although little has been known about this miRNA, we found a report that miR-6088 displays a significantly upregulated expression in patients with stroke compared with NC (Gui et al., 2019). Considering that stroke is a brain disease induced by deficient blood supply and will lead to nervous system injury, we inferred that miR-6088 may also participate in the process of neurodegeneration. Additionally,

miR-6088 was identified as one of the differentially methylated genes with high relevance to Parkinson's disease and neurodegeneration (Marsh et al., 2016), which provides strong support for the crucial role of miR-6088 in pathological processes of the nervous system.

Another important miRNA (miR-4327) is significantly associated with dementia, especially AD, through literature review. In the decision rules, we found that high expression of miR-4327 will lead to the classification of dementia, while relatively low expression indicates the normal cohort. As demonstrated by a miRNA expression profile experiment with Down syndrome, the expression level of miR-4327 was significantly higher in the case group than in the control group, suggesting that dysregulated miR-4327 may be related to abnormal development (Karaca et al., 2018). Individuals with Down syndrome usually show characteristics of damaged brain and intellectual disability, suggesting that miR-4327 affects brain development and results in several pathological processes including neurodegeneration. Moreover, using miRDB website tools, we found that the *OTUD1* gene is predicted as one of the target genes of miR-4327. *OTUD1* encodes a deubiquitinase, and mutations in this gene were reported to be associated with the

TABLE 6 | Performance of key classifiers without SMOTE.

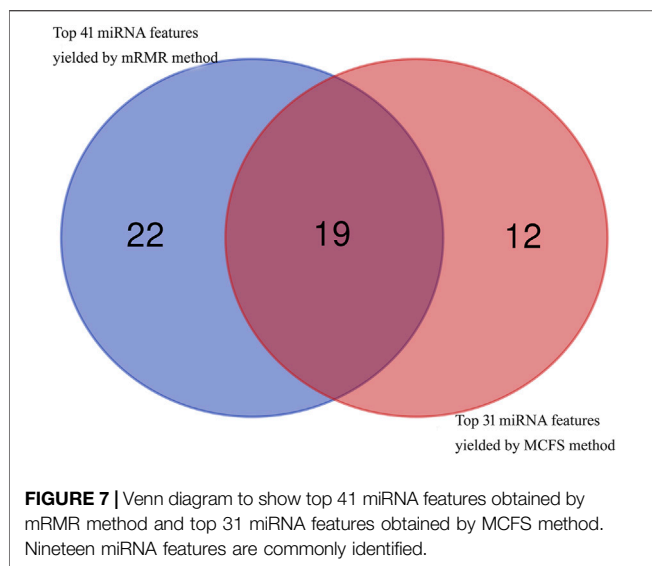
Feature selection method	Classification algorithm	Number of features	ACC	MCC
mRMR	Random forest	106	0.691	0.323
	Random forest	41	0.690	0.313
	PART	72	0.550	0.158
MCFS	Random forest	106	0.690	0.319
	Random forest	31	0.691	0.317
	PART	89	0.547	0.162



development of neurological phenotypes including ataxia with cerebellar atrophy and dementia (De Roux et al., 2016). On the basis of this finding, *OTUD1* is necessary for the normal neurological function, while excessive miR-4327 levels may inhibit *OTUD1* transcription and break the normal expression status. Therefore, the high level of miR-4327 is a risk indicator of dementia, which is consistent with our prediction model.

The high expression levels of miR-208a-5p display a strong indication to the categories of dementia in decision rules, suggesting that this miRNA plays a potential role in the associated processes. Several studies have described the role of miR-208a in cardiovascular diseases; for example, circulating levels of miR-208a are significantly elevated in patients with acute coronary syndrome (De Rosa et al., 2011). MiR-208a

was undetectable in the blood from healthy individuals, while upregulated expression was observed in the plasma of patients with acute myocardial infarction (Wang et al., 2010). Transgenic overexpression of miR-208a in heart tissue led to hypertrophic growth and arrhythmias in mice (Callis et al., 2009), providing reliable experimental evidence regarding the key function of miR-208a in cardiovascular diseases. Healthy brain functioning is dependent on adequate blood supply, while certain vascular diseases will cause brain injury such as VaD. We inferred that high expression of miR-208a first induces disorders in the vascular system that gradually develop into VaD, which is consistent with the decision rules. Our study is the first to present the role of miR-208a in neurodegenerative diseases, and this will contribute to the clinical diagnosis of dementia.



The high expression of miR-520f, one of the identified features implicated in both decision rules, indicates dementia. MiR-520f was found to be significantly increased in the CSF of patients with Huntington's disease compared with controls, suggesting that miR-520f can be used as a CSF biomarker for evaluating treatments (Reed et al., 2018). Huntington's disease is a neurodegenerative disease typically diagnosed in midlife, and this disease shares similar neuropathologic phenotypes to dementia. Thus, we inferred that an elevated level of miR-520f may also influence the pathologic processes of dementia. In addition, miR-520f is also significantly upregulated in multiple system atrophy, and its expression is negatively correlated with the target gene *AKT3* (Kim et al., 2019). *AKT3* has been reported to be related to neuronal insulin resistance in neurodegenerative diseases (Schubert et al., 2004). Taken together, we concluded that miR-520f acts as a transcriptional inhibitor of *AKT3*, and *AKT3* reduction will cause the neuropathologic processes of dementia.

The expression level of miR-1227 can be efficiently used to distinguish the types of dementia and NC in the prediction model, which suggests that miR-1227 is another important functional molecule involved in neurodegeneration. On the basis of a rabbit AD model, the specific expression pattern of miR-1227 was observed, which showed similar profiles to those observed in human AD samples (Liu et al., 2014), indicating the potential role of miR-1227 in AD and other dementia diseases. A recent study reported that *LINC00639*, the target gene of miR-1227, was downregulated in HIV-associated dementia (HAD), a kind of cognitive impairment induced by HIV infection (Li et al., 2018). Even though the pathogenesis of HAD remains unclear, the aberration of certain miRNAs such as miR-1227 can provide novel direction for further research. Similarly, increased expression of miR-1227 was detected in CSF from patients with intracerebral hemorrhage (Shi et al., 2018). In summary, miR-1227 displays distinct expression profiles in many brain injury disorders or dementia, suggesting that it may be an auxiliary diagnostic biomarker for these diseases. These findings confirmed the reliability of our decision rules and

implied that the expression criteria of identified miRNAs can be used in disease risk classification and clinical diagnostic.

4.2 Analysis of the Top Features Identified by mRMR and MCFS Methods

In addition to the quantitative analysis discussed above, we have also identified many miRNAs that can be used as indicators for dementia. As the RF classifier with less features provided slight lower performance than the corresponding optimal RF classifier, miRNA features used in these two RF classifiers with less features were investigated in this section. Based on the mRMR feature list, 41 miRNA features were obtained, whereas 31 miRNA features were accessed from the MCFS feature list. After taking the union of these two feature subsets, 53 different miRNA features were obtained, which are listed in **Supplementary Table S5**. A Venn diagram was plotted to show the distribution of these miRNA features in two feature sets, as shown in **Figure 7**. It can be observed that nineteen miRNA features were commonly identified. These features were thought to be more reliable than others. Some of them were discussed as follows.

MiR-4649-5p exhibits an upregulated expression profile in neurodegenerative disorders (Viswambharan et al., 2017). In amyotrophic lateral sclerosis (ALS), which is a fatal neurodegenerative disease, increasing concentration of miR-4649-5p was observed in the plasma of ALS patients, suggesting that this miRNA can be used in the diagnosis of ALS (Takahashi et al., 2015). On the basis of the miRDB database, we found that miR-4649-5p can target *INSYN2*, a protein coding gene implicated in inhibitory synapses. This synaptic inhibition is fundamental for the functioning of the central nervous system, shaping and orchestrating the flow of information through neuronal networks to generate a precise neural code (Uezu et al., 2016). Therefore, miR-4649-5p plays an important role in neural development, which confirms the reliability of our computational analysis.

MiR-3181 is one of the most related features in our computational analysis, and many studies indicate the close association between this miRNA and vascular diseases. Significantly upregulated miR-3181 was detected in endothelial cells treated with acrolein, which is a component of cigarette smoke and has been implicated in the development of vascular disease, suggesting that this miRNA may improve the diagnosis of vascular disease induced by environmental pollutants (Lee et al., 2015). As discussed previously, the development of vascular disease may be accompanied by brain injury such as VaD, suggesting the role of miR-3181 in dementia. The *TCL1B* gene, which is predicted as one target of miR-3181, showed significant differential expression between Parkinson's disease patients and NC (Infante et al., 2015). *TCL1B* is also an activator of Akt, a kinase involved in neuron survival (Hashimoto et al., 2013), and abnormal Akt signaling has been reported to induce dopamine neuron degeneration (Greene et al., 2011).

The expression profile of miR-128-1-5p is also a strong indicator for the classification in our analysis. MiR-128 is a neuronally enriched miRNA that plays a crucial role in neuronal differentiation and survival (Guidi et al., 2010). The

expression of miR-128 is increased in the hippocampus of AD patients (Lukiw and Pogue, 2007). In addition, upregulated miR-128 can cause a decreased expression of *SNAP25* and lead to the perturbation of neuronal activity (Eletto et al., 2008). These results support the role of miR-128 in neurodegenerative disease. Using RNA sequencing techniques, miR-128 showed decreased expression in Huntington's disease (Martí et al., 2010). MiR-128 displays distinct expression patterns in different neurodegenerative diseases, indicating its potential capability of distinguishing varied disease subtypes and confirming the ability of our prediction model to classify different dementias.

Besides above commonly identified miRNAs, some miRNAs identified by exact one feature selection method (mRMR or MCFS) were also quite essential. For example, miR-185-5p is identified as one of the most relevant features that contribute to the classification. MiR-185 has been suggested to participate in the pathogenesis of major depression, a psychosocial impairment, and finally lead to suicide. It was thought to influence neuronal and circuit formation by regulating target downstream gene, *TrkB-T1*, which has been associated with suicidal behavior (Serafini et al., 2014). This finding suggests the key role of miR-185-5p involved in nervous system development, physiology, and diseases.

In this section, we discussed the verified or speculative functions of miRNAs identified by our computational analysis. All these miRNAs have been confirmed to contribute to distinguishing patients with dementia from healthy and varied disease subtypes. Strikingly, many miRNAs related to vascular diseases usually play a putative role in neurodegenerative diseases. This finding suggests the interaction between these two distinct disease types. In summary, this study presented a novel computational approach to identify potential biomarkers for diagnosis and therapy, and also set up a basic research foundation for further studies on the detailed pathological mechanism of miRNAs in neurodegenerative diseases.

5 CONCLUSION

We employed a computational analysis approach to discovery key miRNA properties that differentiate normal and neurodegenerative disease subgroups in this work. The Boruta feature selection method was utilized to exclude unnecessary miRNA features, and then mRMR and MCFS were used to rank the remaining features. A series of feature subsets was generated from these ranked feature lists using the IFS method, and the sample data containing these feature subsets was used to train the RF and PART classifiers. As a result, the optimal miRNA biomarker set was identified on the basis of the evaluation metrics of classifiers under varying number of features,

and the classification rules were extracted from the optimal PARTs. Finally, the relationship between candidate features including hsa-miR-3184-5p, has-miR-6088, and has-miR-4649 and neurodegenerative diseases was validated in recent studies, confirming the efficacy of our methods and establishing the groundwork for further investigation into the underlying pathogenic mechanisms of miRNAs in neurodegenerative illnesses.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120584>.

AUTHOR CONTRIBUTIONS

TH and Y-DC designed the study. SD, LC and KF performed the experiments. ZL and WG analyzed the results. ZL, WG and SD wrote the manuscript. All authors contributed to the research and reviewed the manuscript.

FUNDING

This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDA26040304, XDB38050200), National Key R&D Program of China (2018YFC0910403), the Fund of the Key Laboratory of Tissue Microenvironment and Tumor of Chinese Academy of Sciences (202002).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.880997/full#supplementary-material>

Supplementary Table S1 | Ranked feature lists analyzed by Boruta feature selection method, mRMR, and MCFS.

Supplementary Table S2 | Performance of different classifiers with different numbers of features obtained by IFS method for mRMR feature list.

Supplementary Table S3 | Performance of different classifiers with different numbers of features obtained by IFS method for MCFS feature list.

Supplementary Table S4 | Rule sets extracted from the optimal PART classifiers based on mRMR and MCFS feature lists.

Supplementary Table S5 | Different miRNAs identified by IFS method on mRMR or MCFS feature lists.

REFERENCES

Asanomi, Y., Shigemizu, D., Akiyama, S., Sakurai, T., Ozaki, K., Ochiya, T., et al. (2021). Dementia Subtype Prediction Models Constructed by

Penalized Regression Methods for Multiclass Classification Using Serum microRNA Expression Data. *Sci. Rep.* 11, 20947. doi:10.1038/s41598-021-00424-1

Bekris, L. M., Lutz, F., Montine, T. J., Yu, C. E., Tsuang, D., Peskind, E. R., et al. (2013). MicroRNA in Alzheimer's Disease: an Exploratory Study in Brain,

- Cerebrospinal Fluid and Plasma. *Biomarkers* 18, 455–466. doi:10.3109/1354750x.2013.814073
- Breiman, L. (2001). Random Forests. *Machine Learn.* 45, 5–32. doi:10.1023/a:1010933404324
- Callis, T. E., Pandya, K., Seok, H. Y., Tang, R.-H., Tatsuguchi, M., Huang, Z.-P., et al. (2009). MicroRNA-208a Is a Regulator of Cardiac Hypertrophy and Conduction in Mice. *J. Clin. Invest.* 119, 2772–2786. doi:10.1172/jci36154
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Jair* 16, 321–357. doi:10.1613/jair.953
- Chen, L., Li, Z., Zhang, S., Zhang, Y. H., Huang, T., and Cai, Y. D. (2022). Predicting RNA 5-methylcytosine Sites by Using Essential Sequence Features and Distributions. *Biomed. Res. Int.* 2022, 4035462. doi:10.1155/2022/4035462
- Chen, L., Wang, S., Zhang, Y.-H., Li, J., Xing, Z.-H., Yang, J., et al. (2017). Identify Key Sequence Features to Improve CRISPR sgRNA Efficacy. *IEEE Access* 5, 26582–26590. doi:10.1109/access.2017.2775703
- Chen, W., Chen, L., and Dai, Q. (2021). iMPT-FDNPL: Identification of Membrane Protein Types with Functional Domains and a Natural Language Processing Approach. *Comput. Math. Methods Med.* 2021, 7681497. doi:10.1155/2021/7681497
- Cogswell, J. P., Ward, J., Taylor, I. A., Waters, M., Shi, Y., Cannon, B., et al. (2008). Identification of miRNA Changes in Alzheimer's Disease Brain and CSF Yields Putative Biomarkers and Insights into Disease Pathways. *Jad* 14, 27–41. doi:10.3233/jad-2008-14103
- Coolen, M., Katz, S., and Bally-Cuif, L. (2013). miR-9: a Versatile Regulator of Neurogenesis. *Front. Cel. Neurosci.* 7, 220. doi:10.3389/fncel.2013.00220
- Danborg, P. B., Simonsen, A. H., Waldemar, G., and Heegaard, N. H. H. (2014). The Potential of microRNAs as Biofluid Markers of Neurodegenerative Diseases - a Systematic Review. *Biomarkers* 19, 259–268. doi:10.3109/1354750x.2014.904001
- De Meyer, G., Shapiro, F., Vanderstichele, H., Vanmechelen, E., Engelborghs, S., De Deyn, P. P., et al. (2010). Diagnosis-independent Alzheimer Disease Biomarker Signature in Cognitively normal Elderly People. *Arch. Neurol.* 67, 949–956. doi:10.1001/archneurol.2010.179
- De Rosa, S., Fichtlscherer, S., Lehmann, R., Assmus, B., Dimmeler, S., and Zeiher, A. M. (2011). Transcoronary Concentration Gradients of Circulating microRNAs. *Circulation* 124, 1936–1944. doi:10.1161/circulationaha.111.037572
- De Roux, N., Carel, J.-C., and Léger, J. (2016). “Congenital Hypogonadotropic Hypogonadism: a Trait Shared by Several Complex Neurodevelopmental Disorders,” in *Puberty from Bench to Clinic* (Karger Publishers), 72–86.
- Ding, S., Wang, D., Zhou, X., Chen, L., Feng, K., Xu, X., et al. (2022). Predicting Heart Cell Types by Using Transcriptome Profiles and a Machine Learning Method. *Life* 12, 228. doi:10.3390/life12020228
- Draminski, M., Rada-Iglesias, A., Enroth, S., Wadelius, C., Koronacki, J., and Komorowski, J. (2007). Monte Carlo Feature Selection for Supervised Classification. *Bioinformatics* 24, 110–117. doi:10.1093/bioinformatics/btm486
- Eletto, D., Russo, G., Passiatore, G., Del Valle, L., Giordano, A., Khalili, K., et al. (2008). Inhibition of SNAP25 Expression by HIV-1 Tat Involves the Activity of Mir-128a. *J. Cel. Physiol.* 216, 764–770. doi:10.1002/jcp.21452
- Frank, E., Hall, M., Trigg, L., Holmes, G., and Witten, I. H. (2004). Data Mining in Bioinformatics Using Weka. *Bioinformatics* 20, 2479–2481. doi:10.1093/bioinformatics/bth261
- Frank, E., and Witten, I. H. (1998). “Generating Accurate Rule Sets without Global Optimization,” in *Computer Science Working Papers* (Hamilton, New Zealand: University of Waikato, Department of Computer Science).
- Geekiyana, H., and Chan, C. (2011). MicroRNA-137/181c Regulates Serine Palmitoyltransferase and in Turn Amyloid, Novel Targets in Sporadic Alzheimer's Disease. *J. Neurosci.* 31, 14820–14830. doi:10.1523/jneurosci.3883-11.2011
- Gilad, S., Meiri, E., Yogev, Y., Benjamin, S., Lebanony, D., Yerushalmi, N., et al. (2008). Serum microRNAs Are Promising Novel Biomarkers. *PLoS one* 3, e3148. doi:10.1371/journal.pone.0003148
- Gorodkin, J. (2004). Comparing Two K-Category Assignments by a K-Category Correlation Coefficient. *Comput. Biol. Chem.* 28, 367–374. doi:10.1016/j.compbiolchem.2004.09.006
- Greene, L. A., Levy, O., and Malagelada, C. (2011). Akt as a Victim, Villain and Potential Hero in Parkinson's Disease Pathophysiology and Treatment. *Cell Mol Neurobiol* 31, 969–978. doi:10.1007/s10571-011-9671-8
- Gui, Y., Xu, Z., Jin, T., Zhang, L., Chen, L., Hong, B., et al. (2019). Using Extracellular Circulating microRNAs to Classify the Etiological Subtypes of Ischemic Stroke. *Transl. Stroke Res.* 10, 352–361. doi:10.1007/s12975-018-0659-2
- Guidi, M., Muiños-Gimeno, M., Kagerbauer, B., Martí, E., Estivill, X., and Espinosa-Parrilla, Y. (2010). Overexpression of miR-128 Specifically Inhibits the Truncated Isoform of NTRK3 and Upregulates BCL2 in SH-Sy5y Neuroblastoma Cells. *BMC Mol. Biol* 11, 95. doi:10.1186/1471-2199-11-95
- Hanchuan Peng, H., Fuhui Long, F., and Ding, C. (2005). Feature Selection Based on Mutual Information Criteria of max-dependency, max-relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Machine Intell.* 27, 1226–1238. doi:10.1109/tpami.2005.159
- Hashimoto, M., Suizu, F., Tokuyama, W., Noguchi, H., Hirata, N., Matsuda-Lennikov, M., et al. (2013). Protooncogene TCL1b Functions as an Akt Kinase Co-activator that Exhibits Oncogenic Potency *In Vivo*. *Oncogenesis* 2, e70. doi:10.1038/oncsis.2013.30
- Hou, X., Gong, X., Zhang, L., Li, T., Yuan, H., Xie, Y., et al. (2019). Identification of a Potential Exosomal Biomarker in Spinocerebellar Ataxia Type 3/ Machado-Joseph Disease. *Epigenomics* 11, 1037–1056. doi:10.2217/epi-2019-0081
- Iliffe, S., Robinson, L., Brayne, C., Goodman, C., Rait, G., Manthorpe, J., et al. (2009). Primary Care and Dementia: 1. Diagnosis, Screening and Disclosure. *Int. J. Geriatr. Psychiatry* 24, 895–901. doi:10.1002/gps.2204
- Infante, J., Prieto, C., Sierra, M., Sánchez-Juan, P., González-Aramburu, I., Sánchez-Quintana, C., et al. (2015). Identification of Candidate Genes for Parkinson's Disease through Blood Transcriptome Analysis in LRRK2-G2019s Carriers, Idiopathic Cases, and Controls. *Neurobiol. Aging* 36, 1105–1109. doi:10.1016/j.neurobiolaging.2014.10.039
- Karaca, E., Aykut, A., Ertürk, B., Durmaz, B., Güler, A., Büke, B., et al. (2018). Diagnostic Role of MicroRNA Expression Profile in the Prenatal Amniotic Fluid Samples of Pregnant Women with Down Syndrome. *Balkan Med. J.* 35, 163–166. doi:10.4274/balkanmedj.2017.0511
- Kim, T., Valera, E., and Desplats, P. (2019). Alterations in Striatal microRNA-mRNA Networks Contribute to Neuroinflammation in Multiple System Atrophy. *Mol. Neurobiol.* 56, 7003–7021. doi:10.1007/s12035-019-1577-3
- Kohavi, R. (1995). “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” in *International Joint Conference on Artificial Intelligence* (Lawrence Erlbaum Associates), 1137–1145.
- Kostopoulou, O., Delaney, B. C., and Munro, C. W. (2008). Diagnostic Difficulty and Error in Primary Care-Aa Systematic Review. *Fam. Pract.* 25, 400–413. doi:10.1093/fampra/cm071
- Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2018). miRBase: from microRNA Sequences to Function. *Nucleic Acids Res.* 47, D155–D162. doi:10.1093/nar/gky1141
- Kursa, M. B., and Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *J. Stat. Softw.* 36, 1–13. doi:10.18637/jss.v036.i11
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of Novel Genes Coding for Small Expressed RNAs. *Science* 294, 853–858. doi:10.1126/science.1064921
- Lee, S. E., Yang, H., Son, G. W., Park, H. R., Cho, J.-J., Ahn, H.-J., et al. (2015). Identification and Characterization of MicroRNAs in Acrolein-Stimulated Endothelial Cells: Implications for Vascular Disease. *Biochip J.* 9, 144–155. doi:10.1007/s13206-015-9303-3
- Li, D., Bao, P., Yin, Z., Sun, L., Feng, J., He, Z., et al. (2018). Exploration of the Involvement of LncRNA in HIV-Associated Encephalitis Using Bioinformatics. *PeerJ* 6, e5721. doi:10.7717/peerj.5721
- Li, H., Habes, M., Wolk, D. A., Fan, Y., and Initiative, A. S. D. N. (2019). A Deep Learning Model for Early Prediction of Alzheimer's Disease Dementia Based on Hippocampal Magnetic Resonance Imaging Data. *Alzheimer's & Dement.* 15, 1059–1070. doi:10.1016/j.jalz.2019.02.007
- Li, X., Lu, L., Lu, L., and Chen, L. (2022a). Identification of Protein Functions in Mouse with a Label Space Partition Method. *Mbe* 19, 3820–3842. doi:10.3934/mbe.2022176
- Li, Z., Wang, D., Liao, H., Zhang, S., Guo, W., Chen, L., et al. (2022b). Exploring the Genomic Patterns in Human and Mouse Cerebellums via Single-Cell

- Sequencing and Machine Learning Method. *Front. Genet.* 13, 857851. doi:10.3389/fgene.2022.857851
- Licher, S., Leening, M. J. G., Yilmaz, P., Wolters, F. J., Heeringa, J., Bindels, P. J. E., et al. (2019). Development and Validation of a Dementia Risk Prediction Model in the General Population: an Analysis of Three Longitudinal Studies. *Ajp* 176, 543–551. doi:10.1176/appi.ajp.2018.18050566
- Liu, H., and Setiono, R. (1998). Incremental Feature Selection. *Appl. Intelligence* 9, 217–230. doi:10.1023/a:1008363719778
- Liu, Q. Y., Chang, M. N., Lei, J. X., Koukiekolo, R., Smith, B., Zhang, D., et al. (2014). Identification of microRNAs Involved in Alzheimer's Progression Using a Rabbit Model of the Disease. *Am. J. Neurodegener. Dis.* 3, 33–44.
- Liu, W., and Wang, X. (2019). Prediction of Functional microRNA Targets by Integrative Modeling of microRNA Binding and Target Expression Data. *Genome Biol.* 20, 18. doi:10.1186/s13059-019-1629-z
- Lukiw, W. J., and Pogue, A. I. (2007). Induction of Specific Micro RNA (miRNA) Species by ROS-Generating Metal Sulfates in Primary Human Brain Cells. *J. Inorg. Biochem.* 101, 1265–1269. doi:10.1016/j.jinorgbio.2007.06.004
- Marsh, A. G., Cottrell, M. T., and Goldman, M. F. (2016). Epigenetic DNA Methylation Profiling with MSRE: A Quantitative NGS Approach Using a Parkinson's Disease Test Case. *Front. Genet.* 7, 191. doi:10.3389/fgene.2016.00191
- Martí, E., Pantano, L., Bañez-Coronel, M., Llorens, F., Miñones-Moyano, E., Porta, S., et al. (2010). A Myriad of miRNA Variants in Control and Huntington's Disease Brain Regions Detected by Massively Parallel Sequencing. *Nucleic Acids Res.* 38, 7219–7235.
- Mckeith, I. G., Galasko, D., Kosaka, K., Perry, E. K., Dickson, D. W., Hansen, L. A., et al. (1996). Consensus Guidelines for the Clinical and Pathologic Diagnosis of Dementia with Lewy Bodies (DLB). *Neurology* 47, 1113–1124. doi:10.1212/wnl.47.5.1113
- Mistur, R., Mosconi, L., Santi, S. D., Guzman, M., Li, Y., Tsui, W., et al. (2009). Current Challenges for the Early Detection of Alzheimer's Disease: Brain Imaging and CSF Studies. *J. Clin. Neurol.* 5, 153–166. doi:10.3988/jcn.2009.5.4.153
- Prince, M., Comas-Herrera, A., Knapp, M., Guerchet, M., and Karagiannidou, M. (2016). World Alzheimer Report 2016: Improving Healthcare for People Living with Dementia: Coverage, Quality and Costs Now and in the Future. *Alzheimer's Dis. Int.* 1, 1. doi:10.13140/RG.2.2.22580.04483
- Reed, E. R., Latourelle, J. C., Bockholt, J. H., Bregu, J., Smock, J., Paulsen, J. S., et al. (2018). MicroRNAs in CSF as Prodromal Biomarkers for Huntington Disease in the PREDICT-HD Study. *Neurology* 90, e264–e272. doi:10.1212/wnl.0000000000004844
- Robinson, L., Tang, E., and Taylor, J.-P. (2015). Dementia: Timely Diagnosis and Early Intervention. *Bmj* 350, h3029. doi:10.1136/bmj.h3029
- Schubert, M., Gautam, D., Surjo, D., Ueki, K., Baudler, S., Schubert, D., et al. (2004). Role for Neuronal Insulin Resistance in Neurodegenerative Diseases. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3100–3105. doi:10.1073/pnas.0308724101
- Serafini, G., Pompili, M., Hansen, K. F., Obrietan, K., Dwivedi, Y., Shomron, N., et al. (2014). The Involvement of microRNAs in Major Depression, Suicidal Behavior, and Related Disorders: a Focus on miR-185 and miR-491-3p. *Cel Mol Neurobiol* 34, 17–30. doi:10.1007/s10571-013-9997-5
- Seshadri, S., Fitzpatrick, A. L., Ikram, M. A., Destefano, A. L., Gudnason, V., Boada, M., et al. (2010). Genome-wide Analysis of Genetic Loci Associated with Alzheimer Disease. *Jama* 303, 1832–1840. doi:10.1001/jama.2010.574
- Shi, Q., Ge, D., Yang, Q., Wang, L., and Fu, J. (2018). MicroRNA Profiling of Cerebrospinal Fluid from Patients with Intracerebral Haemorrhage. *Front. Lab. Med.* 2, 141–145. doi:10.1016/j.flm.2019.07.001
- Shigemizu, D., Akiyama, S., Asanomi, Y., Borojevich, K. A., Sharma, A., Tsunoda, T., et al. (2019a). Risk Prediction Models for Dementia Constructed by Supervised Principal Component Analysis Using miRNA Expression Data. *Commun. Biol.* 2, 77. doi:10.1038/s42003-019-0324-7
- Shigemizu, D., Akiyama, S., Asanomi, Y., Borojevich, K. A., Sharma, A., Tsunoda, T., et al. (2019b). A Comparison of Machine Learning Classifiers for Dementia with Lewy Bodies Using miRNA Expression Data. *BMC Med. Genomics* 12, 150. doi:10.1186/s12920-019-0607-3
- Shimoda, A., Li, Y., Hayashi, H., and Kondo, N. (2021). Dementia Risks Identified by Vocal Features via Telephone Conversations: A Novel Machine Learning Prediction Model. *PLoS one* 16, e0253988. doi:10.1371/journal.pone.0253988
- Stephan, B. C. M., Kurth, T., Matthews, F. E., Brayne, C., and Dufouil, C. (2010). Dementia Risk Prediction in the Population: Are Screening Models Accurate? *Nat. Rev. Neurol.* 6, 318–326. doi:10.1038/nrneuro.2010.54
- Takahashi, I., Hama, Y., Matsushima, M., Hirotsu, M., Kano, T., Hohzen, H., et al. (2015). Identification of Plasma microRNAs as a Biomarker of Sporadic Amyotrophic Lateral Sclerosis. *Mol. Brain* 8, 67. doi:10.1186/s13041-015-0161-7
- Tang, S., and Chen, L. (2022). iATC-NFMLP: Identifying Classes of Anatomical Therapeutic Chemicals Based on Drug Networks, Fingerprints and Multilayer Perceptron. *Curr. Bioinformatics* 1, 1. doi:10.2174/1574893617666220318093000
- Uezu, A., Kanak, D. J., Bradshaw, T. W. A., Soderblom, E. J., Catavero, C. M., Burette, A. C., et al. (2016). Identification of an Elaborate Complex Mediating Postsynaptic Inhibition. *Science* 353, 1123–1129. doi:10.1126/science.aag0821
- Viswambharan, V., Thanseem, I., Vasu, M. M., Poovathinal, S. A., and Anitha, A. (2017). miRNAs as Biomarkers of Neurodegenerative Disorders. *Biomarkers Med.* 11, 151–167. doi:10.2217/bmm-2016-0242
- Wakabayashi, K., Mori, F., Kakita, A., Takahashi, H., Tanaka, S., Utsumi, J., et al. (2016). MicroRNA Expression Profiles of Multiple System Atrophy from Formalin-Fixed Paraffin-Embedded Samples. *Neurosci. Lett.* 635, 117–122. doi:10.1016/j.neulet.2016.10.034
- Wang, G.-K., Zhu, J.-Q., Zhang, J.-T., Li, Q., Li, Y., He, J., et al. (2010). Circulating microRNA: a Novel Potential Biomarker for Early Diagnosis of Acute Myocardial Infarction in Humans. *Eur. Heart J.* 31, 659–666. doi:10.1093/eurheartj/ehq013
- Yang, Y., and Chen, L. (2022). Identification of Drug-Disease Associations by Using Multiple Drug and Disease Networks. *Cbio* 17, 48–59. doi:10.2174/1574893616666210825115406
- Zhao, X., Chen, L., and Lu, J. (2018). A Similarity-Based Method for Prediction of Drug Side Effects with Heterogeneous Information. *Math. Biosciences* 306, 136–144. doi:10.1016/j.mbs.2018.09.010

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Guo, Ding, Chen, Feng, Huang and Cai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.