Chemometrics and Intelligent Laboratory Systems 215 (2021) 104316

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Adaptive LASSO logistic regression based on particle swarm optimization for Alzheimer's disease early diagnosis

Xinchun Cui [a], Ruyi Xiao [a], Xiaoli Liu [b], Hong Qiao [c], Xiangwei Zheng [d], Yiquan Zhang [a], Jianzong Du [e,*]

[a] School of Computer Science, Qufu Normal University, 276826, Rizhao, China
[b] Department of Neurology, Affiliated Zhejiang Hospital, Zhejiang University School of Medicine, 310013, Hangzhou, China
[c] Business School, Shandong Normal University, 250014, Jinan, China
[d] School of Information Science and Engineering, Shandong Normal University, 250014, Jinan, China
[e] Department of Respiratory Medicine, Affiliated Zhejiang Hospital, Zhejiang University School of Medicine, 310013, Hangzhou, China

ARTICLE INFO

Keywords:
Adaptive LASSO
MRI image
Particle swarm optimization
Alzheimer's disease
Mild cognitive impairment

ABSTRACT

Accurate classification of Alzheimer's disease (AD) and its prodromal stage mild cognitive impairment (MCI) play key roles in computer-assisted intervention for the diagnosis of AD. However, not all features of AD data will lead to a good classification result, because there are always some unrelated and redundant features. To solve this problem, an adaptive LASSO logistic regression model based on particle swarm optimization(PSO-ALLR)is proposed. This algorithm consists of two stages. In the first stage, the particle swarm optimization (PSO) algorithm is used for global search to remove redundant features and reduces the computational time for the later stage. In the second stage, the adaptive LASSO serves as a local search to select the most relevant features for AD classification.We evaluate the performance of the proposed method on 197 subjects from the baseline MRI data of ADNI database. The proposed method achieves a classification accuracy of 96.27%, 84.81%, and 76.13%, for AD vs. HC, MCI vs. HC, and cMCI vs. sMCI, respectively.

1. Introduction

Alzheimer's disease (AD) is a neurological brain disease. The typical symptoms of the disease are cognitive and memory decline, which seriously affects people's daily lives [1]. According to statistics, there are about 50 million AD patients worldwide, and this number is still growing rapidly. It is estimated that this number will reach 13.8 million by 2050 [2].The rapid increase in the number of AD patients and other forms of dementia bring a major challenge to health and social care systems. Until today, there is still no effective drug to treat the disease. Early diagnosis and early intervention are the only methods that can be relied on. Therefore, the early diagnosis of ad has attracted the attention of scholars from all over the world. Mild cognitive impairment (MCI) is an early stage of AD. It is estimated that 40%–60% of individuals over the age of 58 with MCI have potential AD pathology. Approximately 15% of MCI patients convert to AD each year [3]. Accurate diagnosis of MCI and prediction of the risk process of its conversion to AD are essential. The early diagnosis and timely therapy of AD might be effective to delay the conversion of MCI to AD. The diagnosis of this stage is very difficult, because the clinical symptoms of MCI are not obvious.

In recent studies, machine learning methods based on MRI biomarkers are used for the diagnosis of AD [4–6]. Among numerous machine learning algorithms, logistic regression(LR) is a widely used discriminative method [7]. LR has a direct probabilistic interpretation. In addition to the class label information it can obtain direct classification probabilities [8]. However, the small number of samples and more features in AD data make it difficult for logistic regression to classify Alzheimer's disease. What's more, many irrelevant features have an impact on the accuracy of classification.

For the above problems, the sparse logistic regression models with different sparse penalty terms is proposed such as LASSO [9], SCAD [10] and Elastic net [11]. Regularization methods are an important embedded technique for feature selection and model learning at the same time [12, 13]. Most widely used regular terms include the least absolute shrinkage and selection operator (LASSO). Koh et al. [14] introduced the L_1 regularized logistic regression as a special case to solve large-scale problems. The LASSO can shrink the regression coefficients to zero, thereby selecting some important features simultaneously [15–17]. However,

* Corresponding author.
E-mail address: djzdjz@163.com (J. Du).

https://doi.org/10.1016/j.chemolab.2021.104316
Received 15 November 2020; Received in revised form 6 March 2021; Accepted 10 April 2021
Available online 16 April 2021
0169-7439/

LASSO may make selected features inconsistent in some cases and the estimated parameters may have some deviations. What's more, LASSO can't obtain the sufficiently sparse results. In this paper, we propose a novel method for AD early diagnosis. The method consists of two stages. In the first stage, PSO can remove redundant and irrelevant features and reduce the computational time for the later stage. PSO is a powerful global search method. It has lower cost and faster convergence speed [18]. In the second stage, adaptive LASSO is mainly used to optimize logistic regression, where adaptive weights can penalize different feature coefficients to make feature selection consistent in the L_1 penalty. Adaptive LASSO as a local search method can further determine which features are important among the remaining features and evaluate the model. In this stage, the highest discriminative features are selected. What's more, in our previous work, we mainly use $L_{1/2}$ regularization to sparsely optimize the logistic regression model. $L_{1/2}$ regularization can make the logistic regression produce sparse solutions, thereby selecting important features for classification [19]. However, although sparse logistic regression with $L_{1/2}$ regularization has achieved better classification results on AD, this method cannot address the highly correlated features. Compared with $L_{1/2}$ regularization, the newly proposed two-stage sparse logistic regression can select the optimal feature subset with high classification accuracy by combining PSO and adaptive LASSO regularization. Such method would not only save computational costs, but will also enable doctors to identify a small subset of features related to AD and target only a small number of features in designing less expensive experiments. Therefore, the main contributions of this article include: (1) The combination of PSO and adaptive LASSO logistic regression is proposed; (2) PSO is used for global feature search and Adaptive LASSO is trained for local feature filtering to choose the most relevant features of AD; (3) The proposed new algorithm is used in the early diagnosis of AD, and experiments and analysis are carried out. (4) The biomarkers closely related to AD are identified, which can assist in AD classification.

The rest of this work is arranged as follows. In the second part, we introduce the details of the experimental data. The methodology used in this article is described in the third part. The fourth part introduces the experimental results and discusses them further. The last part is the conclusion of this article.

1.1. ADNI database

The dataset used in our article is derived from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database(http://www.loni.ucla.edu/ADNI). ADNI was found in 2003 by the National Institute of Biomedical Imaging and Bioengineering. It is a non-profit organization [20]. ADNI provides unlimited data access and encourages researchers to develop potential methods for the analysis of AD. Magnetic resonance imaging(MRI) is a widely used imaging mode in the diagnosis and prediction of AD [21–23]. It is able to make better comparisons among different soft tissues. Therefore, we use MRI images for analysis. We selected MRI images of 197 subjects in the ADNI database, including 51 AD, 50 healthy controls (HC) and 96 MCI (including 51 converted MCI (cMCI) and 45 stable MCI (sMCI)). Table 1 presents detailed information about these subjects.

Table 1
Statistical information of subjects (mean standard ± deviation).

Diagnosis	Subjects	Age	Gender(F/M)	MMSE	CDR
AD	51	75.8 ± 7.5	23/28	23.6 ± 2.2	0.7 ± 0.3
HC	50	77.8 ± 6.8	27/23	28.8 ± 1.4	0.0 ± 0.0
cMCI	51	72.5 ± 6.5	26/25	26.7 ± 1.3	0.5 ± 0.0
sMCI	45	71.9 ± 7.6	20/25	27.3 ± 1.6	0.5 ± 0.0

Note:CDR: clinical dementia rating scale, 0 = no dementia, 0.5 = suspected dementia, 1 = mild dementia, MMSE: Concise mental state examination scale.

2. Methods

2.1. Image preprocessing and feature extraction

The MRI image downloaded from the ADNI database requires a series of image preprocessing and extracts the gray matter volume of 90 regions of interest as effective features. The specific image preprocessing and feature extraction process are shown in Fig. 1.

For each subject, T1-weighted MRI was first pre-processed by an anterior commissure-posterior commissure (AC-PC) correction using MIPAV software. Then, N3 algorithm was used to correct the intensity inhomogeneity followed by skull stripping and cerebellum removing. In addition, each MRI was segmented into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) by using VBM8 toolbox. Next, by warping the Automated Anatomical Labeling (AAL) template, for each subject, we parcellated the brain space into 90 regions of interest (ROIs). Lastly, we calculated the gray matter volume of 90 regions of interest as features.

2.2. Logistic regression model

As a special nonlinear model, logistic regression is mainly used to solve classification problems [24]. In our article, we implement LR for the binary classification problem. Supposing that $x_i = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in})^T$ is i-th sample vector of the matrix X. Define a classifier $f = e^x/1 + e^x$and the expression of the logistic regression model is:

$$\pi_i = p(y_i|x_i) = f\left(x_i^T\theta\right) = \frac{\exp\left(x_i^T\theta\right)}{1+\exp(x_i^T\theta)} \tag{1}$$

Where, $\theta = (\theta_0, \theta_1, \theta_2, \ldots, \theta_n)$ represents coefficient matrix of $(n+1)\times 1$. y_i is a return variable with a value of 1 or 0. $y_i = 1$represents the disease class and $y_i = 0$ represents the non-disease class. $\pi_i \in (0,1)$ represents the return probability of the classifier label y_i. If $\pi_i \geq 0.5$, the sample is divided into disease categories. If $\pi_i \geq 0.5$, the sample is divided into healthy categories. The log-likelihood can be expressed as:

$$\log\prod_{i=1}^{n} p(y_i|x_i) = \sum_{i=1}^{n}(y_i\log(\pi_i) + (1-y_i)\log(1-\pi_i)) \tag{2}$$

The loss function based on Eq. (2) is defined as:

$$J\left(\theta\right) = -\frac{1}{m}\sum_{i=1}^{m}\left(y_i\log\left(\pi_i\right) + \left(1-y_i\right)\log\left(1-\pi_i\right)\right) \tag{3}$$

The estimation of the vector θ is obtained by minimizing Eq.(3):

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}}\left[-\frac{1}{m}\sum_{i=1}^{m}\left(y_i\log\left(\pi_i\right) + \left(1-y_i\right)\log\left(1-\pi_i\right)\right)\right] \tag{4}$$

2.3. Adaptive LASSO logistic regression model

When the dimension of the experimental data is higher than the number of training samples, logistic regression is prone to be overfitting. To obtain a robust classifier, the penalization techniques for logistic regression are proposed as:

$$J\left(\theta\right) = -\frac{1}{m}\sum_{i=1}^{m}\left(y_i\log\left(\pi_i\right) + \left(1-y_i\right)\log\left(1-\pi_i\right)\right) + \lambda\sum_{j=1}^{n}P(\theta_j) \tag{5}$$

Where, $\lambda > 0$ is a tuning parameter. Larger value of λ means more zero rows in the weight matrix. A few important features are retained. $P(\theta_j)$ is the penalty term. Zou et al. [25] proposed the adaptive LASSO. It has oracle properties. The sparse logistic regression with adaptive LASSO is

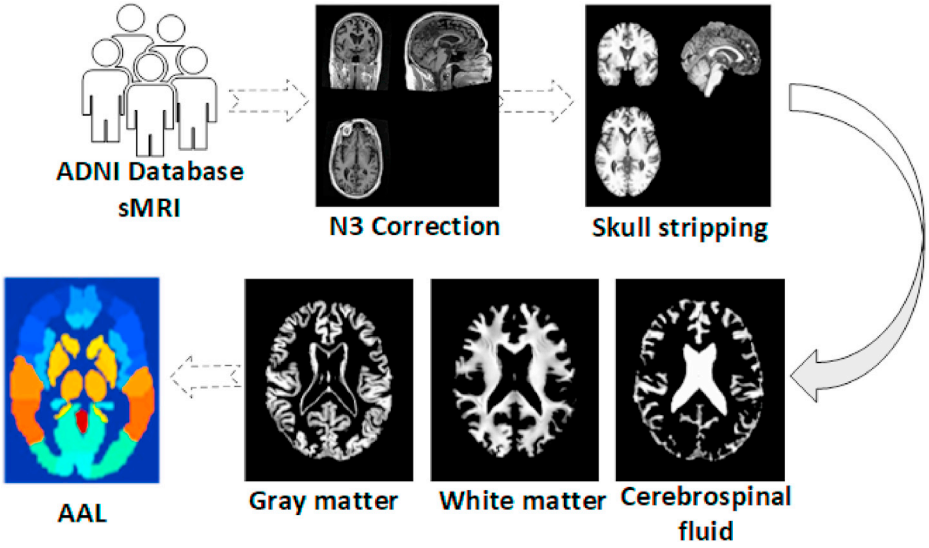

Fig. 1. Image preprocessing and feature extraction process.

defined as:

$$J\left(\theta\right) = -\frac{1}{m}\sum_{i=1}^{m}\left(y_i\log\left(\pi_i\right) + \left(1-y_i\right)\log\left(1-\pi_i\right)\right) + \lambda\sum_{\mathrm{j}=1}^{n} w_j\left|\theta_j\right| \quad (6)$$

Where, $w = (w_1, w_2, \ldots, w_n)^T$ is a $n \times 1$weight vector. The adaptive weight gives different weights to different coefficients in L_1 penalty. It depends on root n-consistent initial values of $\theta^*$and $w_j = \left(\left|\theta_j^*\right|\right)^{-\gamma}$where $\gamma > 0$. The initial weight of adaptive LASSO is obtained by maximum likelihood estimation. The estimation of the vector θ is shown as:

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}}\left[-\frac{1}{m}\sum_{i=1}^{m}\left(y_i\log\left(\pi_i\right) + \left(1-y_i\right)\log\left(1-\pi_i\right)\right) + \lambda\sum_{j=1}^{n} w_j\left|\theta_j\right|\right] \quad (7)$$

2.4. The gradient descent algorithm for the adaptive LASSO logistic regression

Since logistic regression has no analytical solution, it can only be solved by iterative methods. In our article, the gradient descent method is used to solve the sparse logistic regression. Algorithm 1 shows the solution process.

Algorithm 1. the gradient descent for the adaptive LASSO logistic regression

Step 1: The initial points $\theta_0 \in R$ and parameter $\varepsilon \geq 0$, let $k = 0$.

Step 2: If $J(\theta_k) - J(\theta_k + \partial d) \leq \varepsilon$ or satisfy the maximum iteration, algorithm to stop otherwise, go to the next step.

Step 3: Let $d = -g$, $g = \sum_{j=1}^{n}(sigmoid(X\theta) - y)^{'} X^j + \lambda\omega_j sign(\theta_j)$.

Step 4: Calculate the iteration step size $\partial > 0$.

Step 5: Let $\theta_{k+1} = \theta_k + \partial\ d$, $k = k+1$ return step2.

2.5. Particle swarm optimization feature selection stage

Particle swarm optimization (PSO) is global search algorithm based on swarm intelligence. In PSO feature selection stage, each particle represents a subset of features. Particle motion is determined by position vector and velocity vector. In n-dimensional space, the position of the i-th particle is $X_i = (x_i^1, x_i^2, \ldots, x_i^n)$. The speed is $V_i = (v_i^1, v_i^2, \ldots, v_i^n)$. Where, n represents the number of features. The current best position of a particle is $PB_i = [pb_i^1, pb_i^2, \ldots, pb_i^n]$, and the best global position of the neighbor is $GB_i = [gb_i^1, gb_i^2, \ldots, gb_i^n]$. The particle's velocity is updated as follows:

$$v_i^j = wv_i^j + k_1 r_1\left(\mathrm{pb}_i^j - x_i^j\right) + k_2 r_2\left(\mathrm{gb}_i^j - x_i^j\right) \quad (8)$$

Where w is the inertia weight ranging between 0 and 1, $k_1$and k_2 are learning factors, r_1 and r_2 are random numbers ranging between 0 and 1.

In feature selection, a particle represents a potential solution (i.e., feature subset) in an n-dimensional space. The particles are represented using a binary bit string with length n, where n is the total number of features. The position of the particle updates on the basis of the following formulae:

$$x_i^j = \begin{cases} 1, & if \quad sigmoid\left(v_i^j\right) > rand() \\ 0, otherwise \end{cases} \quad (9)$$

Where, $sigmoid(v_i^j)$ is $1/1 + e^{-v_i^j}$. Setting that the position threshold $rand()$ is a random number and uniformly distributed in [0,1].

2.6. The proposed method

Among many features that lead to AD, only a small number of features play a major role. Therefore, selecting the most discriminative features in the classification of AD can achieve the better prediction results. In this paper, PSO along with adaptive LASSO logistic regression is used to improve the classification prediction of AD. The proposed method can select the optimal feature subset for better classification. The objective function is defined by:

$$Fitness = Accuracy\ \ of\ \ the\ \ LR\ \ with\ \ adaptive\ \ LASSO\ \ model \tag{10}$$

In this paper, the fitness of the objective function is specified as the classification accuracy of the logistic regression with the adaptive LASSO. A pseudo code for the proposed method is described in Algorithm 2.

Algorithm 2. Pseudo code of our proposed method for AD classification

```
1:Begin
2:    Initialize the position x and velocity v of each particle, maxiter.
3:    for i=1 to m do
4:      Calculate the pbest of particle i.
5:    end for
6:    Update the gbest.
7:    While(t< maxiter)
8:        for i=1 to m do
9:          for j=1to n do
10:           Update the velocity v according to Eq (8).
11:           Update the position x according to Eq (9).
12:         end for
13:         Compute the fitness value for each particle according to Eq (10).
14;       end for
15:       end while
16:   Return the best selected features and classification accuracies on the test set.
17:end
```

3. Results and discussion

3.1. Experiment setting

In this article, we implement three classification tasks: AD subjects versus HC subjects (AD vs. HC), MCI subjects versus HC subjects (MCI vs. HC) and cMCI versus sMCI (cMCI vs. sMCI). To achieve a fair comparison, two procedures were set up. First, we take 70% of data through the training process and 30% through the testing of each clinical group (i.e., the AD/HC, MCI/HC and cMCI/sMCI groups).

Second, we perform ten-fold cross-validation in the training sets to select the optimal parameter λ. To avoid bias due to the random distribution of samples, we repeated the experiment 50 times. We report the average performances in terms of accuracy (ACC), sensitivity (SEN), specificity (SPE), receiver operating characteristic curve (ROC) and area under the receiver operating characteristic (AUC) on the test set. The specific formula is defined as:

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \tag{11}$$

$$SEN = \frac{TP}{TP + FN} \tag{12}$$

$$SPE = \frac{TN}{TN + FP} \tag{13}$$

ROC curve is a powerful tool for the study of the performance of the classifier. On the ROC curve, the horizontal axis is false positive rate (FPR) and the vertical axis is true positive rate (TPR). The formula is given by:

$$TPR = \frac{TP}{TP + FN} \tag{14}$$

$$FPR = \frac{FP}{TN + FP} \tag{15}$$

The true positive (TP) stands for the number of patients who are

Table 2
Different methods in AD/MCI classification comparison.

Subjects	Method	ACC (%)	SEN (%)	SPE (%)	Selected features
AD vs. HC	PSO-LR	80.82	78.57	83.71	60
	ALLR	91.50	80.50	92.20	35
	PSO-ALLR	**96.27**	**93.33**	**95.78**	**17**
MCI vs. HC	PSO-LR	77.47	78.57	72.16	42
	ALLR	81.93	88.71	77.00	29
	PSO-ALLR	**84.81**	**90.00**	**85.71**	**23**
cMCI vs. sMCI	PSO-LR	66.10	66.67	64.70	41
	ALLR	73.00	69.61	75.25	30
	PSO-ALLR	**76.13**	**78.00**	**86.50**	**16**

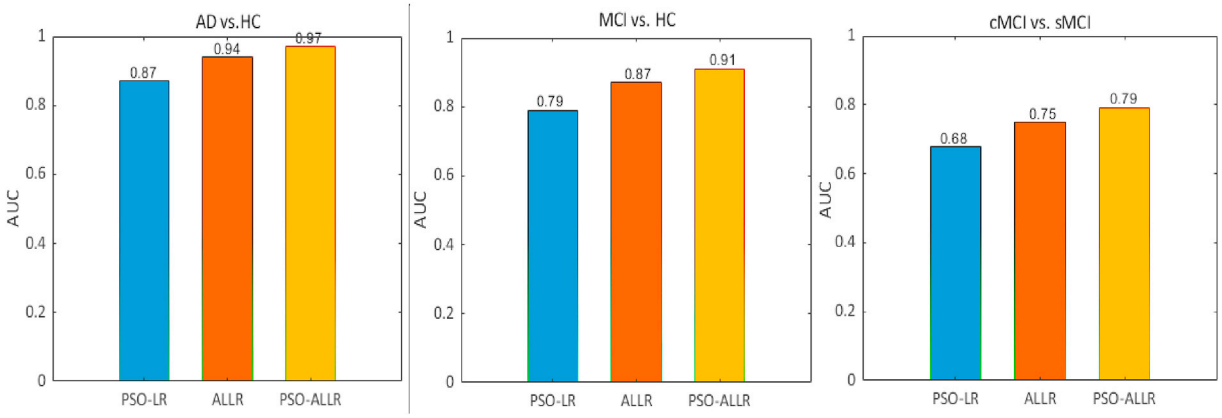

Fig. 2. AUC of AD/MCI classification by different methods.

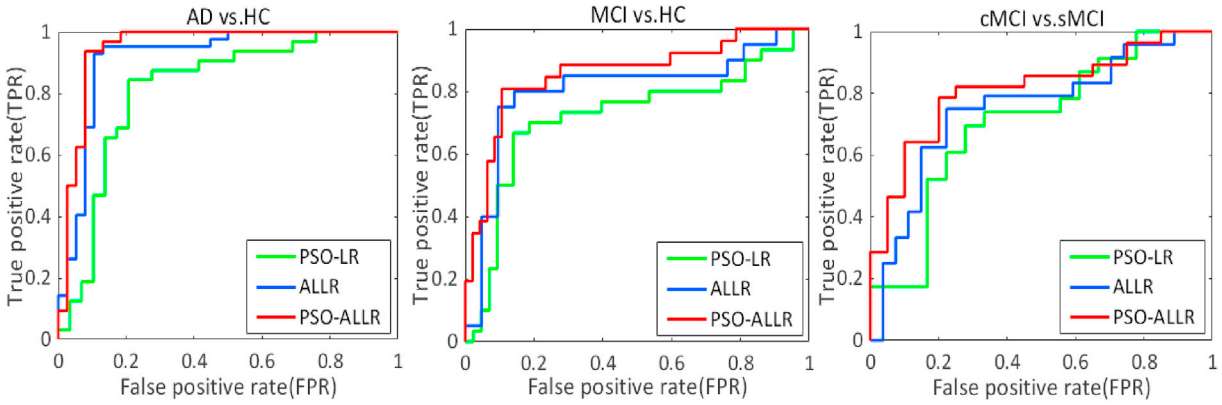

Fig. 3. ROC curves of AD/MCI classification by different method.

correctly classified into disease categories. The true negative (TN) is the number of healthy people with the correct classification of the health class. False positive (FP) is the number of healthy people who are divided into sick patients. False negatives (FN) is the number of sick patients classified as healthy people.

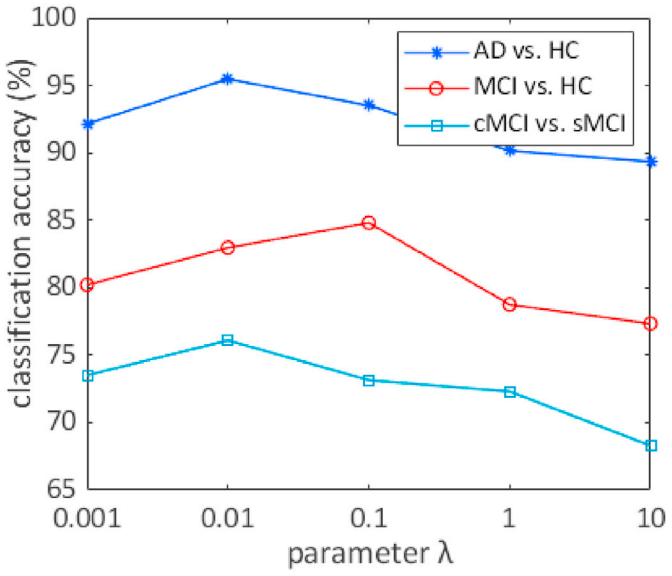

Fig. 4. Classification accuracy under different parameter λ

3.2. Experiment results

This part presents a summary and the discussion of the results. The proposed method (PSO-ALLR) is applied to three groups of classification tasks and is compared with PSO-LR (Logistic regression based on PSO) and ALLR (adaptive LASSO logistic regression).

Table 2 shows the corresponding experimental results. It can be seen from Table 2 that, our method selected 17 features in AD vs. HC, 23 features in MCI vs. HC, and 16 features in cMCI vs. sMCI. The number of features selected by our proposed method is the smallest compared with another two comparative approaches. PSO can remove redundant and irrelevant features. The adaptive LASSO can further determine which features are important among the remaining features. What's more, in terms of classification performance, the classification accuracy achieved by our method is 96.27% in AD vs. HC, which is higher than that of another two logistic regression models. In MCI vs. HC, the classification accuracy is 84.81%. We also carried out experiments on classifying cMCI from sMCI. As a prodromal stage of AD, MCI has a high conversation risk, so it is necessary to identify cMCI from sMCI. Early diagnosis and intervention can delay the conversion of MCI to AD. As can be seen from Table 2, our method obtained the classification accuracy of 76.13%, which is 5% higher than maximum accuracy achieved by the comparison method. This indicates that PSO removes redundant features and retains those features representing a high individual correlation with AD. The combination of PSO and adaptive LASSO logistic regression can be better used for the diagnosis of AD.

For the metrics of sensitivity and specificity, the higher the sensitivity, the lower the chance of mis-diagnosing AD/MCI patients; also the higher the specificity, the lower the chance of mis-diagnosing HC to AD/MCI. It can be seen from Table 2 that, our proposed method showed

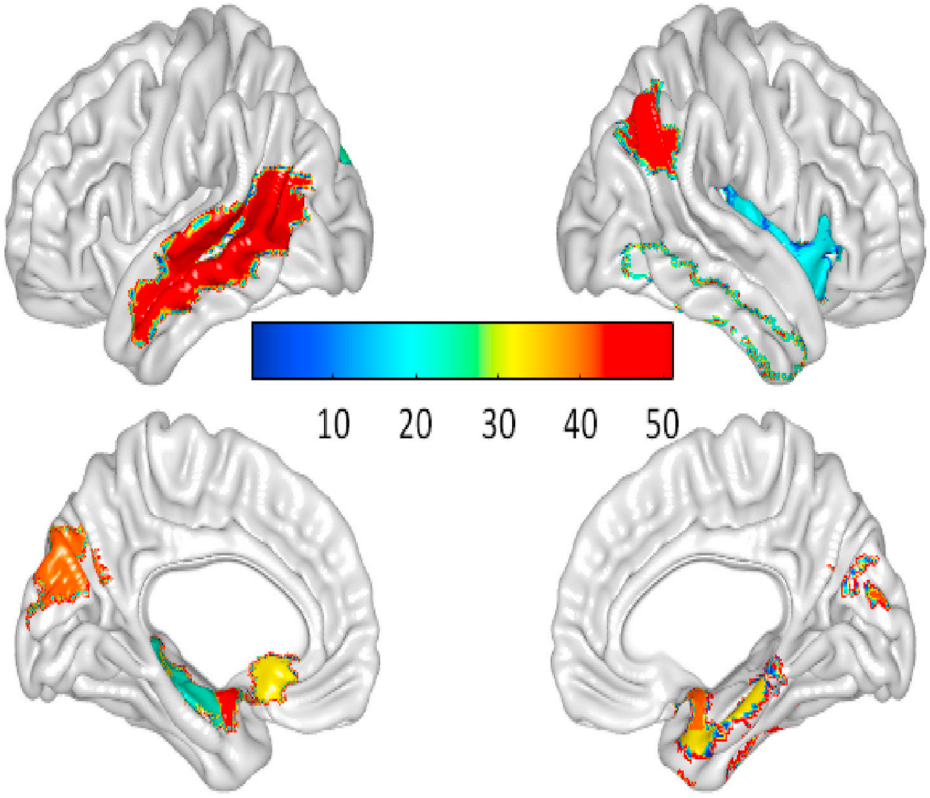

Fig. 5. The most discriminative brain regions in MCI vs. HC.

Table 3
The top 10 discriminative brain regions identified by our proposed method in MCI vs. HC

Brain regions	Selected times
Angular_R	50
Amygdala_L	50
Cuneus_L	48
Temporal_Mid_L	48
Amygdala_R	45
Olfactory_L	42
ParaHippocampal_R	42
Tmporal_Inf_R	38
Hippocampus_L	33
Precentral_L	31

higher sensitivity and specificity in three classification tasks. Our method is less likely to mis-diagnose subjects with AD/MCI, compared with comparison method.

Regarding AUC and ROC curve, as can be seen from Fig. 2, our proposed showed the best AUC of 0.96 in AD vs. HC, 0.87 in MCI vs. HC, and 0.76 in cMCI vs. sMCI. Compared with ALLR, the proposed method increased AUC by 0.03 (AD vs. HC), 0.04 (MCI vs. HC), and 0.04 (cMCI vs. sMCI). Fig. 3 shows the corresponding ROC curve in three classification tasks. Through ROC curve, we can intuitively see that the proposed method has excellent diagnostic power.

3.3. Evaluation with the parameter

In the adaptive LASSO logistic regression, the parameter λ is a regularization control parameter, which controls the sparsity of the model. By selecting optimal parameter λ, the classifier can select the most discriminating brain areas and has better performance. Fig. 4 gives the classification accuracies of our proposed method under different values of the parameterλin the three classification tasks. The horizontal axis represents different values of λ. The vertical axis represents classification accuracies obtained under different parameter values. In this paper, the parameter λ is set from $\{10^{-3},...,10\}$. As we can see from Fig. 4, a proper selection of parameter λwill improve the performance of the proposed. In AD vs. HC, when $\lambda \in \{0.001, 0.01\}$, our method achieves good classification performance. In MCI vs. HC, when $\lambda \in \{0.01, 0.1\}$, our method obtains the high classification accuracy. In cMCI vs. sMCI, our method can better identify cMCI from sMCI when $\lambda \in \{0.001, 0.1\}$. Therefore, we look for the optimal parameter values in the above three ranges. Our method achieves the best classification performance when $\lambda = 0.005$(AD

Table 4
Classification accuracy of different classification methods on AD, MCI, HC.

Method	Subjects (AD/HC)	AD vs. HC(%) ACC | SEN | SPE			Subjects (MCI/HC)	MCI vs. HC(%) ACC | SEN | SPE			Subjects (cMCI/sMCI)	cMCI vs. sMCI(%) ACC | SEN | SPE		
Min et al., 2014	97/128	91.64	88.56	93.85	–	–	–	–	117/117	72.41	72.12	72.58
Li et al., 2015	51/52	91.40	–	–	99/52	77.40	–	–	43/56	57.40	–	–
Yu et al., 2016	50/52	90.60	–	–	97/52	75.10	–	–	–	–	–	–
Ben et al., 2017	45/52	88.16	77.58	93.26	58/52	76.92	66.71	81.15	–	–	–	
Zeng et al., 2018	92/92	81.25	–		–	–	–		95/82	69.23	–	–
Xiao et al., 2020	51/50	93.33	92.25	94.75	96/50	82.75	86.67	78.57	51/45	72.87	81.25	65.45
PSO-ALLR	**51/50**	**96.27**	**93.33**	**95.78**	**96/50**	**84.81**	**90.00**	**85.71**	**51/45**	**76.13**	**78.00**	**86.50**

vs. HC), $\lambda = 0.1$ (MCI vs. HC) and $\lambda = 0.02$ (cMCI vs. sMCI).

3.4. The most discriminative brain regions

Apart from introducing the classification performance of our method, we also report the AD-related brain regions selected by our proposed method in MCI vs. HC. Fig. 5 plots the some frequently selected brain regions in MCI vs. HC. Table 3 summarizes the discriminative brain regions and the corresponding number of selections. They are known to be related to AD [26–32]. For example, hippocampus is related to human memory and learning. It is the first brain region to be damaged in relation to AD disease. Amygdala controls human emotions and cognition. ParaHippocampal has an important relationship with cognition and emotion. This shows that our proposed method can help to find brain regions related to AD so as to better assist the diagnosis of AD.

3.5. Comparison with other methods

To further reflect the advantages of our proposed method, we list some representative methods in recent years. Table 4 represents the classification results obtained by other methods, including SVM [33], DBN [34], Multi-task learning [35], Multiple Kernel Learning [36], SDPSO-SVM-PCA [37], Sparse logistic regression [18]. Although the size of the data set and the method of feature extraction may be different, the data comes from ADNI database. So, it is worth comparing the classification performance. Since several studies use multimodal biomarker, we report their results using only MRI data if available; Otherwise, we report their results using multimodal data. In Table 4, our proposed method achieves the highest accuracy, sensitivity and specificity in three classification tasks. Compared with Min et al.'s results, our method improves classification accuracy by 4.63% in AD vs. HC. In particular, in the cMCI vs. sMCI, our method improves by 3.72%. What's more, compared with our previous method, the proposed method increased ACC by 2.94% (AD vs. HC), 2.06% (MCI vs. HC), and 3.26% (cMCI vs. sMCI). This further proves the advantages of our proposed method in AD classification.

4. Conclusions

In our article, we proposed a novel method for the diagnosis of AD using a PSO algorithm combined with an adaptive LASSO logistic regression. Our method consists of two stages, which combines PSO and adaptive LASSO logistic regression. In the first stage, PSO can remove unimportant features. In the second stage, adaptive LASSO selects the most discriminative brain regions from the remaining features and achieves better classification performance. We evaluate our proposed method based on ADNI dataset. It is worth noting that the accuracy of our proposed method for AD vs. HC and MCI vs. HC are 96.27%, 84.81% respectively. In particular, we use the method for the classification of cMCI and sMCI to obtain the classification accuracy of 76.13%, which is valuable for the timely diagnosis and treatment of MCI. The experimental results show that compared with several other classification methods, our method has good competitiveness in AD classification. In the future work, we will further consider the optimization of the logistic regression model to improve the classification performance of the model, and better apply to AD diagnosis problems.

Funding

This work is partially supported by National Natural Science Foundation of China (71971190); Postgraduate Education Quality Curriculum Construction Project of Shandong Province (SDYKC19178).

Author statement

Xinchun Cui, **Xiaoli Liu:** Original idea, Methodology. **Ruyi Xiao:** Writing - Original Draft. **Hong Qiao**, **Xiangwei Zheng:** Investigation. **Jianzong Du:** Supervision. **Yiquan Zhang:** Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C.Y. Wee, P.T. Yap, D. Zhang, K. Denny, J.N. Browndyke, G.G. Potter, K.A. Welsh-Bohmer, L. Wang, D.G. Shen, Identification of MCI individuals using structural and functional connectivity networks, Neuroimage 59 (3) (2012) 2045–2056.
[2] I. Beheshti, H. Demirel, H. Matsuda, AsDNI, Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm, Comput. Biol. Med. 83 (2017) 109–119.
[3] Alzheimer's Association, Alzheimer's disease facts and figures, Alzheimer's Dementia 11 (3) (2015).
[4] J. Zhang, Y. Gao, Y. Gao, B.C. Munsell, D.G. Shen, Detecting anatomical landmarks for fast Alzheimer's disease diagnosis, IEEE Trans. Med. Imag. 35 (12) (2016) 2524–2533.
[5] M. Liu, D. Zhang, E. Adeli, D.G. Shen, Inherent structure-based multiview learning with multitemplate feature representation for alzheimer's disease diagnosis, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 63 (7) (2015) 1473–1482.
[6] F Ahmad, H Zulifqar, T Malik, Classification of Alzheimer disease among susceptible brain regions, Int. J. Imag. Syst. Technol. 29 (3) (2019) 222–233.
[7] Z.Y. Algamal, M.H. Lee, Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification, Expert Syst. Appl. 42 (23) (2015) 9326–9332.
[8] Y. Liang, C. Liu, X.-Z. Luan, K.-S. Leung, T.-M. Chan, Z.-B. Xu, H. Zhang, Sparse logistic regression with a L 1/2 penalty for gene selection in cancer classification, BMC Bioinf. 14 (1) (2013) 198.
[9] R. Tibshirani, Regression shrinkage and selection via the lasso: a retrospective 73 (3) (2011) 273–282.
[10] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties 96 (456) (2001) 1348–1360.
[11] H. Zou, T. Hastie, Regularization and variable selection via the elastic net 67 (2) (2005) 301–320.
[12] H. Zeng, A. Song, Optimizing single-trial EEG classification by stationary matrix logistic regression in brain–computer interface, IEEE Trans. Neural Network Learn. Syst. 27 (11) (2015) 2301–2313.
[13] Z. Qiu, D.J. Miller, G. Kesidis, A maximum entropy framework for semisupervised and active learning with unknown and label-scarce classes, IEEE Trans. Neural Network Learn. Syst. 28 (4) (2016) 917–933.
[14] K. Koh, S.-J. Kim, S. Boyd, An interior-point method for large-scale l1-regularized logistic regression, J. Mach. Learn. Res. 8 (Jul) (2007) 1519–1555.
[15] Z.Y. Algamal, M.H. Lee, Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification, Comput. Biol. Med. 67 (2015) 136–145.
[16] Y. Wang, W. Liu, L. Caccetta, G. Zhou, Parameter selection for nonnegative l1 matrix/tensor sparse decomposition, Oper. Res. Lett. 43 (4) (2015) 423–426.
[17] Y. Wang, G. Zhou, L. Caccetta, W. Liu, An alternative Lagrange-dual based algorithm for sparse signal reconstruction, IEEE Trans. Signal Process. 59 (4) (2010) 1895–1901.
[18] A. Unler, A. Murat, A discrete particle swarm optimization method for feature selection in binary classification problems, Eur. J. Oper. Res. 206 (3) (2010) 528–539.
[19] R. Xiao, X. Cui, H. Qiao, et al., Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net, Biomed. Signal Process. Contr. 66 (2021).
[20] Initiative AsDN T. Ye, C. Zu, B. Jie, D. Shen, D. Zhang, Discriminative multi-task feature selection for multi-modality classification of Alzheimer's disease, Brain Imag. Behav. 10 (3) (2016) 739–749.
[21] Q. Wang, Y. Zheng, G. Yang, W. Jin, X. Chen, Y. Yin, Multiscale rotation-invariant convolutional neural networks for lung texture classification, IEEE J. Biomed. Health Inform. 22 (1) (2017) 184–195.
[22] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects, Neuroimage 104 (2015) 398–412.
[23] G.A. Papakostas, A. Savio, M. Graña, V.G. Kaburlasos, A lattice computing approach to Alzheimer's disease computer assisted diagnosis based on MRI data, Neurocomputing 150 (2015) 37–42.
[24] R. Wang, N. Xiu, C. Zhang, Greedy projected gradient-Newton method for sparse logistic regression, IEEE Trans. Neural Network Learn. Syst. PP(99) (2019) 1–12.
[25] H. Zou, JJotAsa, The adaptive lasso and its oracle properties 101 (476) (2006) 1418–1429.
[26] B. Cheng, M. Liu, D. Shen, Z. Li, D. Zhang, Multi-domain transfer learning for early diagnosis of Alzheimer's disease, Neuroinformatics 15 (2) (2017) 115–132.
[27] W. Dai, O.L. Lopez, O.T. Carmichael, J.T. Becker, L.H. Kuller, H.M. Gach, Mild cognitive impairment and alzheimer disease: patterns of altered cerebral blood flow at MR imaging, Radiology 250 (3) (2009) 856–866.

[28] H. Matsuda, Voxel-based morphometry of brain MRI in normal aging and Alzheimer's disease, Aging Dis. 4 (1) (2013) 29.
[29] C.Y. Wee, P.T. Yap, D. Shen, Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns, Hum. Brain Mapp. 34 (12) (2013) 3411–3425.
[30] G. Karas, P. Scheltens, S. Rombouts, R. Van Schijndel, M. Klein, B. Jones, W. Van Der Flier, H. Vrenken, F.J.N. Barkhof, Precuneus atrophy in early-onset Alzheimer's disease: a morphometric structural MRI study 49 (12) (2007) 967–976.
[31] S.P. Poulin, R. Dautoff, J.C. Morris, L.F. Barrett, B.C. Dickerson, AsDNI, Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity, Psychiatr. Res. Neuroimaging 194 (1) (2011) 7–13.
[32] K. Zhao, Y.H. Ding, Y. Han, Y. Fan, A.F. Alexander-Bloch, T. Han, D. Jin, B. Liu, J. Lu, C.Y. Song, et al., Independent and reproducible hippocampal radiomic biomarkers for multisite Alzheimer?s disease: diagnosis, longitudinal progress and biological basis, Sci. Bull. 65 (13) (2020) 1103–1113.
[33] R. Min, G. Wu, J. Cheng, Q. Wang, D. Shen, Multi-atlas based representations for Alzheimer's disease diagnosis, Hum. Brain Mapp. 35 (10) (2014) 5052–5070.
[34] F. Li, L. Tran, K.-H. Thung, S. Ji, D. Shen, J. Li, A robust deep model for improved classification of AD/MCI patients, IEEE J. Biomed. Health Inform. 19 (5) (2015) 1610–1616.
[35] G. Yu, Y. Liu, D. Shen, Graph-guided joint prediction of class label and clinical scores for the Alzheimer's disease, Brain Struct. Funct. 221 (7) (2016) 3787–3801.
[36] O.B. Ahmed, J. Benois-Pineau, M. Allard, G. Catheline, C.B. Amar, Recognition of alzheimer's disease and mild cognitive impairment with multimodal image-derived biomarkers and Multiple Kernel learning, Neurocomputing 220 (2017) 98–110.
[37] N. Zeng, H. Qiu, Z. Wang, W. Liu, H. Zhang, Y. Li, A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease, Neurocomputing 320 (2018) 195–202.