Biomedical Signal Processing and Control 66 (2021) 102362

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net

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

[a] School of Computer Science, Qufu Normal University, 276800, Rizhao, China
[b] Business School, Shandong Normal University, 250014, Jinan, China
[c] School of Information Science and Engineering, Shandong Normal University, 250014, Jinan, China
[d] Department of Neurology, Zhejiang Hospital (Lingyin District), 310013, Hangzhou, China

ARTICLE INFO

Keywords:
Alzheimer's disease
Mild cognitive impairment
MRI image
Sparse logistic regression

ABSTRACT

Accurate prediction of high-risk group who may convert to Alzheimer's disease (AD) patients is critical for the future treatment of patients. Recently, logistic regression is used for the early diagnosis of AD. However, due to the high-dimensional small sample characteristics of AD data, this brings difficulties to logistic regression-aided diagnosis. To solve the problem, in this paper, we propose sparse logistic regression with the generalized elastic net for the early diagnosis of AD. The generalized elastic net is composed of Lp regularization and L_2 regularization. The Lp regularization can produce sparse solutions. L_2 regularization ensures that the correlated brain regions are in solution. We evaluate our proposed method on 197 subjects from the baseline MRI data of ADNI database. Our proposed method achieves classification accuracy of 96.10, 84.67, and 75.87 %, for AD vs. HC, MCI vs. HC, and cMCI vs. sMCI, respectively. Experimental results show that, compared with previous methods, our proposed method captures distinct brain regions that are significantly related to AD conversion and provides a significant enhancement in AD classification.

1. Introduction

Alzheimer's disease (AD) is an irreversible degenerative neurological disease. The symptoms of this disease are loss of cognitive and memory, which severely affect people's daily life [1]. According to statistics, there are 50 million AD patients and will be close to 13.8 million people 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. Thus, the early diagnosis of AD and preventing disease progression will help reduce the burden of the disease on society.

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]. So, accurate prediction of the risk of MCI to AD and timely treatment might be effective to delay the conversion of MCI to AD. However, compared to AD classification tasks, MCI conversion prediction is relatively difficult. The changes in the brain structure of MCI patients are relatively stable [4].

Recently, machine learning has been widely used for the early diagnosis of AD [5–9]. Among numerous machine learning algorithms, logistic regression (LR) is a discriminative method that is widely used. LR has a direct probabilistic interpretation. In addition to the class label information it can obtain direct classification probabilities [10]. However, in neuroimaging data analysis, the training samples are limited and the dimensions are higher, which bring great challenges to the diagnosis of AD by logistic regression [11] and easily lead to overfitting. So, selecting the most discriminative features to remove redundant features is essential for the classification of AD and prediction on the conversion of MCI. Feature selection methods can be divided into three categories: filter, wrapper and embedded method. Among them, the embedded method is a feature selection method with a penalty term, which can perform feature selection and classifier design at the same time. Many penalty terms are applied to the logistic regression model to prevent the overfitting and improve the generalization performance of the model [12,13]. Koh et al. [14] introduced the L_1 regularized logistic regression for high-dimensional data classification. The L_1 regularization can shrink the regression coefficients to zero, thereby selecting some important features simultaneously [15–18]. However, when the

* Corresponding author.
E-mail address: cuixc@nuaa.edu.cn (X. Cui).

https://doi.org/10.1016/j.bspc.2020.102362
Received 11 July 2020; Received in revised form 26 October 2020; Accepted 16 November 2020
Available online 19 February 2021

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

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

correlation between features is high, the L_1 regularization only selects the one feature or a few of them among a group of correlated features. To solve the above problems, Zou et al. [19] proposed elastic net for variable selection which can maintain sparsity and grouping effects. Ryali et al. [20] proposed sparse logistic regression with elastic net regression which combine the L_1 regularization and L_2 regularization for whole brain classification. Schouten et al. [21] use elastic net logistic regression for AD classification. However, comparing to L_1 regularization, the Lp regularization can produce more sparse solutions when choosing the appropriate p value. Li et al. [22] proposed the generalized elastic net regularization and applied it to linear discriminant analysis. Motivated by this, in this paper, we proposed the sparse logistic regression with the generalized elastic net regularization for the early diagnosis of AD. The generalized elastic net regularization which combines Lp regularization and L_2 regularization can produce the sparse solutions and ensure that the correlated brain regions are included in the resulting solutions simultaneously. It can select the most discriminative features for AD/MCI classification. The main contributions of this paper include: (1) Collect experimental data from ADNI database. Through the automatic anatomical labeling (AAL) template, the brain is divided into 90 regions of interest (ROI). We extracted the gray matter volume of 90 regions as features. (2) Use the generalized elastic net regularization method to select the most distinguishing features for classification. (3) The proposed method is used to predict the conversion from MCI to AD, which is essential for the early diagnosis of AD.

It's worth noting that the objective function of logistic regression has no analytical solution, the optimal parameters cannot be directly obtained. In addition, the Lp regularization penalty term is non-convex, which makes it difficult to solve the sparse logistic regression. Coordinate descent is a s "one-at-a-time" optimization algorithm. It is widely used to solve non-convex penalty term [23,24]. In this paper, we use the coordinate descent algorithm to solve the model.

The remainder of this article is as follows. The second part introduces the materials and methods used in our article. The third part shows our experimental results. The fourth part is to discuss the experimental results. The last part is conclusion.

2. Method

2.1. ADNI database

The experimental datasets are derived from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI).ADNI was founded in 2003 by the National Institute of Biomedical Imaging and Bioengineering. It is a non-profit organization [25]. ADNI provides unlimited data access and encourages researchers to develop potential methods for analyzing the progression of early AD. MRI is a widely used imaging mode in the diagnosis of AD [26–28]. It is able to make better comparison between different soft tissues. Therefore, we use structural MR images for analysis. We select MRI images of 197 subjects in the ADNI database, including 51 AD, 50 HC and 96 MCI (including 51 converted MCI (cMCI) and 45 stable MCI (sMCI)). Table 1 presents detailed information about these subjects.

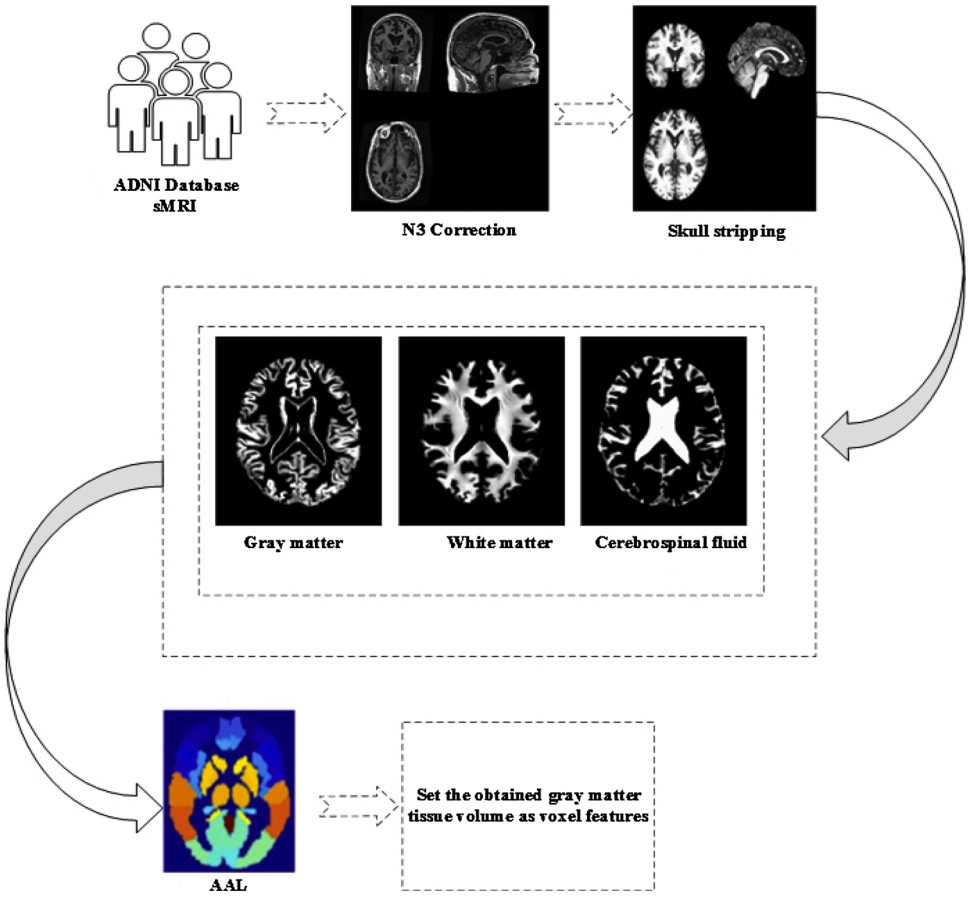

Fig. 1. Image pre-processing and feature extraction.

2.2. Image pre-processing and feature selection

The experimental data downloaded from the ADNI database require a series of image preprocessing and feature extraction. Fig. 1 shows the image pre-processing and feature extraction of MRI image. All MRI images were pre-processed by first performing an anterior commissure-posterior commissure (AC-PC) correction using the MIPAV software. The AC-PC corrected images were resampled to $256 \times 256 \times 256$, and the N3 algorithm was used to correct intensity inhomogeneity. An accurate and robust skull stripping was performed using VBM8 toolbox. Then, we used VBM8 toolbox to segment brain into three different tissue types: gray matter (GM), white matter (WM)and cerebrospinal fluid (CSF). Next, by warping the Automated Anatomical Labeling (AAL) template, for each subject, we parcellated the brain space into 90 regions of interest (ROIs). Finally, we compute the GM tissue volume from the subject's MRI image as a feature.

2.3. Logistic regression model

As a special nonlinear model, logistic regression is mainly used to solve classification problems [29]. In this paper, we implement LR for the binary classification problem. Suppose we have m samples, where $x_i = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in})^T$ is i-th input pattern with dimensionality$n$. Then the expression of the logistic regression model is:

$$\pi_i = p(y_i|x_i) = sigmoid(x_i^T\theta + \theta_0) = \frac{\exp(x_i^T\theta + \theta_0)}{1 + \exp(x_i^T\theta + \theta_0)}. \tag{1}$$

Where, $\theta \in R^n$ is the whole parameter vector. $y_i$is a return variable, containing two values either 1 or 0. It is obtained by $I(\pi_i > 0.5)$, where $I(.)$ is an indicator function.$\pi_i \in (0,1)$represents the return probability of the classifier label y_i. The log-likelihood can be expressed as:

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

The loss function based on (2) is shown as:

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

2.4. Sparse logistic regression with the generalized elastic net regularization

When faced with high-dimensional small sample data, logistic regression is prone to overfitting. To reduce the number of features and obtain a robust classifier, the penalization techniques for logistic regression were proposed as:

$$J(\theta) = -\sum_{i=1}^{m} (y_i \log(\pi_i) + (1 - y_i)\log(1 - \pi_i)) + \lambda \sum_{j=1}^{n} P(\theta_j). \tag{4}$$

Where,$\lambda > 0$is a tuning parameter and $P(\theta_j)$ is the regular term that constraints the feature coefficient estimates. In many regular terms, the generalized elastic net can select the important features by setting some feature coefficients to 0 and maintaining the grouping effects. The generalized elastic net is a combination of Lp regularization and L_2 regularization. The Lp regularization can produce more sparse solutions, which make some feature coefficients to 0. When the pairwise correlations between two brain regions are very high, L_2 regularization can ensure that two brain regions are retained or deleted at the same time. The loss function of a sparse logistic regression with the generalized elastic net regularization is defined as:

$$\begin{aligned} J(\theta) &= -\sum_{i=1}^{m} (y_i \log(\pi_i) + (1 - y_i) log(1 - \pi_i)) \\ &+ \lambda_1 \sum_{j=1}^{n} |\theta_j|^p + \lambda_2 \sum_{j=1}^{n} \theta_j^2 \end{aligned} \tag{5}$$

Where, the parameters$\lambda_1$and λ_2 control the model's sparsity and group effect respectively [30]. In this paper, $\lambda_1$and λ_2 are adjusted by the ten-fold cross-validation (CV) method in the training set.

2.5. The optimization algorithm for sparse logistic regression

In this paper, we use the coordinate descent algorithm to solve the sparse logistic regression based on generalized elastic net. The coordinate descent algorithm is an efficient method for solving regularization models. Since, the Eq. (5) does not have closed-form solution, we transformed it into a quadratic function. Differentiating Eq. (5) with respect to $\theta_j$yields following formulae:

$$\frac{\partial J(\theta)}{\partial \theta} = \sum_{i=1}^{m} (\pi_i - y_i)x_i = X(\pi - y) \tag{6}$$

$$\frac{\partial J^2(\theta)}{\partial \theta \partial \theta^T} = \sum_{i=1}^{m} x_i^T k_i x_i = X^T K X \tag{7}$$

Where, $k_i = \pi_i \cdot (1 - \pi_i)$and $K$is the diagonal matrix of diagonal elements k_i. In Eq. (5), θsatisfies the following conditions.

$$\frac{\partial J(\theta)}{\partial \theta} = X(\pi - y) = 0 \tag{8}$$

In order to obtain the solution of Eq. (8), we use Newton-Raphson to iteratively calculate.

$$\begin{aligned} \theta^{new} &= \theta^{old} - \left(\frac{\partial J(\theta)}{\partial \theta \partial \theta^T}\right)^{-1} \frac{\partial J(\theta)}{\partial \theta} \\ &= \theta^{old} + \left(X^T K X\right)^{-1} X^T (y - \pi) \\ &= \left(X^T K X\right)^{-1} X^T K z \end{aligned} \tag{9}$$

Algorithm I: The coordinate descent algorithm for sparse logistic regression with the generalized elastic net

Step1: Initialize $\theta_0 \leftarrow \dfrac{\sum_{i=1}^{m} k_i (z_i - \sum_{j=1}^{n} \theta_j x_{ij})}{\sum_{i=1}^{m} k_i}$.

Step 2: Update $\theta_j(m)$ and cycle over $j = 1,\ldots,n$.

Step 2.1: Compute $\lambda_{critical}$ for j-th coordinate.

Step 2.2: Update $\theta_j(m) \rightarrow Newton\ A1g\ orithm(Eq(13), Eq(14))$,when $\lambda_1 < \lambda_{critical}$.

Step 3: Let $m = m+1$, $\theta(m+1) \leftarrow \theta(m)$.

Repeat Step 2,3 until $\sum_{j=1}^{n} \left(\left|\theta_j(m+1)\right| - \left|\theta_j(m)\right|\right) < 10^{-8}$.

Table 2
Different p values in AD/MCI classification comparison.

Methods	AD vs. HC			MCI vs. HC			cMCI vs. sMCI		
	ACC (Dim)	SEN	SPE	ACC (Dim)	SEN	SPE	ACC (Dim)	SEN	SPE
LR-L_2	90.25(90)	87.50	92.85	79.41(90)	85.89	74.75	68.75(90)	73.35	62.50
LR-L_1	90.79(40)	88.21	92.85	80.10(52)	86.21	75.87	69.45(45)	75.00	64.01
Lp + L_2($p = 0.2$)	93.95(25)	91.25	94.82	80.98(28)	85.51	75.36	70.75(28)	64.21	75.75
Lp + L_2($p = 0.4$)	**96.10(32)**	94.21	**96.12**	82.72(33)	86.51	79.34	69.73(43)	64.91	75.00
Lp + L_2($p = 0.6$)	90.33(48)	88.21	91.87	78.69(30)	84.83	74.25	72.85(53)	72.25	70.25
Lp + L_2($p = 0.8$)	92.13(50)	91.45	94.75	83.33(46)	81.21	78.15	**75.87(47)**	**73.71**	**77.51**
Lp + L_2($p = 1.0$)	93.00(45)	94.15	95.38	**84.67(47)**	**90.75**	**82.61**	72.12(50)	71.41	75.31

Where, $z = X\theta^{old} + K^{-1}(y - \pi)$. Newton-Raphson iteration is to solve the least square problem.

$$\theta^{new} = \underset{\theta}{\operatorname{argmin}}(z - X\theta)^T K(z - X\theta) \tag{10}$$

The iterative form of the generalized elastic net logistic regression model is as follow.

$$\theta^{new} = \underset{\theta}{\operatorname{argmin}}(z - X\theta)^T K(z - X\theta) + \lambda_1 \sum_{j=1}^{n} |\theta_j|^p + \lambda_2 \sum_{j=1}^{n} \theta_j^2 \tag{11}$$

Eq. (11) can be transformed into the following form.

$$L(\theta_j) = \min \sum_{i=1}^{m} k_i \left(z_i - \theta_0 - \sum_{j \neq t} x_{ij}\theta_j \right)^2 + \lambda_1 \sum_{j=1}^{n} |\theta_j|^p + \lambda_2 \sum_{j=1}^{n} \theta_j^2 \tag{12}$$

The differentiating Eq. (12) with respect to $\theta_t$gives the following expression.

$$\frac{\partial L}{\partial \theta_j} = \left(\sum_{i=1}^{m} x_{ij}^2 k_i + 2\lambda_2 \right) \left(\theta_j - \frac{\sum_{i=1}^{m} x_{ij}^2 k_i \sum_{i=1}^{n} x_{ij} \left(y_i - \sum_{t \neq j} \theta_t x_{it} \right)}{\left(\sum_{i=1}^{m} x_{ij}^2 k_i + 2\lambda_2 \right) \sum_{i=1}^{m} x_{ij}^2} \right) + \lambda_1 \times p \times sign(\theta_j) |\theta_j|^{p-1} \tag{13}$$

$$\frac{\partial^2 L}{\partial \theta_j^2} = \sum_{i=1}^{m} x_{ij}^2 k_i + 2\lambda_2 + \lambda_1 \cdot p \cdot (p-1) \cdot |\theta_j|^{p-2} \tag{14}$$

Let $u_j = \sum_{i=1}^{m} x_{ij}^2 k_i + 2\lambda_2, c_j = \frac{\sum_{i=1}^{m} x_{ij}^2 k_i \sum_{i=1}^{n} x_{ij} \left(y_i - \sum_{j \neq t} \theta_t x_{it} \right)}{\left(\sum_{i=1}^{m} x_{ij}^2 k_i + 2\lambda_2 \right) \sum_{i=1}^{m} x_{ij}^2}$.

Eq. (13) can be expressed as.

$$L'\left(\theta_j\right) = u_j\left(\theta_j - c_j\right) + \lambda_1 \cdot p \cdot sign\left(\theta_j\right) |\theta_j|^{p-1} \tag{15}$$

Derivative of Eq. (13) can be obtained.

$$L''\left(\theta_j\right) = u_j + \lambda_1 p \left(p - 1\right) |\theta_j|^{p-2} \tag{16}$$

We need to find $\lambda_{critical}$ for which Eq. (12) has minima. The following equations are used to be solved

$$\begin{cases} L(\theta_j) - \frac{u}{2} c^2 = 0 \\ L'(\theta_j) = 0 \end{cases} \tag{17}$$

The solution is as follow.

$$\begin{cases} \theta_j = \frac{2 - 2p}{2 - p} \bullet c \\ \lambda_{critical} = \frac{u|c|}{2 - p} \bullet \left(\frac{2 - 2p}{2 - p} \bullet |c| \right)^{1-p} \end{cases} \tag{18}$$

The algorithm is stopped when it exceeds maximum number of iterations or maximum relative difference between $\theta_j(m+1)$ and $\theta_j(m)$ falls below $\varepsilon = 10^{-8}$. The procedure of the coordinate descent algorithm for the SLR is given as algorithm I.

3. Results

3.1. Experiment setting

In this paper, 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 assess classification performance of different methods, the program consists of two nested ten-fold cross-validation. In the outer loop, we divide the data set into 10 parts, one of which is used for testing. In the inner loop, we perform ten-fold cross-validation in the training sets to select the optimal parameters $\lambda_1$and λ_2. The classification performance of all methods is quantified by calculation accuracy (ACC), sensitivity (SEN) and specificity (SPE). The specific formula is defined as:

$$ACC = \frac{TP + TN}{TP + FN + TN + FP} \tag{19}$$

$$SEN = \frac{TP}{TP + FN} \tag{20}$$

$$SPE = \frac{TN}{TN + FP} \tag{21}$$

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

3.2. Experiment results

Table 2 lists the classification results via different methods on AD vs. HC and MCI vs. HC. In AD vs. HC, our proposed method achieves classification accuracy of 96.10 %, sensitivity of 94.21 %, and specificity of 96.12 %. In MCI vs. HC, our method achieves classification accuracy of 84.67 %, sensitivity of 90.75 %, and specificity of 82.61 %. We also carried out experiments to predict MCI conversion. As an early 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. This is of important clinical and practical significance. In cMCI vs. sMCI, the classification accuracy, sensitivity and specificity of

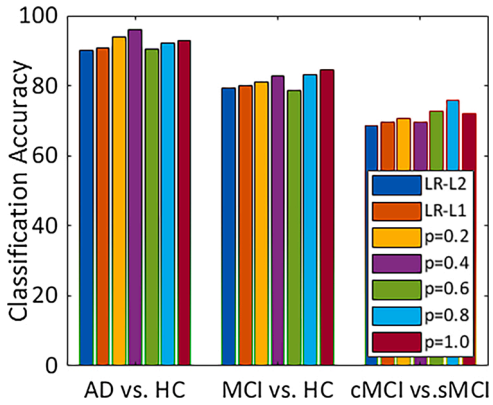

Fig. 2. The accuracy of different methods in three classification tasks.

our method are 75.87 %, 73.71 % and 77.51 % respectively.

4. Discussion

4.1. Evaluation with different p values

In order to assess the classification performance of our proposed method at different p-values, we compare it with other methods including LR with L_2 regularization and LR with L_1 regularization. The classification accuracy for three classification tasks is computed under the optimal parameters. Table 2 lists the classification results for the different p values on AD vs. HC, MCI vs. HC and cMCI vs. sMCI. As can be seen from the Table 2, when p is 0.4, our proposed achieves the best classification performance in AD vs. HC compared with other methods. In MCI vs. HC, our method obtains a good classification effect when p is 1. So, in different classification tasks, by choosing an appropriate p value, our proposed method can achieve good classification performance. In particular, in cMCI vs. sMCI, when p is 0.8, the classification accuracy obtained by our proposed method on cMCI and sMCI is 75.12 %. Comparing with logistic regression based on L_1 regularization, the classification accuracy achieved by our proposed method is more than nearly 4 percentage points. The classification result demonstrates that our proposed method can better identify cMCI from sMCI under the optimal p value and better predict the disease progression.

Fig. 2 shows the corresponding histogram of the proposed method in the three sets of classification tasks. Through the histogram, we can intuitively see that our proposed method can achieve a better classification performance under an appropriate p value, which shows the better diagnostic capabilities than other methods.

4.2. Evaluation with the parameters

In Eq. (5), the parameters $\lambda_1$and λ_2 control the degree of Lp regularization and L_2 regularization, respectively. By selecting optimal parameter values, the classifier can have better performance and select the most discriminating brain areas. We used ten-fold cross-validation in the training sets to select the optimal parameters $\lambda_1$and $\lambda_2\lambda_1$and λ_2 are selected from the set $\{10^{-2},10^{2}\}$. Every value in the span can participate in the calculation. Different $\lambda_1$and $\lambda_2$values produced different classification effects. Optimal values are selected by the highest classification accuracy. Fig. 3 shows the classification accuracy obtained by selecting different parameter values in the three classification tasks. Different colors represent different classification accuracy. The horizontal axis represents different values of λ_1. The vertical axis represents different values of λ_2.

From Fig. 3, we can see that in the classification of AD vs. HC, our proposed method can achieve the best classification accuracy when λ_1, $\lambda_2 \in \{10^{-2},1\}$. In MCI vs. HC, when $\lambda_1,\lambda_2 \in \{10^{-2},10^{-1}\}$, the classification result is optimal. In cMCI vs. sMCI, when $\lambda_1,\lambda_2 \in \{10^{-1},1\}$, the classification results within this range are stable and optimal. Therefore, the initializations of two parameters are set to the above three ranges

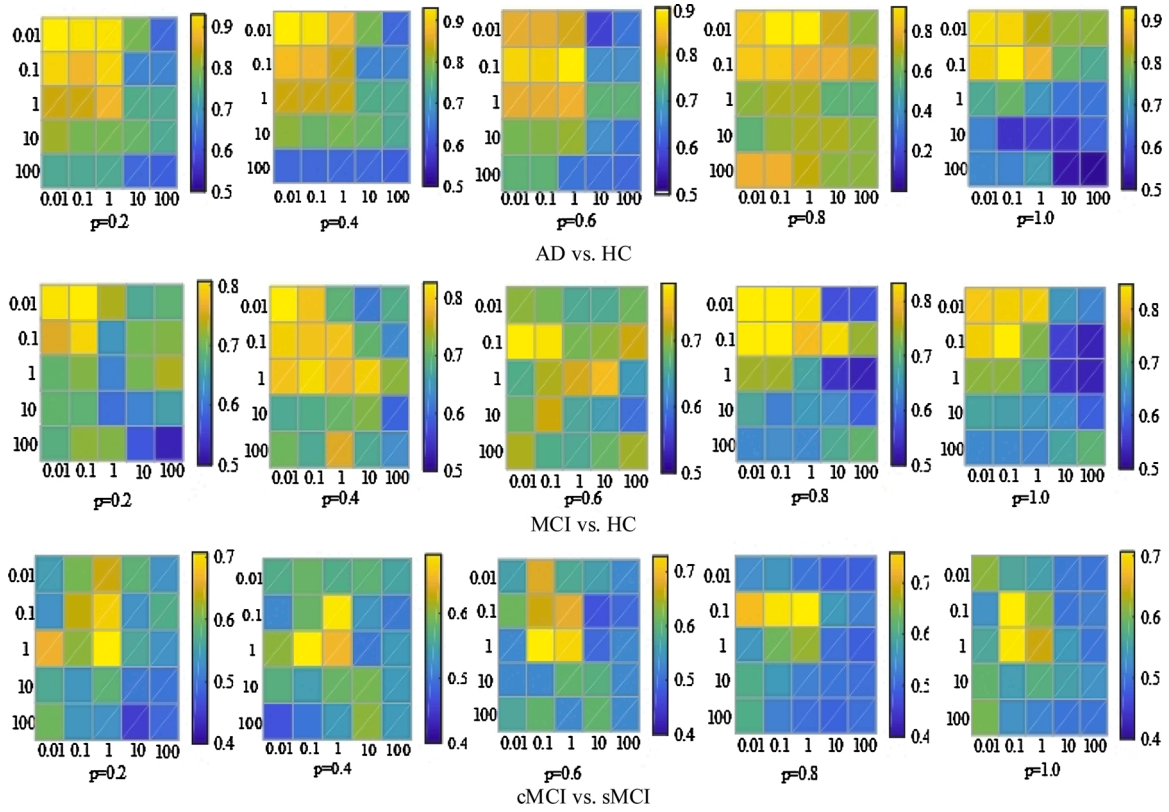

Fig. 3. Classification accuracy under different parameters.

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

Brain regions	Weight
Hippocampus_L	0.3478
Amygdala_L	0.3278
Middle temporal gyrus_L	0.3154
Precuneus_R	0.2791
ParaHippocampal gyrus_R	0.2477
Superior occipital gyrus_L	0.1941
Precentral gyrus_R	0.1921
Insula_L	0.1908
Angular_L	0.1867
Cuneus_L	0.1725

during the parameter selection process. In the experiment, our proposed method achieves the best classification performance for AD vs. HC when $p = 0.4$and $\lambda_1 = 0.01$,$\lambda_2 = 0.05$. In MCI vs. HC, when $p = 1$ and$\lambda_1 = \lambda_2 = 0.1$, our method obtains the high classification accuracy. In cMCI vs. sMCI, when $p = 0.8$ and$\lambda_1 = 0.5$,$\lambda_2 = 0.3$, our proposed method can better classify cMCI from sMCI. Throughout the above analysis, in different classification tasks, selecting optimal parameter values can improve classification performance under the appropriate p value.

4.3. The most discriminative brain regions

In addition to introducing the classification performance of our proposed method, we also report the AD-related brain regions selected by our method. To identify brain regions that can be regarded as potential imaging biomarkers, we first identified the important brain areas based on their weight. Each feature is multiplied by a feature weight, and we judge the importance of the feature according to the size of the weight. We select the feature of highest weight as the most discriminative feature. Table 3 summarizes brain regions that are the most discriminative to classify sMCI from cMCI. Fig. 4 plots the brain regions selected in cMCI vs. sMCI. They are known to be related to AD [31–35]. For example, Hippocampal atrophy, or shape change, is one of the main hallmarks of AD. The amygdala plays an important role in emotional expressions, memory processing and managing stimulatory input. The precuneus is a key area for memory impairment of AD patients. The above experimental results indicate that the generalized elastic net with optimal p value can select the most discriminative brain area and remove redundant information to boost the accuracy and better be applied to the early diagnosis of AD.

4.4. Comparison with other methods

To further evaluate the advantages of our proposed method, we list some representative methods in the recent years. Table 4 represents the classification results obtained by other methods, including SVM [36], DBN [37], Multi-task learning [38], Multiple Kernel Learning [39], Random Forest (RF) [40], SDPSO-SVM-PCA [41], CNN and RNN [42]. Although the size of the data set and the method of feature extraction may be different, the data come from the ADNI database. So, it is worth comparing the classification performance. In Table 4, our proposed method achieves better classification accuracies in three classification tasks. This further proves the effectiveness of our proposed method.

5. Conclusions

In this paper, the sparse logistic regression with the generalized elastic net is proposed for the early diagnosis of AD. The generalized elastic net is able to identify the most discriminative brain areas and improve the classification performance. We evaluate our proposed method based on ADNI datasets. It is worth noting that the accuracy of our proposed method for AD vs. HC and MCI vs. HC are 95.61 % and 84.67 % respectively. In particular, we used this method to perform classification experiments on cMCI and sMCI, and obtained a classification accuracy of 75.87 %, which is essential for MCI conversion prediction. More importantly, our method can find disease-related biomarkers that can be used for disease diagnosis. In future work, we will consider using better optimization algorithms to train the model to achieve faster prediction speed and better classification performance.

CRediT authorship contribution statement

Ruyi Xiao: Data curation, Writing - original draft. **Xinchun Cui:** Methodology, Writing - review & editing. **Hong Qiao:** Investigation. **Xiangwei Zheng:** Investigation. **Yiquan Zhang:** Validation. **Chenghui Zhang:** Supervision. **Xiaoli Liu:** Supervision.

Declaration of Competing Interest

The authors report no declarations of interest.

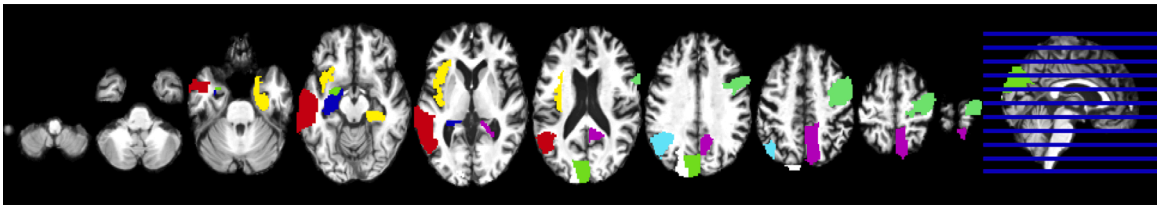

Fig. 4. The most discriminative brain regions identified by the proposed method in cMCI vs. sMCI.

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

Author	Methods	AD vs. HC (%)	Subjects (AD/HC)	MCI vs. HC (%)	Subjects (MCI/HC)	cMCI vs. sMCI (%)	Subjects (cMCI/sMCI)
Min et al., 2014 [36]	SVM	91.64	97/128	–	–	72.41	117/117
Li et al., 2015 [37]	DBN	91.40	51/52	77.40	99/52	57.40	43/56
Yu et al., 2016 [38]	Multi-task learning	92.60	50/52	80.00	97/52	–	–
Ahmedet et al., 2017 [39]	Multiple Kernel Learning	90.20	45/52	75.49	58/52	–	–
Ruiz et al., 2018 [40]	RF	82.79	188/229	71.92	401/229	67.39	86/312
Zeng et al., 2018 [41]	SDPSO-SVM-PCA	81.25	92/92	–	–	69.23	95/82
Basaia et al., 2019 [42]	CNN + RNN	91.33	198/229	–	–	71.71	167/236
–	proposed	96.10	51/50	84.67	96/50	75.87	51/45

Acknowledgments

This work is partially supported by National Nature Science Foundation of China (71971190); High-quality Course for Graduate Education of Shandong Province (Digital Image Processing, SDYKC19178). Shandong Social Science Planning Research Project (20CSDJ20).

References

[1] C.Y. Wee, P.T. Yap, D.Q. Zhang, et al., Identification of MCI individuals using structural and functional connectivity networks, Neuroimage 59 (3) (2012) 2045–2056.
[2] I. Beheshti, H. Demirel, H. Matsuda, et al., 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 Dement. 11 (3) (2015).
[4] C. Lian, M. Liu, J. Zhang, et al., Hierarchical Fully Convolutional Network for Joint Atrophy Localization and Alzheimer's Disease Diagnosis Using Structural MRI, 2018.
[5] Y. Cho, J. Seong, Y. Jeong, et al., Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data, Neuroimage 59 (3) (2012) 2217–2230.
[6] M. Liu, D. Zhang, D. Shen, Ensemble sparse classification of Alzheimer's disease, Neuroimage 60 (2) (2012) 1106–1116.
[7] B. Cheng, M. Liu, D. Shen, et al., Multi-domain transfer learning for early diagnosis of Alzheimer's disease, Neuroinformatics 15 (2) (2017) 115–132.
[8] Q. Zhu, N. Yuan, J. Huang, et al., Multi-modal AD classification via self-paced latent correlation analysis, Neurocomputing 355 (2019) 143–154.
[9] N. Zeng, H. Qiu, Z. Wang, et al., A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease, Neurocomputing 320 (2018) 195–202.
[10] Y. Liang, C. Liu, X.Z. Luan, et al., Sparse logistic regression with a L 1/2 penalty for gene selection in cancer classification, BMC Bioinformatics. 14 (1) (2013) 198.
[11] R. Wang, N. Xiu, C. Zhang, Greedy Projected gradient-Newton method for sparse logistic regression, IEEE Trans. Neural Netw. Learn. Syst. 31 (2) (2019) 527–538.
[12] 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 Netw. Learn. Syst. 28 (4) (2016) 917–933.
[13] F. Liu, X. Huang, C. Gong, et al., Indefinite kernel logistic regression with concave-inexact-convex procedure, IEEE Trans. Neural Netw. Learn. Syst. 30 (3) (2018) 765–776.
[14] K. Koh, S.J. Kim, S. Boyd, An interior-point method for large-scale l1-regularized logistic regression, J. Mach. Learn. Res. 8 (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, et al., Parameter selection for nonnegative l1 matrix/tensor sparse decomposition, Oper. Res. Lett. 43 (4) (2015) 423–426.
[17] Y. Wang, G. Zhou, L. Caccetta, et al., An alternative Lagrange-dual based algorithm for sparse signal reconstruction, IEEE Trans. Signal Process. 59 (4) (2010) 1895–1901.
[18] Y. Wang, W. Liu, L. Caccetta, et al., Parameter selection for nonnegative l1 matrix/tensor sparse decomposition, Oper. Res. Lett. 43 (4) (2015) 423–426.
[19] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Sat. Soc. B 67 (2) (2005) 301–320.
[20] S. Ryali, K. Supekar, D.A. Abrams, et al., Sparse logistic regression for whole-brain classification of fMRI data, Neuroimage 51 (2) (2010) 752–764.
[21] T.M. Schouten, M. Koini, F. deVos, et al., Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease, Neuroimage Clin. 11 (2016) 46–51.
[22] C.N. Li, M.Q. Shang, Y.H. Shao, et al., Sparse L1-norm two dimensional linear discriminant analysis via the generalized elastic net regularization, Neurocomputing 337 (2019) 80–96.
[23] R. Mazumder, J.H. Friedman, T. Hastie, Sparsenet: coordinate descent with nonconvex penalties, J. Am. Stat. Assoc. 106 (495) (2011) 1125–1138.
[24] J. Klimaszewski, M. Korzeń, Optimization of l p-regularized linear models via coordinate descent, Schedae Inform. 25 (2016) 61.
[25] T. Ye, C. Zu, B. Jie, et al., Discriminative multi-task feature selection for multi-modality classification of Alzheimer's disease, Brain Imaging Behav. 10 (3) (2016) 739–749.
[26] Q. Wang, Y. Zheng, G. Yang, et al., Multiscale rotation-invariant convolutional neural networks for lung texture classification, IEEE J. Biomed. Health Inform. 22 (1) (2017) 184–195.
[27] E. Moradi, A. Pepe, C. Gaser, et al., Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects, Neuroimage 104 (2015) 398–412.
[28] G.A. Papakostas, A. Savio, M. Graña, et al., A lattice computing approach to Alzheimer's disease computer assisted diagnosis based on MRI data, Neurocomputing 150 (2015) 37–42.
[29] R. Wang, N. Xiu, C. Zhang, Greedy projected gradient-Newton method for sparse logistic regression, IEEE Trans. Neural Netw. Learn. Syst. 31 (2) (2019) 527–538.
[30] A. Zakariya Yahya, L. Muhammad Hisyam, Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification, Comput. Biol. Med. 67 (2015) 136–145.
[31] G.F. Busatto, G.E. Garrido, O.P. Almeida, et al., A voxel-based morphometry study of temporal lobe gray matter reductions in Alzheimer's disease, Neurobiol. Aging 24 (2) (2003) 221–231.
[32] B. Jie, D. Zhang, B. Cheng, et al., Manifold regularized multitask feature learning for multimodality disease classification, Hum. Brain Mapp. 36 (2) (2015) 489–507.
[33] H.J.A. Matsuda, Disease, Voxel-based morphometry of brain MRI in normal aging and Alzheimer's disease, Aging Dis. 4 (1) (2013) 29–37.
[34] Y. Fan, N.K. Batmanghelich, C.M. Clark, et al., Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline, Neuroimage 39 (4) (2008) 1731–1743.
[35] C.Y. Wee, P. Yap, D.Q. Zhang, et al., Identification of MCI individuals using structural and functional connectivity networks, Neuroimage 59 (3) (2012) 2045–2056.
[36] R. Min, G. Wu, J. Cheng, et al., Multi-atlas based representations for Alzheimer's disease diagnosis, Hum. Brain Mapp. 35 (10) (2014) 5052–5070.
[37] F. Li, L. Tran, K.H. Thung, et al., A robust deep model for improved classification of AD/MCI patients, IEEE J. Biomed. Health Inform. 19 (5) (2015) 1610–1616.
[38] G. Yu, Y. Liu, D.G. Shen, Graph-guided joint prediction of class label and clinical scores for the Alzheimer's disease, Brain Struct. Funct. 221 (7) (2016) 3787–3801.
[39] O.B. Ahmed, J. Benois-Pineau, M. Allard, et al., Recognition of Alzheimer's disease and mild cognitive impairment with multimodal image-derived biomarkers and multiple kernel learning, Neurocomputing 220 (2017) 98–110.
[40] E. Ruiz, J. Ramirez, J.M. Górriz, et al., Alzheimer's disease computer-aided diagnosis: histogram-based analysis of regional MRI volumes for feature selection and classification, J. Alzheimer's Dis. 65 (3) (2018) 819–842.
[41] N. Zeng, H. Qiu, Z. Wang, et al., A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease, Neurocomputing 320 (2018) 195–202.
[42] R. Cui, M. Liu, RNN-based longitudinal analysis for diagnosis of Alzheimer's disease, Comput. Med. Imaging Graph. 73 (2019) 1–10.