Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-09738-0

Early diagnosis model of Alzheimer's Disease based on sparse logistic regression

Ruyi Xiao [1] · Xinchun Cui [1] · Hong Qiao [2] · Xiangwei Zheng [3] · Yiquan Zhang [1]

Received: 8 April 2020 / Revised: 14 August 2020 / Accepted: 26 August 2020
Published online: 25 September 2020

Abstract
Accurate classification of Alzheimer's Disease (AD) and its prodromal stage, i.e., mild cognitive impairment (MCI) are critical for the effective treatment of AD. However, compared with AD classification tasks, predicting the conversion of MCI to AD is relatively difficult. as there are only minor differences among MCI groups. What's more, in brain imaging analysis, the high dimensionality and relatively small number of subjects brings challenges to computer-aided diagnosis of AD and MCI. Many previous researches focused on the identification of imaging biomarkers for AD diagnosis. In this paper, we introduce sparse logistic regression for the early diagnosis of AD. Sparse logistic regression (SLR) uses $L_{1/2}$ regularization to impose a sparsity constraint on logistic regression. The $L_{1/2}$ regularization is considered a representative of Lq regularization, where fewer but informative key brain regions are applied for the classification of AD/MCI. We evaluated the SLR on 197 subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Experimental results showed that the SLR improves the classification performance of AD/MCI compared other classical methods.

Keywords Sparse logistic regression · Mild cognitive impairment · Alzheimer's disease · MR image

1 Introduction

Alzheimer's disease (AD) is a neurological brain disease, with symptoms such as loss of cognitive and memory, which severely affect people's daily life [40]. According to

✉ Xinchun Cui
cuixinchun@qfnu.edu.cn

[1] School of Information Science and Engineering, Qufu Normal University, Rizhao 276800, China

[2] Business School, Shandong Normal University, Jinan 250014, China

[3] School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China

statistics, there are 50 million AD patients in the world, which is likely to double by 2050 [30]. Mild cognitive impairment (MCI) is the transitional stage between healthy elderly and AD [15]. It is estimated that 15% to 20% of individuals over the age of 65 have MCI. Nearly 10% to 15% of MCI patients convert to AD each year [31]. So, an accurate prediction of future conversion from MCI to AD is essential for the early diagnosis of AD to delay the deterioration of the disease [17].

In the past few years, many machine learning methods have been used for the early detection of diseases [7, 8, 10, 11]. For example, Chiranji et al. [9] combined possibilistic fuzzy c-mean and intuitionistic fuzzy c-mean for the diagnosis of breast cancer, featuring high accuracy with clustering and breast cancer detection. Kauser et al. [1] proposed an integrated scheme for heart disease diagnosis which combined cuckoo search and rough set. Tripathy et al. [34] used rough set with formal concept analysis for intelligent medical diagnosis. In order to minimize the decision rules, they use two processes to mine suitable rules. Luck et al. [36] combined adaptive genetic algorithm and logistic regression for the early diagnosis of AD. Adaptive genetic algorithm was used for feature selection, and logistic regression was used for classification. Liu et al. [22] used multi-task learning approach for feature selection to solve the high-dimensional problem of neuroimaging data. A multi-kernel support vector machine was used for classification. Previous studies have shown that it is not only important but also challenging to extract the representative biomarkers for classification in high-dimensional neuroimage data analysis. Recently, logistic regression (LR) is a widely used method of machine learning, aiming at solving classification problems [3, 39]. In addition to the class label information it can obtain direct classification probabilities [21]. Faced with the problems mentioned above, many generalized models and methods have been proposed [23, 32]. Zhang et al. [48] used L_2-Regularized logistic regression for MCI classification. However, the L_2 regularization did not produce sparse solutions. Koh et al [19] 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 multiple important features simultaneously [37]. However, L_1 regularization may make the estimated parameters biased [26]. For the above problems, Xu et al. [42] proposed the $L_{1/2}$ regularization. As a representative of Lq ($0 < q < 1$)regularization, the $L_{1/2}$ regularization boasts decent properties, such as unbiasedness, sparsity and oracle properties [43]. In this paper, we used logistic regression with $L_{1/2}$ regularization for the early diagnosis of AD. SLR can better adapt to the characteristics of high-dimensional small samples of AD datasets and achieve better classification performance with minimal features. The main contributions of this paper include: (1) The $L_{1/2}$ regularization was used to select the most discriminative features for classification (2) The SLR was applied to predict the conversion of MCI to AD, allowing timely treatment to delay disease progression (3) Experimental data was collected from ADNI database. Through the Automated Anatomical Labeling (AAL) template [35], the brain was divided into 90 regions of interest (ROI). We extracted the gray matter volume of 90 regions as features.

The rest of this work is arranged as follows. The second part of this paper introduced related works. The third part described the proposed method. The fourth part presented the experimental results with a discussion. Finally, the last part of the presentation offered a conclusion.

2 Related works

In this paper, we implemented LR for the binary classification. Let $x_i = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in})^T$ be the i-th sample vector of matrix $\mathbf{X}$. Then the formula of the logistic regression model is given by:

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

where, $\theta = (\theta_0, \theta_1, \theta_2, \ldots, \theta_n)$ represents coefficient matrix of $(n+1)\times 1$. π_i is the predicted probability. $y_i \in \{0, 1\}$ is a response variable, which is then obtained by $I(\pi_i > 0.5)$, where $I(.)$ is an indicator function.

The log-likelihood can be expressed as:

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

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

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

In brain imaging analysis, logistic regression is prone to overfitting caused by high dimensionality but with a small number of training samples. To reduce the number of features and obtain a robust classifier, the penalization techniques for logistic regression is given by:

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

where, $\lambda > 0$ is a tuning parameter and $P(\theta_j)$ is the regularization term. Many regularizations methods are proposed to solve overfitting, such as LASSO, SCAD and Elastic net regularization.

3 Methods

Figure 1 is an overview of method consisting of: (1) image preprocessing and feature extraction (2) SLR model-training (3) SLR-based classification of AD, MCI and HC. The following is a detailed description of the approach.

3.1 ADNI database

The experimental data set used in this paper was 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. As a non-profit organization [44], ADNI provides unlimited data access and encourages researchers to develop potential methods for analyzing the progression of early AD. Magnetic resonance imaging (MRI) is a widely used imaging mode in the diagnosis of AD [28, 29, 38], a modality that facilitates comparisons between different soft tissues. Therefore, we used structural MR 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) who converted to AD within 24 months and 45 stable MCI (sMCI) who has not converted to AD within 24 months) Table 1 presents detailed information about these subjects.

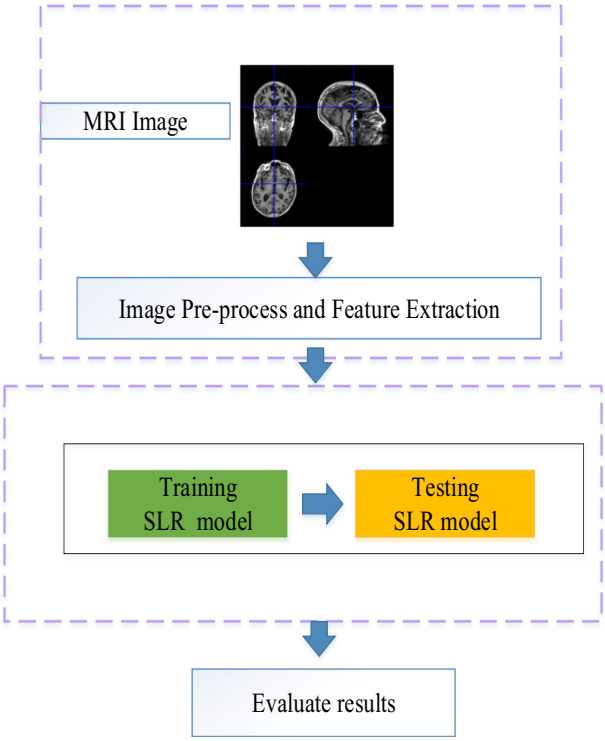


Fig. 1 Overview of the method

3.2 Image preprocessing and feature extraction

The MRI images downloaded from the ADNI database required a series of image preprocessing, that extracted the gray matter volume of 90 regions of interest as effective features. The effective features were input into the classifier for classification. The overview of image preprocessing and feature extraction are shown in Fig. 2.

In this paper, we used SPM8 software (http://www.fil.ion.ucl.ac.uk/spm/) and VBM8 toolbox to preprocess the MRI images. The preprocessing process consisted of 5 steps, i.e., (1) skull stripping (removing non-brain tissue), (2) spatial standardization and segmentation (grey matter (GM), white matter (WM) and cerebrospinal fluid (GSF)), (3) modulation (The density feature

Table 1 Statistical information of subjects

Diagnosis	Subjects	Age (Mean ± Standard Deviation)	Gender(F/M)	MMSE (Mean ± Standard Deviation)	CDR (Mean ± Standard Deviation)
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

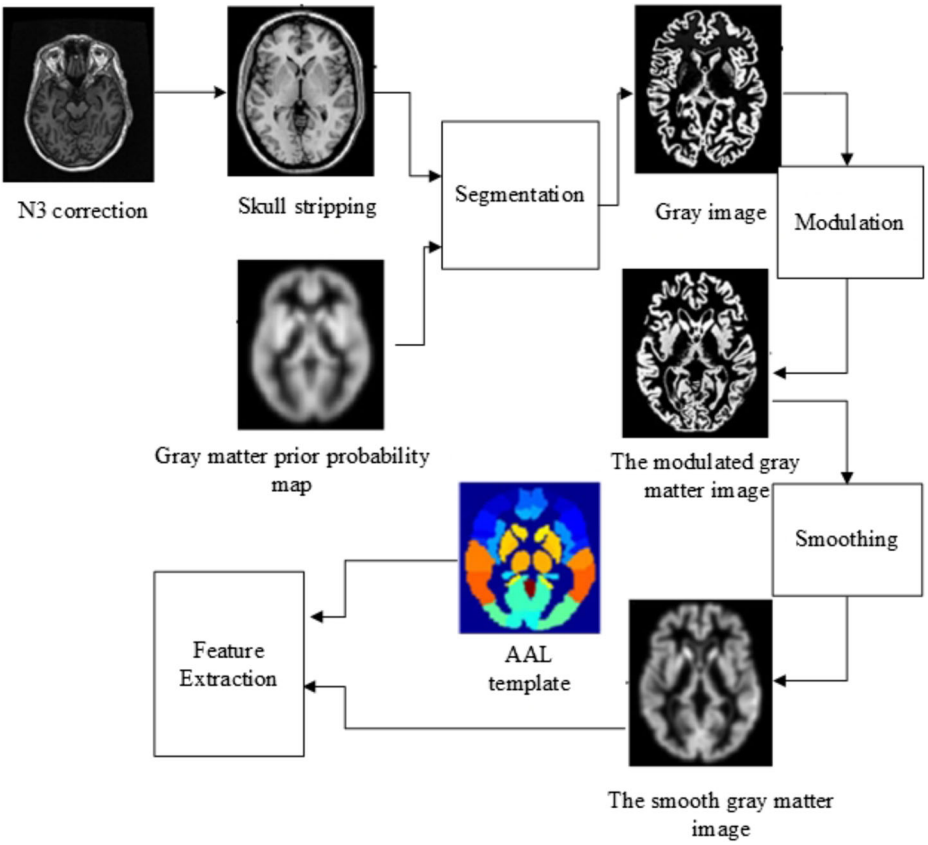


Fig. 2 Image preprocessing and feature extraction

was transformed into a volume feature), (4) smoothing (removing noise from the image), (5) registration (registering each subject's gray matter volume map onto an AAL template) (6) selecting the gray matter volume of the 90 auto-labeled regions of interest (ROI) as features.

3.3 Sparse logistic regression model

The $L_{1/2}$ regularization can yield the most sparse solution and select important features for classification. The loss function of a logistic regression with $L_{1/2}$ regularization is given by:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left(y_i \log(\pi_i) + (1-y_i)\log(1-\pi_i)\right) + \lambda \sum_{j=1}^{n} \left|\theta_j\right|^{1/2}. \quad (5)$$

Where, λ is regularization parameters that control the model's sparsity and group effect [46]. In this paper, λ was adjusted by ten-fold cross-validation (CV) in the training set.

3.4 A coordinate descent algorithm for sparse logistic regression

The $L_{1/2}$ regularization is non-convex. In solving non-convex problems, coordinate descent algorithm showed significant efficiency and convergence [4]. The coordinate descent

algorithm [16] is a "one -at -a-time" approach. The univariate half thresholding operator of the coordinate descent algorithm is as follows.

$$\theta_j = Half\left(\omega_j, \lambda\right) = \begin{cases} \frac{2}{3}\omega_j\left(1+\cos\left(\frac{2\left(\pi-\varphi\left(\omega_j\right)\right)}{3}\right)\right) & if\left|\omega_j\right| > \frac{3}{4}(\lambda)^{\frac{2}{3}} \\ 0 & otherwise \end{cases}. \tag{6}$$

The Eq. (3) is linearized by one-term Taylor series expansion:

$$L(\theta,\lambda) \approx \frac{1}{2n}\sum_{i=1}^{n}(Z_i - X_i\theta)\, W_i(Z_i - X_i\theta) + \lambda\sum_{j=1}^{n}|\theta_j|^{1/2}. \tag{7}$$

Where $Z_i = X_i\widetilde{\theta} + \frac{Y_i - f\left(X_i\widetilde{\theta}\right)}{f\left(X_i\widetilde{\theta}\right)\left(1-f\left(X_i\widetilde{\theta}\right)\right)}$ is the estimated response, $\omega_i = f\left(X_i\widetilde{\theta}\right)\left(1-f\left(X_i\widetilde{\theta}\right)\right)$ is the weight, $f\left(X_i\widetilde{\theta}\right) = \exp\left(X_i\widetilde{\theta}\right)/\left(1+\exp\left(X_i\widetilde{\theta}\right)\right)$ is a evaluate value under current parameters. The procedure of the coordinate descent algorithm for the SLR is given by:

Algorithm: The coordinate descent algorithm for sparse logistic regression

Step1	Initialize all $\theta_j = 0\ (j = 1,2,...,p)$ and X,y and λ are chosen by 10-fold cross-validation.
Step2	Calculate $Z(m)$, $W(m)$ and the loss function Eq (7) based on $\theta(m)$.
Step3	Update each $\theta_j(m)$, and cycle over $j = 1,2,...,p$.
Step3.1	Calculate $Z_i^{(j)} \leftarrow \sum_{k\neq j} x_{ik}\theta_k(m)$ and $\omega_j(m) \leftarrow \omega_j(m)x_{ij}(Z_i(m) - Z_i^{(j)}(m))$.
Step3.2	Update $\theta_j(m)$ by Eq (6).
Step 4	Let $m = m+1$, $\theta(m+1) = \theta(m)$, repeat Step2,3 until $\sum_{i=1}^{p}(\left\|\theta_i(m+1)\right\| - \left\|\theta_i(m)\right\| < 10^{-8})$.

4 Experiment results

4.1 Experiment setting

In this paper, we implemented 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). In order to achieve a fair comparison, two procedures were set up. First, each dataset was randomly distributed into training sets and test sets, at a ratio of 7:3. Second, ten-fold cross-validation was performed in the training sets to select optimal parameter λ. To avoid bias caused by random distribution of samples, we repeated the experiment 50 times. The classification performance of all methods was quantified by calculating accuracy (ACC), sensitivity (SEN), specificity (SPE), receiver operating characteristic curve (ROC) and area under the receiver operating characteristic (AUC).

4.2 Classification results

This part presents a summary and the discussion of the results. To reduce feature dimensions and select the discriminate brain regions, we used the $L_{1/2}$ regularization to ensure the sparse solution. We compared the classification performance of SLR on AD with LR, LR with L_2 regularization (LR-L_2) and LR with L_1 regularization (LR-L_1).

Table 2 lists the classification results via different methods on AD vs. HC and MCI vs. HC. In AD vs. HC, the SLR had a classification accuracy of 93.33%, sensitivity of 92.25%, specificity of 94.75% and the AUC of 0.96, higher than those of L_2 and L_1. In MCI vs. HC, the SLR had a classification accuracy of 82.75%, sensitivity of 86.67%, specificity of 78.57% and the AUC of 0.89. The classification performance of SLR showed a better classification performance than other methods.

Figure 3 shows the corresponding ROC curves. Through the ROC curve, we can further investigate the performance of SLR on AD classification. SLR performs better than all the other methods in the two classification tasks, indicating that $L_{1/2}$ regularization can find out the most discriminant features to boost the accuracy of SLR.

We also carried out experiments on classification 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, with clinical and practical significance. The corresponding results are shown in Table 3.

Table 2 Comparison of different methods in AD/MCI classification

Method	AD vs. HC				MCI vs. HC			
	ACC (%)	SEN (%)	SPE (%)	AUC	ACC (%)	SEN (%)	SPE (%)	AUC
LR	80.00	78.75	82.85	0.87	70.21	75.75	57.14	0.74
LR-L_2	90.25	87.50	92.85	0.93	79.41	85.89	74.75	0.84
LR-L_1	90.79	88.21	92.85	0.94	80.10	86.21	75.87	0.87
SLR	93.33	92.25	94.75	0.96	82.75	86.67	78.57	0.89

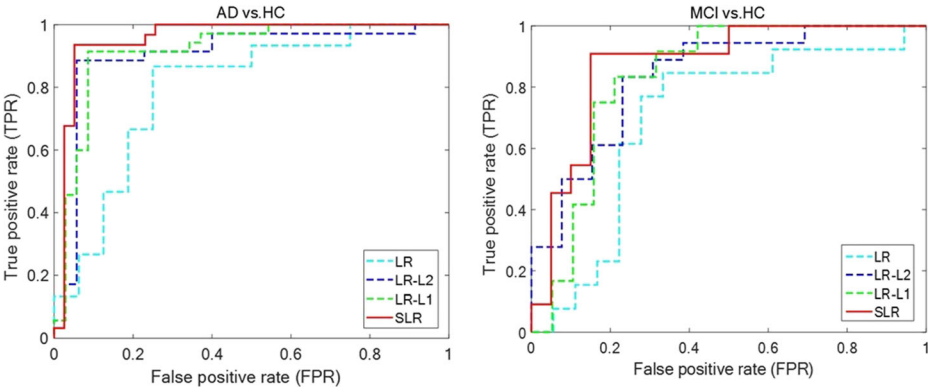


Fig. 3 ROC curves of different methods for classification of AD and MCI

From Table 3, in cMCI vs. sMCI, the classification accuracy, sensitivity and specificity of SLR are 72.87%, 81.25%, 65.45% respectively. Compared with LR-L_1, the classification accuracy of SLR is more than 3 percentage points higher.

Figure 4 shows the ROC curve obtained by different methods in cMCI and sMCI classification. It can be further proved by the ROC curve and the AUC values shown in Table 3 that SLR can cope with challenges brought by the early diagnosis of AD, and is better tool for cMCI and sMCI classification.

4.3 The most discriminative brain regions

In addition to introducing the classification performance, we also reported the AD-related brain regions selected by the SLR method. Figure 5 plots some frequently selected brain regions for AD vs. HC, which are known to be related to AD [6, 13, 18, 25, 41]. For example, the hippocampus is related to human memory and learning. It is the first brain region to be damaged in relation to AD disease. The amygdala controls people's emotions and cognition. Some studies also investigated the role of the parahippocampal gyrus and precuneus in patients with AD [14, 47].

4.4 Comparison with other methods

To further reflect the advantages of SLR, we listed some classical methods in recent years. Table 4 shows the classification results obtained by other methods. In Table 4, the SLR always has the highest classification accuracies in AD vs. HC, MCI vs. HC and cMCI vs. sMCI., further proving the advantages of SLR in AD classification.

Table 3 Comparison of different methods incMCI andsMCI

Method	cMCI vs. sMCI			
	ACC (%)	SEN (%)	SPE (%)	AUC
LR	56.25	62.50	50.00	0.61
LR-L_2	68.75	73.35	62.50	0.68
LR-L_1	69.45	75.00	64.01	0.71
SLR	72.87	81.25	65.45	0.75

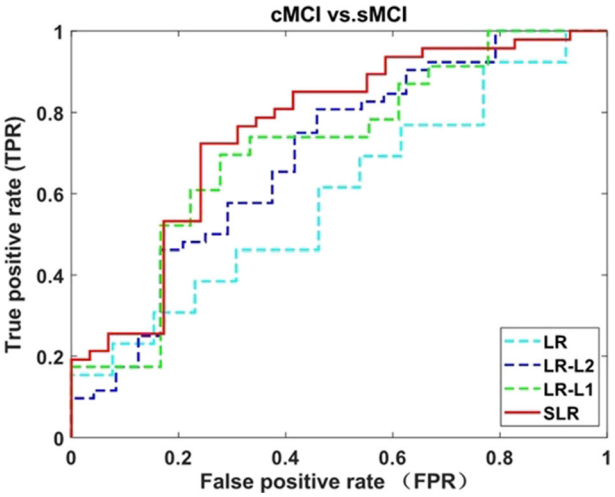
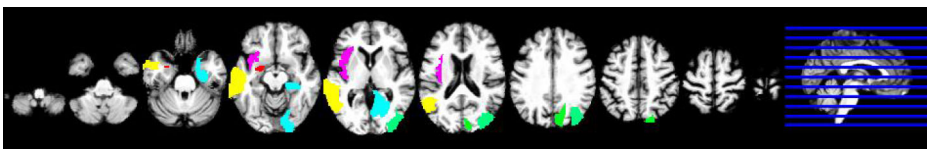


Fig. 4 ROC curves of different methods in cMCI/sMCI classification

4.5 General discussion

In this study, we used sparse logistic regression for AD vs. HC, MCI vs. HC and cMCI vs. sMCI classifications. It is worth noting that the SLR method obtains classification accuracies of 93.33% and 82.75% for AD vs. HC and MCI vs. HC classification with the most discriminative features selecting by $L_{1/2}$ regularization. This proves the $L_{1/2}$ regularization a useful method to discover the disease-related biomarkers and improve classification performance. As to predicting the conversion of MCI to AD, a challenge task to the study, it is necessary to distinguish cMCI from sMCI, so that sMCI subjects can get timely treatment for possible delay of progressing to AD. The classification accuracy obtained by SLR in cMCI vs. sMCI is 72.87%, an improvement of 4% (vs. LR-L_2) and 3% (vs. LR-L_1) respectively. SLR has a desirable performance for MCI conversion prediction. We also compared the classification accuracy of SLR with other studies. Table 4 lists the corresponding comparison results. The performance of the SLR was highly competitive compared to other methods reported in the literature for AD classification and the prediction of MCI-to-AD conversion.

Although the proposed method demonstrated a desirable performance for AD detection and MCI conversion prediction, there are also some limitations in the present study. Firstly, we used only single-mode data for our experiments. Secondly, this article only took into account the two-class classification scenario (i.e., AD vs. HC, MCI vs. HC, cMCI vs. sMCI), while SLR for muli-class classification was ignored. In the future study, we will use multimodal data for experiments and test the proposed model for three classifications.



AD vs.HC

Fig. 5 The most discriminative brain regions identified by the proposed SLR method for AD vs. HC

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

Author	Methods	AD vs. HC (%)	MCI vs. HC (%)	cMCI vs. sMCI (%)
Min et al. 2014 [27]	SVM	91.64	–	72.41
Li et al. 2015 [20]	DBN	91.40	77.40	57.40
Yu et al. 2016 [45]	Multi-task learning	92.6	80.00	–
Ahmedet et al. 2017 [2]	Multiple Kernel Learning	90.20	75.49	–
Ruiz et al. 2018 [33]	Random forest	82.79	71.92	67.39
Basaia et al. 2019 [12]	CNN + RNN	91.33	–	71.71
Liu et al. 2020 [24]	Multi-model deep learning framework	88.90%	76.25	–
Jie et al. 2020 [5]	Embedded feature selection+ SVM	–	–	71.17%
–	SLR	93.33	82.75	72.87

5 Conclusion

In this paper, we investigated the efficiency of SLR method in identifying AD subjects, MCI subjects, and HC subjects. To reduce feature dimensions and find out the discriminative brain regions, we used the $L_{1/2}$ regularization to produce sparse solution. We evaluated the SLR method based on ADNI dataset. The experimental results indicated that the performance of the SLR can compete strongly with state-of-the-art techniques. In the future study, we will consider further optimization of the logistic regression model to improve the classification performance of the model, and better apply it in AD diagnosis.

Acknowledgments This work is partially supported by National Nature Science Foundation of China (71971190); Ministry of Education Humanities and Social Science Project (11YJCZH021, 15YJCZH111). Shandong Social Science Planning Research Project (17CHLJ41, 16CTQJ02, 18CHLJ34).

Compliance with ethical standards

Conflict of interest The authors declare no financial or commercial conflict of interest.

References

1. Acharjya D (2020) A hybrid scheme for heart disease diagnosis using rough set and cuckoo search technique. J Med Syst 44(1):27
2. Ahmed OB, Benois-Pineau J, Allard M, Catheline G, Amar CB, Initiative AsDN (2017) Recognition of Alzheimer's disease and mild cognitive impairment with multimodal image-derived biomarkers and multiple kernel learning. Neurocomputing 220:98–110
3. Algamal ZY, Lee MH (2015) Penalized logistic regression with the adaptive LASSO for gene selection in high-dimensional cancer classification. Expert Syst Appl 42(23):9326–9332
4. Breheny P, Huang J (2011) Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. Ann Appl Stat 5(1):232–253
5. Cai J, Hu L, Liu Z, Zhou K, Zhang H (2020) An embedded feature selection and multi-class classification method for detection of the progression from mild cognitive impairment to Alzheimer's disease. J Med Imaging Health Inform 10(2):370–379
6. Cheng B, Liu M, Shen D, Li Z, Zhang D, Initiative AsDN (2017) Multi-domain transfer learning for early diagnosis of Alzheimer's disease. Neuroinformatics 15(2):115–132
7. Chowdhary CL, Acharjya D (2016) A hybrid scheme for breast cancer detection using intuitionistic fuzzy rough set technique. Int J Healthcare Info Syst Inform 11(2):38–61

8. Chowdhary CL, Acharjya D (2017) Clustering algorithm in possibilistic exponential fuzzy c-mean segmenting medical images. J Biomimet Biomater Biomed Eng 30:12–23
9. Chowdhary CL, Acharjya D (2018) Segmentation of mammograms using a novel intuitionistic possibilistic fuzzy c-mean clustering algorithm. Nature Inspired Computing 75-82
10. Chowdhary CL, Acharjya D (2020) Segmentation and feature extraction in medical imaging: a systematic review. Procedia Comp Sci 167:26–36
11. Chowdhary CL, Sai GVK, Acharjya D (2016) Decrease in false assumption for detection using digital mammography. Comput Intell Data Min 2:325–333
12. Cui R, Liu M, Imaging AsDNIJCM, Graphics (2019) RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. Comput Med Imaging Graph 73:1–10
13. Dai W, Lopez OL, Carmichael OT, Becker JT, Kuller LH, Gach HM (2009) Mild cognitive impairment and alzheimer disease: patterns of altered cerebral blood flow at MR imaging. Radiology 250(3):856–866
14. Del Sole A, Clerici F, Chiti A, Lecchi M, Mariani C, Maggiore L, Mosconi L, Lucignani G (2008) Individual cerebral metabolic deficits in Alzheimer's disease and amnestic mild cognitive impairment: an FDG PET study. Eur J Nucl Med Mol Imaging 35(7):1357–1366
15. Eshkoor SA, Hamid TA, Mun CY, Ng CK (2015) Mild cognitive impairment and its management in older people. Clin Interv Aging 10:687
16. Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33(1):1
17. Gómez-Sancho M, Tohka J, Gómez-Verdejo V, Initiative AsDN (2018) Comparison of feature representations in MRI-based MCI-to-AD conversion prediction. Magn Reson Imaging 50:84–95
18. Karas G, Scheltens P, Rombouts S, Van Schijndel R, Klein M, Jones B, Van Der Flier W, Vrenken H, Barkhof FJN (2007) Precuneus atrophy in early-onset Alzheimer's disease: a morphometric structural MRI study. Neuroradiology 49(12):967–976
19. Koh K, Kim S-J, Boyd S (2007) An interior-point method for large-scale l1-regularized logistic regression. J Mach Learn Res 8:1519–1555
20. Li F, Tran L, Thung K-H, Ji S, Shen D, Li J (2015) A robust deep model for improved classification of AD/MCI patients. IEEE J Biomed health Inform 19(5):1610–1616
21. Liang Y, Liu C, Luan X-Z, Leung K-S, Chan T-M, Xu Z-B, Zhang H (2013) Sparse logistic regression with a L 1/2 penalty for gene selection in cancer classification. BMC Bioinform 14(1):198
22. Liu F, Wee C-Y, Chen H, Shen D (2014) Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's disease and mild cognitive impairment identification. NeuroImage 84:466–475
23. Liu F, Huang X, Gong C, Yang J, Suykens JA (2018) Indefinite kernel logistic regression with concave-inexact-convex procedure. IEEE Transact Neur Netw Learn Syst 30(3):765–776
24. Liu M, Li F, Yan H, Wang K, Ma Y, Shen L, Xu M, Initiative AsDN (2020) A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. NeuroImage 208:116459
25. Matsuda H (2013) Voxel-based morphometry of brain MRI in normal aging and Alzheimer's disease. Aging Dis 4(1):29–37
26. Meinshausen N, Yu B (2009) Lasso-type recovery of sparse representations for high-dimensional data. Ann Stat 37(1):246–270
27. Min R, Wu G, Cheng J, Wang Q, Shen D, Initiative AsDN (2014) Multi-atlas based representations for Alzheimer's disease diagnosis. Hum Brain Mapp 35(10):5052–5070
28. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J, Initiative AsDN (2015) Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. Neuroimage 104:398–412
29. Papakostas GA, Savio A, Graña M, Kaburlasos VG (2015) A lattice computing approach to Alzheimer's disease computer assisted diagnosis based on MRI data. Neurocomputing 150:37–42
30. Patterson C (2018) The state of the art of dementia research: new frontiers. World Alzheimer Report
31. Petersen RC (2016) Mild cognitive impairment. CONTINUUM: Lifelong Learn Neurol 22(2):404
32. Qiu Z, Miller DJ, Kesidis G (2016) A maximum entropy framework for semisupervised and active learning with unknown and label-scarce classes. IEEE Transact Neur Netw Learn Syst 28(4):917–933
33. Ruiz E, Ramírez J, Górriz JM, Casillas J, Initiative AsDN (2018) Alzheimer's disease computer-aided diagnosis: histogram-based analysis of regional MRI volumes for feature selection and classification. J Alzheimers Dis 65(3):819–842
34. Tripathy B, Acharjya D, Cynthya V (2013) A framework for intelligent medical diagnosis using rough set with formal concept analysis
35. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15(1):273–289

36. Vandewater L, Brusic V, Wilson W, Macaulay L, Zhang P (2015) An adaptive genetic algorithm for selection of blood-based biomarkers for prediction of Alzheimer's disease progression. BMC Bioinform 16(18):S1
37. Wang Y, Liu W, Caccetta L, Zhou G (2015) Parameter selection for nonnegative l1 matrix/tensor sparse decomposition. Oper Res Lett 43(4):423–426
38. Wang Q, Zheng Y, Yang G, Jin W, Chen X, Yin Y (2017) Multiscale rotation-invariant convolutional neural networks for lung texture classification. IEEE J Biomed health Inform 22(1):184–195
39. Wang R, Xiu N, Zhang C (2019) Greedy projected gradient-Newton method for sparse logistic regression. IEEE Transact Neur Netw Learn Syst 31(2):527–538
40. Wee C-Y, Yap P-T, Zhang D, Denny K, Browndyke JN, Potter GG, Welsh-Bohmer KA, Wang L, Shen DJN (2012) Identification of MCI individuals using structural and functional connectivity networks. Neuroimge 59(3):2045–2056
41. Wee CY, Yap PT, Shen D, Initiative AsDN (2013) Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns. Hum Brain Mapp 34(12):3411–3425
42. Xu Z, Zhang H, Wang Y, Chang X, Liang Y (2010) L 1/2 regularization. SCIENCE CHINA Inf Sci 53(6): 1159–1169
43. Xu Z, Zhang H, Wang Y, Chang X, Liang YJSCIS (2010) L 1/2 regularization. SCIENCE CHINA Inf Sci 53(6):1159–1169
44. Ye T, Zu C, Jie B, Shen D, Zhang D, Initiative AsDN (2016) Discriminative multi-task feature selection for multi-modality classification of Alzheimer's disease. Brain Imaging Behav 10(3):739–749
45. Yu G, Liu Y, Shen D (2016) Graph-guided joint prediction of class label and clinical scores for the Alzheimer's disease. Brain Struct Funct 221(7):3787–3801
46. Zakariya Yahya A, Muhammad Hisyam L (2015) Regularized logistic regression with adjusted adaptive elastic net for gene selection in high dimensional cancer classification. Comput Biol Med 67:136–145
47. Zhang D, Wang Y, Zhou L, Yuan H, Shen D, Initiative AsDN (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage 55(3):856–867
48. Zhang X, Hu B, Ma X, Xu L (2015) Resting-state whole-brain functional connectivity networks for mci classification using l2-regularized logistic regression. IEEE Transact Nanobiosci 14(2):237–247