



# Large Margin and Local Structure Preservation Sparse Representation Classifier for Alzheimer's Magnetic Resonance Imaging Classification

Runmin Liu<sup>1</sup>, Guangjun Li<sup>1\*</sup>, Ming Gao<sup>2</sup>, Weiwei Cai<sup>3,4</sup> and Xin Ning<sup>5</sup>

<sup>1</sup> College of Sports Engineering and Information Technology, Wuhan Sports University, Wuhan, China, <sup>2</sup> College of Sports Science and Technology, Wuhan Sports University, Wuhan, China, <sup>3</sup> School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China, <sup>4</sup> AiTech Artificial Intelligence Research Institute, Changsha, China, <sup>5</sup> Institute of Semiconductors, Chinese Academy of Sciences, Beijing, China

## OPEN ACCESS

### Edited by:

Yuanpeng Zhang,  
Nantong University, China

### Reviewed by:

Lijun Xu,  
Nanjing Institute of Technology (NJIT),  
China  
Min Shi,  
Fuzhou University of International  
Studies and Trade, China

### \*Correspondence:

Guangjun Li  
liguangjun@whsu.edu.cn

### Specialty section:

This article was submitted to  
Alzheimer's Disease and Related  
Dementias,  
a section of the journal  
Frontiers in Aging Neuroscience

**Received:** 08 April 2022

**Accepted:** 09 May 2022

**Published:** 25 May 2022

### Citation:

Liu R, Li G, Gao M, Cai W and  
Ning X (2022) Large Margin and Local  
Structure Preservation Sparse  
Representation Classifier  
for Alzheimer's Magnetic Resonance  
Imaging Classification.  
Front. Aging Neurosci. 14:916020.  
doi: 10.3389/fnagi.2022.916020

Alzheimer's disease (AD) is a progressive dementia in which the brain shrinks as the disease progresses. The use of machine learning and brain magnetic resonance imaging (MRI) for the early diagnosis of AD has a high probability of clinical value and social significance. Sparse representation classifier (SRC) is widely used in MRI image classification. However, the traditional SRC only considers the reconstruction error and classification error of the dictionary, and does not consider the global and local structural information between images, which results in unsatisfactory classification performance. Therefore, a large margin and local structure preservation sparse representation classifier (LMLS-SRC) is developed in this manuscript. The LMLS-SRC algorithm uses the classification large margin term based on the representation coefficient, which results in compactness between representation coefficients of the same class and a large margin between representation coefficients of different classes. The LMLS-SRC algorithm uses local structure preservation term to inherit the manifold structure of the original data. In addition, the LMLS-SRC algorithm imposes the  $\ell_{2,1}$ -norm on the representation coefficients to enhance the sparsity and robustness of the model. Experiments on the KAGGLE Alzheimer's dataset show that the LMLS-SRC algorithm can effectively diagnose non AD, moderate AD, mild AD, and very mild AD.

**Keywords:** Alzheimer's disease, sparse representation classifier, image classification, magnetic resonance imaging, KAGGLE Alzheimer's dataset

## INTRODUCTION

Alzheimer's disease (AD) is a chronic progressive neurodegenerative disease that usually progresses slowly in the early stages and gets worse over time (Katabathula et al., 2021). AD often occurs in the elderly. The initial symptoms are easy to forget recent events. With the development of the disease, the symptoms may include language problems, disorientation, mood swings, loss of self-care ability, etc., which will eventually seriously affect the daily life of the elderly. Currently, about 90 million people worldwide suffer from AD of varying degrees. It is estimated that by 2050, the number of AD patients will reach 300 million (Wong, 2020). The specific symptoms of very mild AD are progressive decline in memory or other cognitive functions, but do not affect the ability of daily living. According to statistics, about 10–15% of very mild AD will eventually transform into AD (Porsteinsson et al., 2021). Current scientific and clinical research has not yet clearly identified the

pathogenesis and etiology of AD, and there is no fully effective treatment drug. AD is uncontrollable and irreversible after being diagnosed. However, if patients can be intervened and treated in the early stage of mild cognitive impairment (MCI), it is hoped that the onset of AD will be delayed by 5 years, and even stop the progression of AD in the stage of MCI, and no longer worsen into AD, reducing the number of patients with AD by 40% (Venugopalan et al., 2021).

In the past decade, neuroimaging techniques have been widely used in the classification and prediction of AD. Among them, magnetic resonance imaging (MRI) is a non-contact imaging technology that can provide detailed three-dimensional anatomical images of the brain and provide effective information for the classification and prediction of AD (Al-Khuzai and Duru, 2021). The AD classification algorithms based on machine learning usually extract the required features from the collected medical images by manual or semi-manual methods. Various parts of the brain regions of AD patients will atrophy to varying degrees due to the progression of the disease process. The volume, shape and texture information of the hippocampus, gray matter, white matter, and cerebral cortex of the brain are important features to distinguish AD and healthy people (Lee et al., 2020; Gao, 2021). To classify AD MRI images, some studies extract the volume information of the whole brain or part of the brain. Some scholars segment different regions of the brain and take the volume of each segment as features. According to the anatomical automatic labeling brain region template, some researchers divide the entire brain or part of the brain region into multiple regions and then obtain the features for each region. AD Patients often experience cerebral cortex atrophy and ventricular enlargement, and early AD patients usually have hippocampal atrophy (van Oostveen and de Lange, 2021). Therefore, some scholars use the volume information of different regions of interest such as the hippocampus as features based on medical prior knowledge. Another common feature extraction method is the morphometric measurement method, which is often implemented based on MRI images and PET images. For example, Al-Khuzai and Duru (2021) took the overall shape of the brain in MRI images as features. Katabathula et al. (2021) used the shape information of the hippocampus as features. Brain gully dilation is often seen in AD patients. Furthermore, texture features are also widely used in MRI images. Gao (2021) extracted the grayscale co-occurrence matrix of images as features. Hett et al. (2018) used 3D Gabor filter to extract and classify multi-directional texture features of MRI images.

Classifiers such as sparse representation classifier (SRC), logistic regression (LR), support vector machine (SVM), and decision tree (DT) are widely used in AD MRI image classification. For example, Kruthika et al. (2019) used a multi-level classifier to classify AD MRI images. They first used a naive Bayes classifier, and then used SVM as secondary classification to classify the data with confidence lower than the threshold. Liu et al. (2015) proposed a multi-view learning algorithm based on inherent structure of mild cognitive impairment (MCI) MRI images, which used the multi-view features of MCI images to train multiple SVMs, and then fused and discriminated each classifier result. Altaf et al. (2018) used SVM, random forest, and

*K*-nearest neighbor (KNN) to train AD classifiers, respectively, and the final classification result was the weighted sum of the results of each classifier. Yao et al. (2018) used the idea of hierarchical classification to classify AD MRI images. They initially classified samples into four classes (AD, healthy, MCI, converted MCI), then they trained several binary classifiers (AD and converted MCI, healthy and MCI), and finally got a classifier that can classify all samples into four classes. Pan et al. (2019) proposed an algorithm to integrate multi-level features based on FDG-PET images, and simultaneously considered the region features and connectivity between regions to classify AD or MCI from healthy people. Finally, multiple SVMs were used for voting classification, and good results had been achieved in multiple binary classification tasks.

Magnetic resonance imaging image features usually suffer from high dimensionality and small sample size, which may lead to overfitting in data-driven machine learning methods (Jiang et al., 2019). To solve this problem, most existing methods adopt feature selection or feature representation to exploit the potential knowledge of data. Sparse representation is one of the widely used feature representation methods. Sparse representation can explore potential relationships within the data (Gu et al., 2021). Chang et al. (2015) proposed a dictionary learning algorithm based on sparse decomposition of stacked prediction. They used the spatial pyramid matching method to encode representation coefficients, and used SVM to classify the pathological state of tumors. Shi et al. (2013) developed a multi-modal SRC algorithm for lung histopathological image classification, which used genetic algorithm to guide the learning of three sub-dictionaries of color, shape and texture, and then combined sparse reconstruction error and majority voting algorithm for classification of lung histopathology images. He (2019) proposed a spatial pyramid matching algorithm based on joint representation coefficient, which utilized the three color channel information of RGB, and converted the grayscale description operator into a color description operator, which improved the image classification performance. Jiang et al. (2019) extracted features from breast cancer histopathological images based on stacked sparse autoencoder, and used Softmax function to detect cell nuclei in histopathological images. Zhang et al. (2016) realized the fusion of global and local features of the nuclear image, and then combined the ranking and majority voting algorithm to classify the histopathological images of breast cancer. The above algorithms can effectively extract image features by introducing the sparsity of the image, and the extracted features have good reconstruction properties, but they do not have good discriminative ability.

To improve the diagnosis of MCI and AD based on MRI images, we propose large margin and local structure preservation sparse representation classifier (LMLS-SRC) in this manuscript. The traditional SRC only uses the classification error term to control the classification accuracy, and does not fully consider the class label information of the representation coefficients. Different from the traditional SRC, the LMLS-SRC algorithm introduces the classification margin term of representation coefficients into the sparse representation classifier, so that the similar representation coefficients are

compact in the representation space, and the dissimilar representation coefficients are separated as much as possible in the representation space. Experiments on the KAGGLE Alzheimer's dataset verify the advantages of our algorithm. Major contributions of this manuscript are highlighted below: (1) Considering the global information of the data by using the large margin term, the obtained dictionary is discriminative, and the representation coefficient has the small intra-class distance and large inter-class distance. (2) The local structure preservation term is introduced, which can inherit the manifold structure of the original data. (3) The  $\ell_{2,1}$ -norm term on the representation coefficients is used, which can enhance the sparsity and robustness of the representation coefficients.

## BACKGROUNDS

### Dictionary-Based Sparse Representation Classifier

Using SRC algorithm in image classification, how to design effective dictionary and representation coefficient for feature representation is the key factor to determine the algorithm performance (Wright et al., 2009). There are three aspects considered in the design of SRC algorithm: (1) The reconstruction error of the representation coefficients is small, so that the samples are as close to the original samples as possible in the sparse representation; (2) The representation coefficients are constrained to make the representation coefficients as sparse as possible; (3) The discrimination term should be considered to better extract more discriminative information of data (Jiang et al., 2013).

Let  $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_K] \in \mathbf{R}^{d \times N}$  be the  $K$ -classes training sample set,  $\mathbf{X}_k = [\mathbf{x}_1, \dots, \mathbf{x}_{N_k}]$  be the  $k$ -th class training sample subset,  $k = 1, 2, \dots, K, N = N_1 + N_2 + \dots + N_K$ .  $d$  is the dimensional of samples. The SRC algorithm for image classification can be represented as,

$$\min_{\mathbf{D}, \mathbf{A}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda g(\mathbf{A}) + \eta f(\mathbf{D}, \mathbf{A}, \mathbf{Y}), \quad (1)$$

where  $\mathbf{Y}$  is the class label matrix of  $\mathbf{X}$ .  $\mathbf{D} \in \mathbf{R}^{d \times m}$  is the learned dictionary, and  $\mathbf{A} \in \mathbf{R}^{m \times N}$  is the representation coefficient matrix of  $\mathbf{X}$ .  $m$  is the size of dictionary. In model training, the data reconstruction item  $\|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2$  is to ensure the representation ability of the dictionary  $\mathbf{D}$ , so that the reconstruction error of the training data is minimized, and the reconstructed image is as close to the original sample as possible. The regularization term is used to constrain the sparsity of the representation coefficients, which is usually represented as,

$$g(\mathbf{A}) = \|\mathbf{A}\|_p. \quad (2)$$

where  $\|\cdot\|_p$  is the regularization term of the representation coefficient  $\mathbf{A}$  ( $p < 2$ ), which makes the representation coefficient as sparse as possible.  $f(\mathbf{D}, \mathbf{A}, \mathbf{Y})$  is the discriminative function term of representation coefficient for classification to ensure the discriminative ability of  $\mathbf{D}$  and  $\mathbf{A}$ .

To obtain a discriminative dictionary, Yang et al. (2017) developed a supervised Fisher discrimination dictionary learning

(FDDL), which associated the elements in the dictionary with the class labels of the samples based on the Fisher discrimination criterion. Jiang et al. (2013) proposed the discriminative Label consistent K-SVD (LC-KSVD) algorithm. Zhang et al. (2019) proposed a robust flexible discriminative dictionary learning (RFDDL) algorithm based on subspace recovery and enhanced locality. This algorithm improved image representation and classification by enhancing representation coefficient robustness. The computational complexity of the SRC representation coefficient is usually high. To quickly obtain the representation coefficients, Ma et al. (2017) proposed the local sparse representation algorithm, which used the KNN criterion to select  $k$  samples adjacent to the current sample to build a dictionary matrix. In this way, the size of the dictionary is reduced and the process of representation coefficient is greatly accelerated. Similarly, inspired by the KNN criterion, Zheng and Ding (2020) developed a sparse KNN classifier based on group lasso strategy and KSVD algorithm. Wang et al. (2018) proposed a SRC algorithm based on the  $\ell_2$ -norm, which replaced the  $\ell_1$ -norm with the  $\ell_2$ -norm to constrain the coefficients. Ortiz and Becker (2014) proposed an approximate linear SRC algorithm. Authors used least square algorithm to select the training samples corresponding to the absolute values of the  $k$  largest coefficients to build a sub-dictionary.

### KAGGLE Alzheimer's Image Dataset

The experiments in this manuscript are carried out on the KAGGLE Alzheimer's image dataset (Loddo et al., 2022). The KAGGLE Alzheimer's dataset contains a total of four types of MRI images: non AD (3,200 images), very mild AD (2,240 images), mild AD (896 images) and moderate AD (64 images), with the resolution of  $176 \times 208$ . The KAGGLE Alzheimer's dataset does not provide detailed information on patient status. **Figure 1** shows some example images of the KAGGLE Alzheimer's dataset.

## THE PROPOSED ALGORITHM

### Objective Function

The purpose of sparse representation is to represent the sample with as few elements as possible on a given dictionary, so that a more concise representation of the sample can be obtained, and the useful information contained in the sample can be easily obtained. Thus the core problem of sparse representation is how to compute sparse coding coefficients on a given learned dictionary. Compared with the commonly used  $\ell_1$ -norm and  $\ell_2$ -norm,  $\ell_{2,1}$ -norm can improve the robustness of the model and reduce the computational complexity. Thus, we introduce  $\ell_{2,1}$ -norm constraint on representation coefficients in LMLS-SRC, i.e.,

$$\Pi_1 = \arg \min_{\mathbf{D}, \mathbf{A}} \{\|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda_1 \|\mathbf{A}\|_{2,1}^2\}, \quad (3)$$

where  $\lambda_1$  is a constant.

We define a large margin term on representation coefficient that relies on a specific neighborhood size for intra-class and inter-class representation coefficients. The large margin

term minimizes the intra-class distance of the representation coefficient and maximizes the inter-class distance of the representation coefficient, so as to improve the difference between the representation coefficients of different classes. The large margin term on representation coefficient can be written as,

$$f(\mathbf{a}_i) = \arg \min \left\{ \sum_{t \in C_k} \frac{\|\mathbf{a}_i - \mathbf{a}_t\|^2}{N_k} - \sum_{j \notin C_k} \frac{\|\mathbf{a}_i - \mathbf{a}_j\|^2}{N - N_k} \right\}, \quad (4)$$

where  $\sum_{t \in C_k} \frac{\|\mathbf{a}_i - \mathbf{a}_t\|^2}{N_k}$  represents the distance between  $\mathbf{a}_i$  and the sparse representation of the same class.  $\sum_{j \notin C_k} \frac{\|\mathbf{a}_i - \mathbf{a}_j\|^2}{N - N_k}$  represents the distance between  $\mathbf{a}_i$  and the sparse representation of the different class.  $C_k$  is the index set of the  $k$ -th class sample.

We build the intra-class similarity matrix  $Q^w$  and inter-class similarity matrix  $Q^b$  based on representation coefficient. The elements of the matrix  $Q^w$  and matrix  $Q^b$  are expressed as,

$$q_{ij}^w = \begin{cases} 1/N_k, & \text{if } i, j \in C_k \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$$q_{ij}^b = \begin{cases} 1/(N - N_k), & \text{if } i \in C_k, j \notin C_k \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Then the large margin term on representation coefficient can be expressed as,

$$\begin{aligned} \Pi_2 &= \arg \min \frac{1}{N} \sum_{i=1}^N f(\mathbf{a}_i) \\ &= \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N (q_{ij}^w \|\mathbf{a}_i - \mathbf{a}_j\|^2 - q_{ij}^b \|\mathbf{a}_i - \mathbf{a}_j\|^2) \\ &= \frac{1}{N} \left( 2 \sum_{i=1}^N \mathbf{a}_i^2 - 2 \sum_{i=1}^N \sum_{j=1}^N \mathbf{a}_i q_{ij}^w \mathbf{a}_j \right) - \\ &\quad \frac{1}{N} \left( \sum_{i=1}^N \mathbf{a}_i^2 + \sum_{j=1}^N \mathbf{a}_j^2 q_{ij}^b - 2 \sum_{i=1}^N \sum_{j=1}^N \mathbf{a}_i q_{ij}^b \mathbf{a}_j \right) \\ &= \text{tr} \left( \frac{1}{N} \mathbf{A}^T (2\mathbf{I} - 2\mathbf{Q}^w) \mathbf{A} \right) - \text{tr} \left( \frac{1}{N} \mathbf{A}^T (\mathbf{I} + \tilde{\mathbf{Q}}^b - 2\mathbf{Q}^b) \mathbf{A} \right) \\ &= \text{tr} \left( \frac{1}{N} \mathbf{A}^T (\mathbf{I} - 2\mathbf{Q}^w - \tilde{\mathbf{Q}}^b + 2\mathbf{Q}^b) \mathbf{A} \right) \\ &= \text{tr}(\mathbf{A}^T \mathbf{S} \mathbf{A}) \end{aligned} \quad (7)$$

where  $\mathbf{S} = \frac{1}{N} (\mathbf{I} - 2\mathbf{Q}^w - \tilde{\mathbf{Q}}^b + 2\mathbf{Q}^b)$ . The matrix  $\tilde{\mathbf{Q}}^b$  is the diagonal matrix with the element being the column-sum of  $\mathbf{Q}^b$ .

Following the principle of local structure preservation, if two images are close in the original space, they should also have similar representation coefficients. To this end, we construct a similarity matrix  $\mathbf{P}$  that reflects the intrinsic local structure between images. The element of matrix  $\mathbf{P}$  is defined as,

$$p_{ij} = \begin{cases} \exp \left( -\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2} \right), & \text{if } \mathbf{x}_i \in N(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N(\mathbf{x}_i), \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

where  $N(\mathbf{x}_j)$  represents the  $k$  nearest neighbors of  $\mathbf{x}_j$ .

The local structure preservation term on representation coefficient is expressed as,

$$\begin{aligned} \Pi_3 &= \arg \min \sum_{i,j}^N p_{ij} \|\mathbf{a}_i - \mathbf{a}_j\|_2^2 \\ &= \text{tr}(\mathbf{A}^T (\mathbf{P} - \tilde{\mathbf{P}}) \mathbf{A}) \\ &= \text{tr}(\mathbf{A}^T \mathbf{L} \mathbf{A}), \end{aligned} \quad (9)$$

where the graph Laplacian matrix  $\mathbf{L}$  is  $\mathbf{L} = \mathbf{P} - \tilde{\mathbf{P}}$ ,  $\tilde{\mathbf{P}}$  is the diagonal matrix with the element being the row-sum of  $\mathbf{P}$ .

The LMLS-SRC algorithm is a supervised learning model. Using the class labels of all training samples, we use a linear classifier  $\mathbf{W}$  for representation coefficient  $\mathbf{A}$  and dictionary  $\mathbf{D}$ , i.e.,

$$\Pi_4 = \arg \min_{\mathbf{W}, \mathbf{A}} \|\mathbf{W}\mathbf{A} - \mathbf{Y}\|_F^2 + \lambda_5 \|\mathbf{W}\|_F^2. \quad (10)$$

In summary, the objective function of the LMLS-SRC algorithm can be written as,

$$F(\mathbf{D}, \mathbf{A}, \mathbf{W}) = \min_{\mathbf{D}, \mathbf{A}, \mathbf{W}} \Pi_1 + \Pi_2 + \Pi_3 + \Pi_4, \quad (11)$$

i.e.,

$$\begin{aligned} \min_{\mathbf{D}, \mathbf{A}, \mathbf{W}} & \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda_1 \|\mathbf{A}\|_{2,1}^2 + \lambda_2 \text{tr}(\mathbf{A}^T \mathbf{S} \mathbf{A}) \\ & + \lambda_3 \text{tr}(\mathbf{A}^T \mathbf{L} \mathbf{A}) + \lambda_4 \|\mathbf{W}\mathbf{A} - \mathbf{Y}\|_F^2 + \lambda_5 \|\mathbf{W}\|_F^2, \\ \text{s.t.} & \|\mathbf{d}_i\|_2^2 \leq 1, \forall i \end{aligned} \quad (12)$$

where  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ , and  $\lambda_5$  are trade-off parameters.

By alternately optimizing the representation coefficient  $\mathbf{A}$ , dictionary  $\mathbf{D}$  and classifier parameter  $\mathbf{W}$ , the following performance can be obtained as: (1) the dictionary  $\mathbf{D}$  has more sparse representation performance, which enhances the reconstruction of the sample by the dictionary. (2) LMLS-SRC maximizes the distance between different classes of representation coefficients and greatly reduces the similarity between different classes of representation coefficients. (3) The representation coefficient is more discriminative, which is beneficial to the performance of image classification.

## Optimization

(1) Fix  $\mathbf{D}$ ,  $\mathbf{W}$ , and update  $\mathbf{A}$ . Eq. (12) can be written by,

$$\begin{aligned} \min F(\mathbf{A}) &= \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda_1 \|\mathbf{A}\|_{2,1}^2 + \lambda_2 \text{tr}(\mathbf{A}^T \mathbf{S} \mathbf{A}) \\ &+ \lambda_3 \text{tr}(\mathbf{A}^T \mathbf{L} \mathbf{A}) + \lambda_4 \|\mathbf{W}\mathbf{A} - \mathbf{Y}\|_F^2. \end{aligned} \quad (13)$$

According to the definition of  $\ell_{2,1}$ -norm,  $\|\mathbf{A}\|_{2,1}^2 = \text{tr}(\mathbf{A}^T \mathbf{\Omega} \mathbf{A})$ .  $\mathbf{\Omega}$  is a diagonal matrix whose elements are setting by  $\Omega_{ii} = 1/(2\|\mathbf{A}_i\|_2)$  where  $\mathbf{A}_i$  represents the  $i$ -th row of  $\mathbf{A}$ .

Equation (12) can be re-written by,

$$\begin{aligned} \min F(\mathbf{A}) &= \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda_1 \text{tr}(\mathbf{A}^T \mathbf{\Omega} \mathbf{A}) + \lambda_2 \text{tr}(\mathbf{A}^T \mathbf{S} \mathbf{A}) \\ &+ \lambda_3 \text{tr}(\mathbf{A}^T \mathbf{L} \mathbf{A}) + \lambda_4 \|\mathbf{W}\mathbf{A} - \mathbf{Y}\|_F^2. \end{aligned} \quad (14)$$

Setting  $\partial F(\mathbf{A}) / \partial \mathbf{A} = 0$ , we can obtain,

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{A}} &= 2\mathbf{D}^T \mathbf{D} \mathbf{A} - 2\mathbf{D}^T \mathbf{X} + (2\lambda_1 \mathbf{\Omega} + 2\lambda_2 \mathbf{S} + 2\lambda_3 \mathbf{L}) \mathbf{A} \\ &+ 2\lambda_4 (\mathbf{W}^T \mathbf{W} \mathbf{A} - \mathbf{W}^T \mathbf{Y}). \end{aligned} \quad (15)$$

$\mathbf{A}$  can be obtained by the updated by,

$$\begin{aligned} \mathbf{A}^* &= (\mathbf{D}^T \mathbf{D} + \lambda_1 \mathbf{\Omega} + \lambda_4 \mathbf{W}^T \mathbf{W} + \lambda_2 \mathbf{S} + \lambda_3 \mathbf{L})^{-1} \\ &(\lambda_4 \mathbf{W}^T \mathbf{Y} + \mathbf{D}^T \mathbf{X}). \end{aligned} \quad (16)$$

(2) Fix  $\mathbf{A}$ ,  $\mathbf{W}$ , and update  $\mathbf{D}$ . Equation (12) can be written by,

$$\begin{aligned} \min F(\mathbf{D}) &= \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2, \\ \text{s.t.} & \|\mathbf{d}_i\|_2^2 \leq 1, \forall i \end{aligned} \quad (17)$$

We can solve Eq. (17) by the following Lagrangian dual function,

$$\min F(\mathbf{D}, \boldsymbol{\sigma}) = \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \sum_{i=1}^m \gamma_i (\|d_i\|_2^2 - 1), \quad (18)$$

where  $\gamma_i$  is the Lagrange multiplier of  $i$ -th atoms.

We build a diagonal matrix  $\boldsymbol{\Theta}$  with the element  $\Theta_{ii} = \gamma_i$ . Equation (18) can be written by,

$$\min F(\mathbf{D}, \boldsymbol{\Theta}) = \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \text{tr}(\mathbf{D}^T \mathbf{D} \boldsymbol{\Theta}) - \text{tr}(\boldsymbol{\Theta}). \quad (19)$$

Setting  $\partial F(\mathbf{D}, \boldsymbol{\Theta}) / \partial \mathbf{D} = 0$ , we can obtain,

$$\mathbf{D}^* = \mathbf{X}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \boldsymbol{\Theta})^{-1}. \quad (20)$$

(3) Fix  $\mathbf{A}$  and  $\mathbf{D}$ , and update  $\mathbf{W}$ . Equation (12) can be written by,

$$\min F(\mathbf{W}) = \lambda_4 \|\mathbf{W}\mathbf{A} - \mathbf{Y}\|_F^2 + \lambda_5 \|\mathbf{W}\|_F^2. \quad (21)$$

Setting  $\partial F(\mathbf{W}) / \partial \mathbf{W} = 0$ , we can obtain,

$$\mathbf{W}^* = \lambda_4 \mathbf{Y}\mathbf{A}^T (\lambda_4 \mathbf{A}\mathbf{A}^T + \lambda_5 \mathbf{I})^{-1}. \quad (22)$$

**The optimization steps of LMLS-SRC algorithm are shown in Algorithm 1.**

Input: training set  $\mathbf{X}$  and its label matrix  $\mathbf{Y}$ , tolerance error  $\delta$ , maximum number of iterations  $maxiter$ , parameters  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ , and  $\lambda_5$ ,

Output: parameters  $\mathbf{D}$ ,  $\mathbf{A}$ , and  $\mathbf{W}$ .

Initialize: initialize  $\mathbf{D}$  and  $\mathbf{A}$  using the LC-KSVD algorithm,  $\mathbf{W} = \mathbf{I}$ ,  $m = 1$ ,

Calculate matrices  $Q^w$ ,  $Q^b$ , and  $\mathbf{P}$ ;

While not converged and  $m \leq maxiterdo$

    Calculate  $\mathbf{D}(m)$  by Eq. (20);

    Calculate  $\mathbf{A}(m)$  by Eq. (16);

    Calculate  $\mathbf{W}(m)$  by Eq. (22);

    Check the convergence condition  $\frac{|F(\mathbf{D}(m), \mathbf{A}(m), \mathbf{W}(m)) - F(\mathbf{D}(m-1), \mathbf{A}(m-1), \mathbf{W}(m-1))|}{F(\mathbf{D}(m-1), \mathbf{A}(m-1), \mathbf{W}(m-1))} < \delta$

$m = m + 1$

end while

## EXPERIMENTS

### Experimental Settings

In clinical diagnosis, AD classification tasks consist of two categories. The first is the AD binary classification task, which extracts features based on MRI images and uses machine learning models to classify normal individuals and AD patients, which can help doctors diagnose AD patients. The second is the classification of various ADs, especially the diagnosis and identification of mild AD and very mild AD. Early prediction of AD can help to take treatment and intervention measures in the early stage of AD. Therefore, in this manuscript, we design binary, three-class and four-class classification tasks on the KAGGLE Alzheimer's dataset.

Volume analysis is the commonly used feature extraction method in AD classification. Volumetric feature extraction is

divided into two categories: density maps and predefined area methods. AD MRI image is mainly related to the volume of the density map structure, cortical structure, subcortical structure and other regions. In this manuscript, we use FSL (FMRIB software library) toolbox to extract MRI features (Jenkinson et al., 2012). FSL is a library of comprehensive analysis tools for brain imaging data such as MRI, developed by the FMRIB Centre in Oxford. We use the FSL toolbox to calculate the volume, area and thickness characteristics of various brain tissues in brain MRI images. In the comparison experiment, the LMLS-SRC algorithm is compared with SRC (Wright et al., 2009), logistic regression (LR) (Tsangaratos and Ilia, 2016), linear discriminant (LD) (Kim et al., 2011), LC-KSVD, FDDL, and sparse representation-based discriminative metric learning (SRDML) (Zhou et al., 2022). The radial basis function (RBF) kernel is used in LR. The default settings are used to produce test results from these classifiers using the MATLAB classification learner toolbox. The RBF kernel and the regularization parameters for all comparison algorithms range from  $10^{-3}$  to  $10^3$ . The number of dictionary atoms in SRC and dictionary learning is set as the number of training samples. Indicators of classification performance include classification accuracy, sensitivity, specificity, precision, F1-score, and G-mean. We carry out 5-fold cross-validation strategy and record the experimental results.

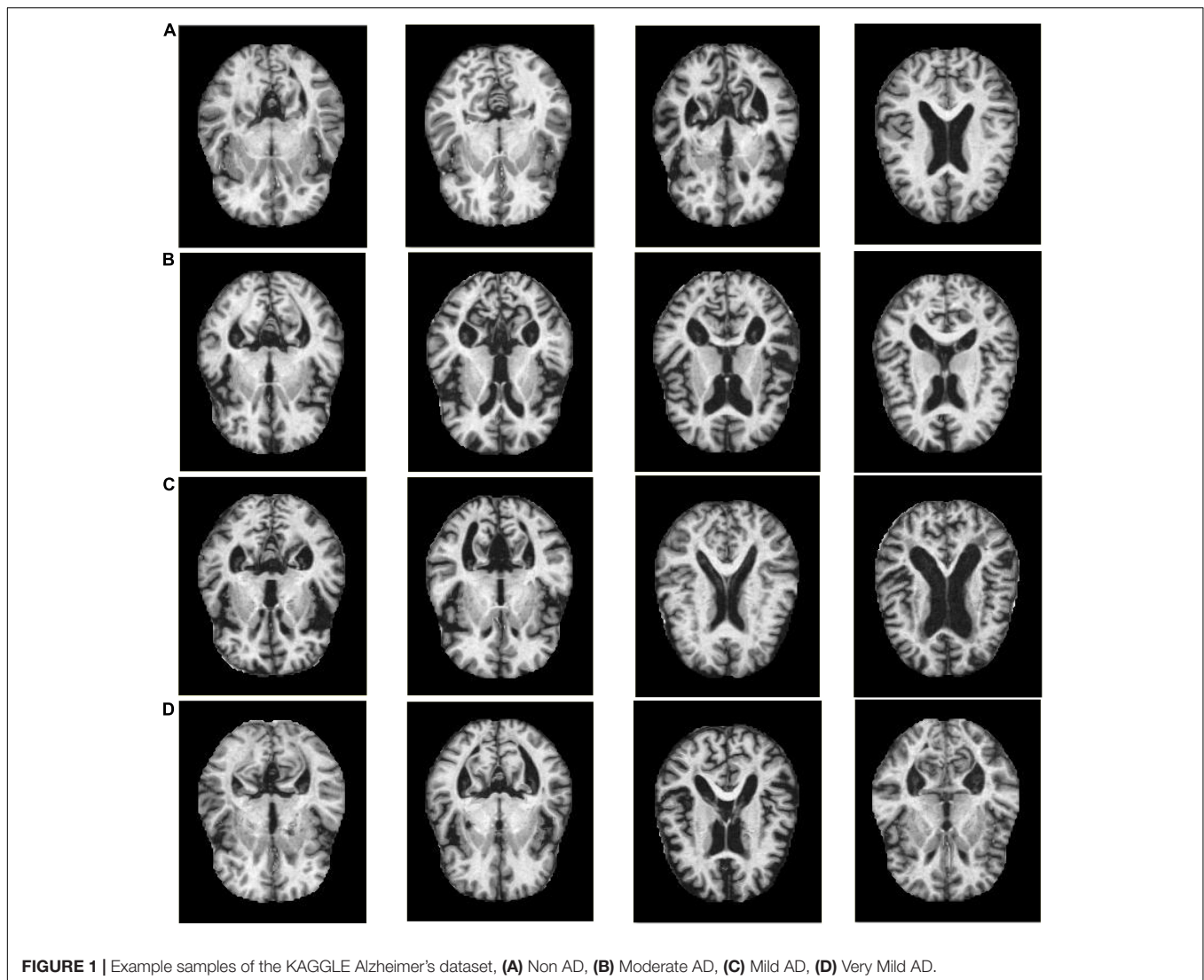
### Experimental Results

(1) Binary classification task. The main goal of this work is to classify brain MRI into AD and non AD classes. We utilized 3,200 and 62 MRI images for non AD and AD classes, respectively. We randomly selected 1,000 MRI images from the non AD class images to increase the moderate AD class dataset to 620 MRI images using data augmentation techniques. The comparative training and test results in binary classification task are shown in **Tables 1, 2**, respectively.

(2) Three-class classification tasks. The main goal of this work was to classify brain MRI into three classes: non AD, mild AD, and moderate AD. Using data augmentation techniques, these three classes of datasets contain 3,200, 700, and 620 images, respectively. We randomly selected 1,000 MRI images from the non AD class. The comparative training and test results in three-class classification task are shown in **Tables 3, 4**, respectively.

(3) Four-class classification tasks. The main goal of this work is to classify brain MRI images into four classes: very mild AD, non AD, mild AD, and moderate AD. Similar to the three-class classification task described, we randomly selected 1,000 MRI images each from non AD class images and very mild AD, respectively, and used data augmentation to increase the moderate dementia dataset to 520 MRI images. The number of images in the four categories of very mild AD, non AD, mild AD, and moderate AD are 1,000, 1,000, 700, and 520, respectively. The comparative training and test results in four-class classification task are shown in **Tables 5, 6**, respectively.

We can see that all the comparison algorithms have the highest classification accuracy in the binary classification task (AD and non AD). It shows that these machine learning algorithms have excellent performance in the classification and diagnosis of AD. It is more practical to classify patients, very mild AD, non



**FIGURE 1** | Example samples of the KAGGLE Alzheimer's dataset. **(A)** Non AD, **(B)** Moderate AD, **(C)** Mild AD, **(D)** Very Mild AD.

AD, mild AD, and moderate AD into four classes, and this classification task is more difficult. The classification accuracy of all the comparison algorithms on the four-class task is slightly lower than that on the two-class task. However, the LMLS-SRC algorithm achieves the best results in these tables, indicating that our algorithm has a great improvement in the diagnosis of AD.

In **Tables 2, 4, 6**, the LMLS-SRC algorithm improves the classification accuracy of the second best algorithm by 2.84, 3.00, and 2.41%, respectively. This shows that the dictionary learned in this study has better reconstruction performance for the samples of same class and better discriminative performance for samples of different classes. KSVD, LC-KSVD, and LMLS-SRC are SRC algorithms. The KSVD and LC-KSVD algorithms only constrain the discriminative ability of the representation coefficients, and do not take into account the large margin between the representation coefficients of different classes. Therefore, the discriminative ability of the learned dictionary obtained by KSVD and LC-KSVD is still weak. The dictionary learned by the LMLS-SRC algorithm in this manuscript is combined with the

classification large margin criterion, which directly constrains the intra-class distance and inter-class distance of the representation coefficients. Compared with the other three algorithms, the inter-class differences of the dictionary learned by our algorithm are more discriminative.

### Parameter Analysis

(1) Convergence analysis. The update of  $\{(\mathbf{D}), (\mathbf{A}), (\mathbf{W})\}$  in the objective function are three convex optimization problems. That is, when other parameters are fixed, the iterative solution of dictionary  $\mathbf{D}$ , representation coefficient  $\mathbf{A}$  and classifier parameter  $\mathbf{W}$  is the convex problem. The solution of dictionary  $\mathbf{D}$  is obtained by Eq. (20). The solution of dictionary  $\mathbf{A}$  is obtained by Eq. (16). The solution of dictionary  $\mathbf{W}$  is obtained by Eq. (22). **Figure 2** shows the convergence of the LMLS-SRC algorithm. As shown in **Figure 2**, it can be seen that the classification accuracy of the LMLS-SRC algorithm tends to be parallel to the X-axis from the 10th iteration. Here, it can be considered that our algorithm converges after 12 iterations.

**TABLE 1 |** The comparative training results (with standard deviation) in binary classification task.

Algorithms	Accuracy	Sensitivity	Specificity	Precision	F1-score	G-mean
LD	81.30 (2.84)	81.99 (3.15)	80.06 (2.80)	80.58 (3.32)	82.26 (3.27)	81.02 (2.97)
LR	82.15 (2.55)	82.62 (2.66)	81.79 (2.70)	82.68 (2.35)	82.51 (2.56)	82.20 (2.60)
SRC	82.10 (2.35)	78.97 (2.01)	77.33 (2.64)	77.63 (1.62)	77.55 (1.43)	78.15 (2.28)
LC-KSVD	80.27 (2.54)	81.34 (2.12)	78.94 (2.63)	80.85 (1.82)	79.93 (2.07)	80.13 (1.59)
FDDL	83.16 (2.64)	84.47 <b>(??)</b>	81.38 (1.83)	85.20 (1.45)	82.86 (1.69)	82.91 (1.54)
SRDML	85.71 (2.15)	85.91 (2.23)	85.09 (1.75)	84.10 (1.88)	85.08 (1.74)	85.50 (1.96)
LMLS-SRC	<b>89.80</b> (2.02)	<b>90.39</b> (1.35)	<b>87.87</b> (2.06)	<b>88.89</b> (1.35)	<b>90.43</b> (1.28)	<b>89.12</b> (1.19)

The bold values in Tables 1–6 are the best experiment results.

**TABLE 2 |** The comparative test results (with standard deviation) in binary classification task.

Algorithms	Accuracy	Sensitivity	Specificity	Precision	F1-score	G-mean
LD	80.92 (2.26)	81.64 (1.69)	80.44 (2.10)	81.45 (2.06)	80.70 (1.62)	81.04 (1.37)
LR	81.61 (1.71)	82.28 (2.58)	80.96 (2.70)	82.81 (1.04)	80.79 (1.88)	81.62 (2.64)
SRC	82.91 (1.75)	83.18 (2.46)	82.86 (2.28)	83.07 (1.16)	82.94 (1.87)	83.02 (2.37)
LC-KSVD	82.15 (2.74)	82.59 (1.38)	80.51 (2.80)	82.78 (2.55)	82.56 (1.96)	81.54 (1.93)
FDDL	82.89 (2.23)	84.26 (1.50)	81.71 (1.43)	84.35 (1.14)	83.23 (2.02)	82.98 (1.46)
SRDML	85.44 (2.14)	87.13 (2.20)	84.35 (2.10)	86.42 (2.74)	85.42 (2.05)	85.73 (2.15)
LMLS-SRC	<b>88.28</b> (2.07)	<b>90.15</b> (2.06)	<b>86.75</b> (1.67)	<b>90.08</b> (1.92)	<b>88.31</b> (1.18)	<b>88.43</b> (1.68)

**TABLE 3 |** The comparative training results (with standard deviation) in three-class classification task.

Algorithms	Accuracy	Sensitivity	Specificity	Precision	F1-score	G-mean
LD	80.13 (2.72)	80.70 (1.92)	80.52 (2.24)	79.57 (2.28)	80.94 (2.34)	80.61 (2.36)
LR	81.31 (2.55)	82.55 (2.30)	80.25 (2.03)	81.17 (2.62)	81.82 (2.19)	81.39 (2.16)
SRC	81.94 (2.20)	82.20 (2.49)	80.46 (2.59)	81.09 (2.10)	81.17 (2.21)	81.33 (2.54)
LC-KSVD	83.80 (1.76)	85.54 (1.68)	81.64 (2.98)	83.94 (2.23)	83.32 (1.80)	83.57 (2.24)
FDDL	84.04 (2.30)	86.12 (2.61)	81.13 (2.33)	83.93 (2.24)	84.32 (2.36)	83.59 (2.47)
SRDML	85.39 (2.33)	86.82 (2.00)	84.88 (2.37)	86.32 (2.05)	86.86 (2.33)	85.85 (2.02)
LMLS-SRC	<b>89.32</b> (1.84)	<b>91.38</b> (1.20)	<b>86.81</b> (2.81)	<b>88.86</b> (2.12)	<b>89.00</b> (1.53)	<b>89.07</b> (1.83)

**TABLE 4 |** The comparative test results (with standard deviation) in three-class classification task.

Algorithms	Accuracy	Sensitivity	Specificity	Precision	F1-score	G-mean
LD	78.47 (2.16)	79.40 (1.99)	77.83 (2.50)	78.94 (2.29)	78.76 (1.60)	78.61 (2.23)
LR	79.43 (2.02)	80.38 (2.56)	78.75 (2.19)	79.50 (1.95)	78.99 (2.18)	79.56 (1.75)
SRC	80.23 (1.79)	80.22 (2.53)	79.26 (2.30)	79.31 (2.54)	79.47 (1.39)	79.74 (2.31)
LC-KSVD	81.72 (1.31)	82.22 (2.34)	80.59 (2.41)	81.19 (2.22)	81.05 (1.35)	81.40 (2.40)
FDDL	82.26 (2.20)	83.12 (2.37)	80.87 (1.42)	82.53 (2.56)	82.39 (2.44)	81.98 (1.84)
SRDML	84.90 (2.27)	85.66 (2.49)	83.86 (1.80)	85.27 (2.83)	85.11 (2.13)	84.76 (2.12)
LMLS-SRC	<b>87.90</b> (1.81)	<b>89.25</b> (2.02)	<b>86.53</b> (2.04)	<b>88.71</b> (1.74)	<b>88.44</b> (1.81)	<b>87.88</b> (2.27)

**TABLE 5 |** The comparative training results (with standard deviation) in four-class classification task.

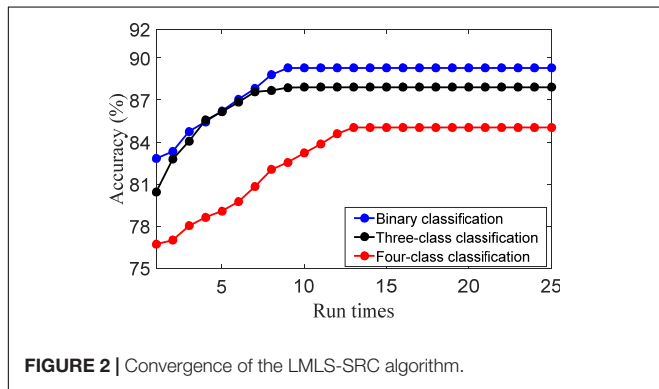
Algorithms	Accuracy	Sensitivity	Specificity	Precision	F1-score	G-mean
LR	79.70 (1.47)	80.06 (2.15)	78.70 (2.20)	81.49 (2.73)	79.23 (1.48)	79.37 (2.10)
LR	80.81 (1.88)	81.71 (1.47)	79.41 (2.09)	80.87 (2.11)	80.40 (1.22)	80.55 (1.27)
SRC	80.86 (2.02)	82.38 (2.29)	79.92 (1.84)	78.97 (1.37)	80.41 (1.62)	81.14 (2.05)
LC-KSVD	82.61 (2.16)	84.10 (1.58)	80.92 (1.55)	82.36 (2.02)	83.52 (2.32)	82.50 (1.59)
FDDL	83.85 (1.56)	84.56 (2.80)	82.70 (2.29)	83.46 (3.09)	84.09 (2.07)	83.63 (2.53)
SRDML	85.91 (2.05)	86.46 (2.63)	83.28 (2.34)	83.39 (2.16)	84.97 (1.55)	84.85 (2.48)
LMLS-SRC	<b>86.58</b> (1.59)	<b>87.64</b> (1.13)	<b>85.93</b> (2.45)	<b>86.93</b> (2.00)	<b>86.21</b> (1.49)	<b>86.78</b> (1.66)

(2) Training set size. The size of the training set usually directly determines the performance of machine learning algorithms. **Figure 3** shows the classification accuracy of the LMLS-SRC algorithm on binary-class, three-class and four-class classification tasks under different training sets of each subclass. The X-axis represents the training sample size  $N$  of each subclass,  $N = [50, 100, \dots, 400]$ . From **Figure 3**, we can see that the accuracy of LMLS-SRC increases with the increase of training samples. When the training sample size of each subset reaches 200, the performance of the LMLS-SRC algorithm is basically stable, indicating that the LMLS-SRC algorithm can achieve better performance without too many training samples.

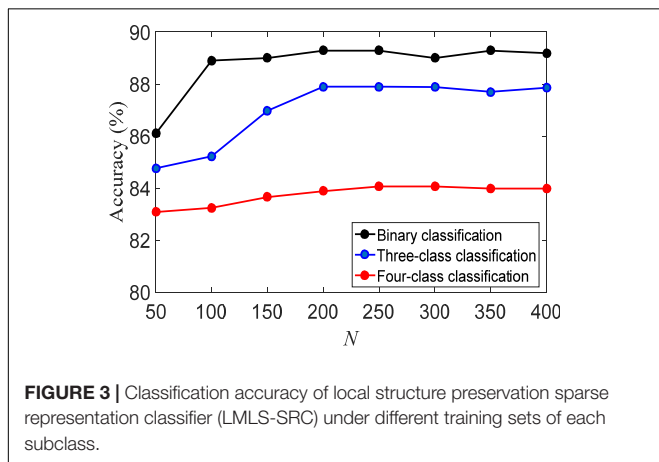
(3) Regularization parameters. The LMLS-SRC algorithm has five regularization parameters  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ , and  $\lambda_5$ , and the regularization parameters are all obtained in  $[1.0E-3, \dots, 1.0E+3]$ .  $\lambda_2$  controls the role of the large margin term.  $\lambda_3$  controls the role of the local structure preservation term.  $\lambda_4$  controls the role of the linear classifier. **Figure 4** shows the classification accuracy of the LMLS-SRC algorithm in the binary, three-class and four-class

**TABLE 6 |** The comparative test results (with standard deviation) in four-class classification task.

Algorithms	Accuracy	Sensitivity	Specificity	Precision	F1-score	G-mean
LD	77.67 (2.22)	79.69 (1.51)	77.51 (2.15)	78.40 (2.52)	77.50 (2.08)	78.59 (1.80)
LR	78.56 (1.89)	79.60 (2.51)	78.47 (1.60)	78.59 (2.74)	78.23 (1.43)	79.03 (2.01)
SRC	79.40 (2.13)	79.77 (2.33)	79.06 (2.68)	80.25 (1.44)	79.15 (2.32)	79.41 (2.50)
LC-KSVD	81.25 (2.40)	81.77 (2.19)	81.34 (1.59)	80.87 (2.36)	81.55 (2.08)	81.55 (1.86)
FDDL	81.45 (1.33)	81.06 (2.00)	80.02 (2.09)	80.67 (2.73)	80.69 (1.25)	80.54 (2.05)
SRDML	83.13 (2.06)	82.10 (2.26)	82.94 (2.04)	83.56 (1.49)	83.17 (1.99)	82.52 (2.15)
LMLS-SRC	<b>85.54</b> (1.59)	<b>86.19</b> (2.03)	<b>84.51</b> (2.12)	<b>86.15</b> (1.63)	<b>85.97</b> (1.06)	<b>85.34</b> (2.07)

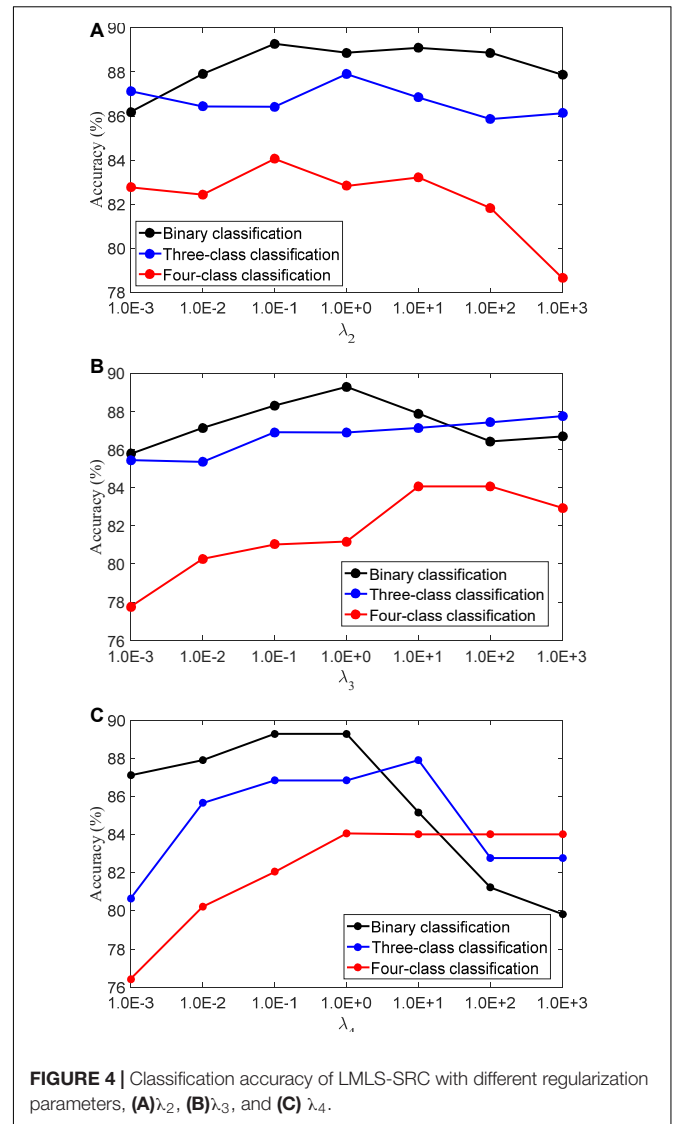


**FIGURE 2 |** Convergence of the LMLS-SRC algorithm.



**FIGURE 3 |** Classification accuracy of local structure preservation sparse representation classifier (LMLS-SRC) under different training sets of each subclass.

tasks with different  $\lambda_2$ ,  $\lambda_3$ , and  $\lambda_4$ , respectively. **Figure 4** shows that the performance of the LMLS-SRC algorithm varies greatly with different  $\lambda_2$ ,  $\lambda_3$ , and  $\lambda_4$ , while fixing the other parameters. Therefore, it is reasonable to use a grid search strategy to optimize the regularization parameters.



**FIGURE 4 |** Classification accuracy of LMLS-SRC with different regularization parameters, (A)  $\lambda_2$ , (B)  $\lambda_3$ , and (C)  $\lambda_4$ .

## CONCLUSION

With the acceleration of the global aging trend, one of the problems brought about is the rapid increase in the number of AD patients. The pathogenesis and effective treatment of AD are still unclear at present. Early detection, classification, and prediction of AD, and targeted care and treatment of patients on this basis can delay the progression of AD. Machine learning algorithms that can automatically extract information and complete inference have good application prospects in AD classification and prediction. Therefore, this manuscript conducts research based on the application of SRC algorithm in AD classification. The research content mainly includes two aspects: model construction and model performance evaluation. The proposed LMLS-SRC algorithm introduces the large margin term and local constraint term in the traditional SRC model, and obtains the dictionary and representation coefficients with discriminative ability while maintaining the data manifold



structure. The effectiveness of the LMLS-SRC algorithm is validated on the KAGGLE Alzheimer's dataset.

Although the LMLS-SRC algorithm shows the advantages compared with some excellent algorithms, there are still some problems to be solved. In the future, we will mainly focus on the following aspects: (1) The LMLS-SRC algorithm belongs to the shallow model. How to design the deep model of the sparse representation algorithm needs to be further studied. (2) In this manuscript, brain MRI images are used as the basic data to study the application of AD classification. Multimodal data can provide richer information, and how to extract AD-related features from multimodal data can be studied in the future. (3) This manuscript uses the volume features extracted by using FSL tool. Extracting various features for AD classification can be done in the next future. (4) In practical applications, image classification often encounters small samples or even a single training sample, and traditional SRC algorithms cannot effectively handle such situations. How to deal with the single training sample is the work to be further studied in the future.

## REFERENCES

- Al-Khuzaie, F., and Duru, A. D. (2021). Diagnosis of Alzheimer disease using 2D MRI slices by convolutional neural network. *Appl. Bion. Biomech.* 2:6690539. doi: 10.1155/2021/6690539
- Altaf, T., Anwar, S. M., Gul, N., Majeed, M. N., and Majid, M. (2018). Multi-class Alzheimer's disease classification using image and clinical features. *Biomed. Signal Process. Control* 43, 64–74. doi: 10.1016/j.bspc.2018.02.019
- Chang, H., Zhou, Y., Borowsky, A., Barner, K., Spellman, P., and Parvin, B. (2015). Stacked predictive sparse decomposition for classification of histology sections. *Int. J. Comput. Vis.* 113, 3–18. doi: 10.1007/s11263-014-0790-9
- Gao, S. S. (2021). Gray level co-occurrence matrix and extreme learning machine for Alzheimer's disease diagnosis. *Int. J. Cogn. Comput. Eng.* 2, 116–129. doi: 10.1016/j.ijcce.2021.08.002
- Gu, X., Zhang, C., and Ni, T. G. (2021). A hierarchical discriminative sparse representation classifier for EEG signal detection. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18, 1679–1687. doi: 10.1109/TCBB.2020.3006699
- He, T. (2019). Image classification based on sparse coding multi-scale spatial latent semantic analysis. *EURASIP J. Image Video Process.* 2019:38. doi: 10.1186/s13640-019-0425-8
- Hett, K., Ta, V. T., Manjón, J. V., and Coupé, P. (2018). Adaptive fusion of texture-based grading for Alzheimer's disease classification. *Comput. Med. Imag. Graph.* 70, 8–16. doi: 10.1016/j.compmedimag.2018.08.002
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015
- Jiang, Y., Chen, L., Zhang, H., and Xiao, X. (2019). Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module. *PLoS One* 14:e0214587. doi: 10.1371/journal.pone.0214587
- Jiang, Z., Lin, Z., and Davis, L. S. (2013). Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 2651–2664. doi: 10.1109/TPAMI.2013.88
- Katabathula, S., Wang, Q., and Xu, R. (2021). Predict Alzheimer's disease using hippocampus MRI data: a lightweight 3D deep convolutional network model with visual and global shape representations. *Alzheimers Res. Ther.* 13:104. doi: 10.1186/s13195-021-00837-0
- Kim, K. S., Choi, H. H., Moon, C. S., and Mun, C. W. (2011). Comparison of *k*-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of Electromyogram signals based on the wrist-motion directions. *Curr. Appl. Phys.* 11, 740–745. doi: 10.1016/j.cap.2010.11.051

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images>.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

RL conceived and developed the theoretical framework of the manuscript. All authors carried out experiment and data process, drafted the manuscript, and approved the submitted version.

- Kruthika, K. R., Maheshappa, H. D., Rajeswari, and Alzheimer's Disease Neuroimaging Initiative (2019). Multistage classifier-based approach for Alzheimer's disease prediction and retrieval. *Inform. Med. Unlocked* 14, 34–42. doi: 10.1016/j.imu.2018.12.003
- Lee, S., Lee, H., and Kim, K. W. (2020). Magnetic resonance imaging texture predicts progression to dementia due to Alzheimer disease earlier than hippocampal volume. *J. Psychiatry Neurosci.* 45, 7–14. doi: 10.1503/jpn.180171
- Liu, M., Zhang, D., Adeli, E., and Shen, D. G. (2015). Inherent structure-based multiview learning with multitemplate feature representation for Alzheimer's disease diagnosis. *IEEE Trans. Biomed. Eng.* 63, 1473–1482. doi: 10.1109/TBME.2015.2496233
- Loddo, A., Butta, S., and Ruberto, C. D. (2022). Deep learning based pipelines for Alzheimer's disease diagnosis: a comparative study and a novel deep-ensemble method. *Comput. Biol. Med.* 141:105032. doi: 10.1016/j.combiomed.2021.105032
- Ma, H., Gou, J., Wang, X., Ke, J., and Zeng, S. (2017). Sparse coefficient-based *k*-nearest neighbor classification. *IEEE Access* 5, 16618–16634. doi: 10.1109/ACCESS.2017.2739807
- Ortiz, E. G., and Becker, B. C. (2014). Face recognition for web-scale datasets. *Comput. Vis. Image Understand.* 118, 153–170. doi: 10.1016/j.cviu.2013.09.004
- Pan, X., Adel, M., Fossati, C., Gaidon, T., and Guedj, E. (2019). Multilevel Feature Representation of FDG-PET Brain Images for Diagnosing Alzheimer's Disease. *IEEE J. Biomed. Health Inform.* 23, 1499–1506. doi: 10.1109/JBHI.2018.2857217
- Porsteinsson, A. P., Isaacson, R. S., Knox, S., Sabbagh, M. N., and Rubino, I. (2021). Diagnosis of Early Alzheimer's Disease: clinical Practice in 2021. *J. Prevent. Alzheimers Dis.* 8, 371–386. doi: 10.14283/jpad.2021.23
- Shi, Y. H., Gao, Y., Yang, Y. B., Zhang, Y., and Wang, D. (2013). Multimodal sparse representation-based classification for lung needle biopsy images. *IEEE Trans. Biomed. Eng.* 60, 2675–2685. doi: 10.1109/TBME.2013.2262099
- Tsangaratos, P., and Ilia, I. (2016). Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: the influence of models complexity and training dataset size. *Catena* 145, 164–179. doi: 10.1016/j.catena.2016.06.004
- van Oostveen, W. M., and de Lange, E. (2021). Imaging Techniques in Alzheimer's Disease: A Review of Applications in Early Diagnosis and Longitudinal Monitoring. *Int. J. Mol. Sci.* 22:2110. doi: 10.3390/ijms22042110
- Venugopalan, J., Tong, L., Hassanzadeh, H. R., and Wang, M. D. (2021). Multimodal deep learning models for early detection of Alzheimer's disease stage. *Sci. Rep.* 11:3254. doi: 10.1038/s41598-020-74399-w
- Wang, K., Hu, H., and Liu, T. (2018). Discriminative kernel sparse representation via *l2* regularisation for face recognition, Image and vision processing and

- display technology. *IET Inst. Eng. Technol.* 54, 1324–1326. doi: 10.1049/el.2018.6727
- Wong, W. (2020). Economic burden of alzheimer disease and managed care considerations. *Am. J. Manag. Care* 26, S177–S183. doi: 10.37765/ajmc.2020.88482
- Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 210–227. doi: 10.1109/TPAMI.2008.79
- Yang, M., Chang, H., Luo, W., and Yang, J. (2017). Fisher discrimination dictionary pair learning for image classification. *Neurocomputing* 269, 13–20. doi: 10.1016/j.neucom.2016.08.146
- Yao, D., Calhoun, V. D., Fu, Z., Du, Y., and Shi, J. (2018). An ensemble learning system for a 4-way classification of Alzheimer's disease and mild cognitive impairment. *J. Neurosci. Methods* 302, 75–81. doi: 10.1016/j.jneumeth.2018.03.008
- Zhang, X. F., Dou, H., Ju, T., Xu, J., and Zhang, S. T. (2016). Fusing heterogeneous features from stacked sparse autoencoder for histopathological image analysis. *IEEE J. Biomed. Health Inform.* 20, 1377–1383. doi: 10.1109/JBHI.2015.2461671
- Zhang, Z., Ren, J., Jiang, W., Zhang, Z., Hong, R., Yan, S., et al. (2019). Joint subspace recovery and enhanced locality driven robust flexible discriminative dictionary learning. *IEEE Trans. Circuits Syst. Video Technol.* 30, 2430–2446. doi: 10.1109/TCSVT.2019.2923007
- Zheng, S., and Ding, C. (2020). A group lasso based sparse KNN classifier. *Pattern Recogn. Lett.* 131, 227–233. doi: 10.1016/j.patrec.2019.12.020
- Zhou, G., Lu, B., Hu, X., and Ni, T. (2022). Sparse representation-based discriminative metric learning for brain MRI image retrieval. *Front. Neurosci.* 15:829040. doi: 10.3389/fnins.2021.829040
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Liu, Li, Gao, Cai and Ning. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*