Computer Methods and Programs in Biomedicine 199 (2021) 105895

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

CMC: A consensus multi-view clustering model for predicting Alzheimer's disease progression

Xiaobo Zhang[a,b,c], Yan Yang[a,b,c,*], Tianrui Li[a,b,c], Yiling Zhang[a,b,c], Hao Wang[a,b,c], Hamido Fujita[d]

[a] School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China
[b] Institute of Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China
[c] National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu 611756, China
[d] Faculty of Software and Information Science, Iwate Prefectural University, Iwate, Japan

ARTICLE INFO

Article history:
Received 4 September 2020
Accepted 29 November 2020

Keywords:
Consensus representation
Nonnegative matrix factorization (NMF)
Multi-view clustering
Alzheimer's disease (AD) progression
Magnetic resonance imaging (MRI)

ABSTRACT

Machine learning has been used in the past for the auxiliary diagnosis of Alzheimer's Disease (AD). However, most existing technologies only explore single-view data, require manual parameter setting and focus on two-class (i.e., dementia or not) classification problems. Unlike single-view data, multi-view data provide more powerful feature representation capability. Learning with multi-view data is referred to as multi-view learning, which has received certain attention in recent years. In this paper, we propose a new multi-view clustering model called Consensus Multi-view Clustering (CMC) based on nonnegative matrix factorization for predicting the multiple stages of AD progression. The proposed CMC performs multi-view learning idea to fully capture data features with limited medical images, approaches similarity relations between different entities, addresses the shortcoming from multi-view fusion that requires manual setting parameters, and further acquires a consensus representation containing shared features and complementary knowledge of multiple view data. It not only can improve the predication performance of AD, but also can screen and classify the symptoms of different AD's phases. Experimental results using data with twelve views constructed by brain Magnetic Resonance Imaging (MRI) database from Alzheimer's Disease Neuroimaging Initiative expound and prove the effectiveness of the proposed model.

1. Introduction

The social population of elderly people is increasingly serious. Various diseases affect our human physical and mental health through the ages, which is heavily impacting on our families and society. For instance, Alzheimer's disease (AD) is a chronic neurodegenerative disease that usually progresses slowly and then worsens dramatically over time. It often happens in the elderly, and its clinical manifestations are characterized by memory impairment, aphasia, visual space disorders, executive dysfunction and other comprehensive dementias [7,12,45]. AD is predicted to affect 1 of 85 people worldwide by 2050, one billion people approximately [4]. Given such huge amount of AD patients, however, the number of clinicians and medical resources is shown very limited. As this, to alleviate the burden (e.g., diagnosis, care, etc.) for the limited resources to help treat the AD patients is appearing dramatically urgent.

Computer-aided diagnosis technologies can provide an alternative way to alleviate doctors' burden. The intuitions here are at least two-folds: (1) with the rapid development of machine learning, the performance of many relevant models has been remarkably improved, and many machine learning models have been exploited and used as auxiliary diagnosis methods to assist the diagnosis and prediction of many diseases including AD [10,30,32,36,61], and (2) a large amount of medical related data have been collected, e.g., medical text data, medical image data, which can be explored to learn an effective machine learning model. Existing medical neuroimaging technologies mainly include magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), and positron emission tomography (PET). Among these, MRI is a type of tomography, which performs magnetic resonance phenomena to sample electromagnetic signals from the test object. The sampled electromagnetic signals are used to represent the health condition of the test object [35]. In this work, we use

* Corresponding author at: School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756, China.
E-mail address: yyang@swjtu.edu.cn (Y. Yang).

https://doi.org/10.1016/j.cmpb.2020.105895
0169-2607/

two different brain MRI data sets by the Alzheimer's Disease Neuroimaging Initiative (ADNI) platform,[1] which provides some publicly available data sets for research works in the world [54]. There are all six subjects corresponding to each stage of AD progression in the two MRI data sets, that is, Control Normal (CN) i.e. stage 1 of AD progression, Subjective Memory Consider (SMC) i.e. stage 2 of AD progression, Early Mild Cognitive Impairment (EMCI) i.e. stage 3 of AD progression, Mild Cognitive Impairment (MCI) i.e. stage 4 of AD progression, Lately Mild Cognitive Impairment (LMCI) i.e. stage 5 of AD progression and confirmed AD (AD) i.e. stage 6 of AD progression [54].

Although several machine learning techniques have been developed to assist AD diagnosis, existing methods may have two weaknesses. First, they mainly focus on classification or prediction problems of AD, which are supervised learning models and require labeled data to train the models. However, it is costly and time-consuming to collect enough AD data with labels in real applications. Second, learning with multiple views (called multi-view learning) is seldom studied for MRI data. While multi-view learning technology can fully learn the effective features of multiple views and improve the accuracy of data prediction. Also, the motivation of multi-view data combination is that different views have different importance and their own prior knowledge. Thus, it is necessary to generate multi-view MRI data and to develop unsupervised learning models for the automatic auxiliary diagnosis of AD.

Multi-view clustering algorithms have been concerned in recent years. Yang et al. introduced existing multi-view clustering algorithms with a taxonomy according to the mechanisms and principles involved in each algorithm [58]. In addition, some more targeted multi-view clustering techniques are also developed, such as binary multi-view clustering method for multiple view image data [64] and incomplete multiview clustering for incomplete multiview data [55]. In summary, multi-view clustering algorithms can effectively improve the learning performance by combining consistent and complementary information among multiple views. Among these algorithms, non-negative matrix factorization (NMF) based methods belong to a type of representative methods. NMF [17] is able to obtain low-dimensional approximations of non-negative matrices to learn effective features for the data. Liu et al. studied the NMF-based multi-view clustering algorithm (MultiNMF) by searching for a factorization which gives compatible clustering settlements across multiple views [26]. However, it requires manually setting weight parameters for multiple views.

To address the above-mentioned problems, in this work, we propose a **C**onsensus **M**ulti-view **C**lustering model (denoted as CMC) for predicting the multi-stages of AD progression to assist medical diagnosis and medical detection. Here the item 'consensus' denotes that fusing mutiple views automatically to learn a unified representation across all views. In brief, the proposed model integrates consensus fusion, NMF, multi-View learning, and clustering jointly. The proposed model basically works as follows. It first creates twelve views from the original MRI data using several feature extraction algorithms with Scale Invariant Feature Transform (SIFT) [24], KAZE [2] and Gabor filter [47]. Then, all the created views are preprocessed with principal component analysis (PCA) [46] and a normalized processing method. Finally, the processed multi-view data are fed into the proposed matrix factorization model to perform the final predicting task. We approach predicting in a clustering way. The major contributions of this paper are summarized as follows:

1) *Data*. It constructs several brain MRI datasets that fully contain multiple views by using SIFT, KAZE and Gabor filter technologies for medical data with a small data size in real scenes. In this way, the number of initial samples is expanded, and the data features are enhanced, which is conducive to the effect of machine learning.
2) *Methodology*. It proposes a novel unsupervised consensus multi-view clustering model (or simply CMC) via non-negative matrix factorization to predict the multi-stages of AD progression. The key novelty is that it explores multi-view data, avoids manually parameter settings and learns a consensus representation across multiple views.
3) *Results*. Experimental results using two real MRI datasets with twelve views demonstrate that the proposed CMC achieves superior performance compared to other baselines. Experimental results also with two initial views comparing to twelve views verify that constructed new views are useful and beneficial for clustering. This also provides a new solution to the medical examination and prevention of AD.

The rest of this paper is organized as follows. Section 2 introduces the related work of the AD prediction based on machine learning methods. Section 3 presents the proposed consensus multi-view clustering (CMC) model. The optimization algorithm for CMC is shown in Section 4. Experiments are conducted in Section 5. Finally, Section 6 concludes this paper.

2. Related work

In order to recognize different AD classes of patients, Huang et al. developed an random forest architecture based on nonlinear supervised sparse regression for predicting a variety of longitudinal AD clinical scores [14]. Liu et al. provided a multi-scale modeling variantto-function-to-network model to inquire into the causal effect of rare noncoding variants for AD [28]. Previtali et al. proposed a novel feature extraction method from brain MRI scans for classifying AD patient [41]. Vaithinathan et al. presented a new algorithm for the classification problems of AD, which belongs to a kind of texture extraction technique with T1 weighted MRI [49]. Richhariya et al. studied an alternative algorithm based on projection axes termed as least squares projection twin support vector clustering for recognizing AD from MRI data [42]. Wang et al. showed a subspace-based sparse feature learning technique with union-of- subspace representation for AD identification from MRI data[53].

AD using multi-model data has also been studied by some researchers. Li et al. considered a multi-modal supervised within-class-similarity distinctive dictionary learning method based on the weighted combination for AD diagnosis by using the neuroimaging data sets[21]. Tong et al. proposed a multi-modality classification model with nonlinear graph fusion to utilize the complementary information among different modal data of AD [48]. Zhang et al. presented a deep learning model with different convolutional neural networks for the auxiliary diagnosis of AD from multi-modal medical images [60]. Hao et al. studied a novel multi-modal neuroimaging feature selection method with consistent metric constraint for the classification of AD diagnosis, which exploits the advantages of random forest strategy and multi-kernel support vector machine [11].

Deep learning technology can integrate feature learning into the model's building process and reduce the incompleteness from the artificially generated features, which also makes contributions to the prediction of AD [6]. Zhang et al. proposed a new deep learning method based on dense convolutional network, which builds upon the ResNeXt, Adam algorithm and takes the advantages of group convolution technologies [62]. Li et al. furnished a 4D deep learning framework by utilizing a series of 3D convolutional neural networks and the long short-term memory network for AD dis-

[1] www.loni.ucla.edu/adni-2, www.loni.ucla.edu/adni-3

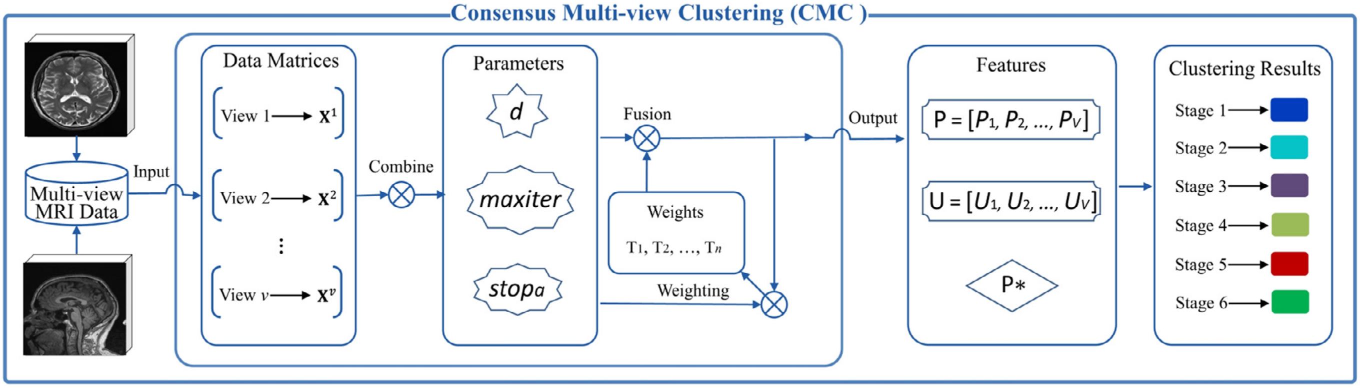


Fig. 1. General framework for our CMC. $\mathbb{X} = [X^1, X^2, \ldots, X^V]$ is the original multi-view data matrices; d is the dimension of consistent representation; *maxiter* denotes the maximum number of iterations; $stop_a$ denotes the stop coefficient.

crimination using 4D fMRI data sets [23]. Lee et al. studied a novel deep learning framework based on magnetic resonance imaging data for AD and mild cognitive impairment diagnosis, which systematically integrated voxel, region, and patch techniques [18]. Liu et al. built a multi-model deep learning method based on convolutional neural network for AD prediction using the structured MRI data [27]. Besides, Zhang et al. developed a discriminative Margin-Sensitive Autoencoder approach for automated AD diagnosis, which mainly implemented the complementary information of multi-view biomedical image features across a semantic-sensitive encoderdecoder paradigm [65].

In addition, other factors related to AD were also studied recently. Fiscon et al. presented a classification approach in a supervised learning manner from the EEG biomedical signals of AD to help medical doctors in the diagnosis formulation [8]. Zhang et al. proposed a novel multi-view learning model for AD detection using the neuroimaging and genetics data [59]. The relation of the AD and the cerebral vasculature were also studied for the auxiliary diagnosis and treatment of AD patients [19,25].

For the early diagnosis of AD, Pais et al. discussed new definitions and challenges in early diagnosis of AD and provided updates on the status of major clinical trials [39]. Li et al. studied a multivariate time series classification model via a neural attention-based deep learning method for AD prediction, which is meaningful for patients in early detection of potential AD risk [20]. Alberdi et al. made a survey to review the multi-modality signals for early AD auxiliary detection [1]. Zhang et al. proposed a new model called strength and similarity guided group sparse representation, which works for recognizing the mild cognitive impairment (i.e. one early stage before confirmed AD) subjects and healthy controls [63]. Pan et al. changed the histogram of oriented gradient descriptor into quantifying the spatial gradients for the detection of AD early stage [40].

Based on the above work, many traditional machine learning and deep learning methods were applied to the AD classification and prediction with the single-mode ultrasound, text information and multi-modality medical data. At the same time, the technologies of early diagnosis of AD have gradually attracted attention and also have been studied to predict the disease severity of AD. However, learning with multiple views for AD is a minority. Thus, this paper constructs multi-view brain MRI data from the original MRI data in experiments and develops a novel clustering method with a consistent expression across multiple views. In the next section, we will detail the methodology of our proposed CMC model for the prediction of multiple stages in AD.

3. Methodology

As an advanced matrix factorization technology, NMF has been used in various machine learning applications, such as recommendation system, image classification and clustering [15,22,31]. Given multi-view data, to capture key information among multiple views and learn the neighborhood relationship between samples, we propose a NMF-based model called CMC (as shorthand for **C**onsensus **M**ulti-view **C**lustering), which can integrate multi-dimension features in multiple views and automatically learn a concensus representation matrix to assist in clustering. The framework of CMC is shown in Fig. 1, which can fuse more features of different views, especially for multi-view MRI data sets with the characteristics of small quantity and small difference, so as to capture much better clustering results than that of single view. The details of the proposed model will be clear shortly. Before that, let us make a brief review to NMF.

NMF [17] has shown obvious competitive edge in machine learning methods. The conventional matrix factorization methods can factorize a matrix and capture two smaller matrices possessing positive and negative factors. As a matter of fact, more negative features from matrices are above rubies, for instance, same message in text file. Hence, it is of importance to factorize one matrix to two nonnegative matrices, that is named as NMF.

Recently, NMF-based techniques have also been studied with development of artificial intelligence. A word embedding was merged into semi-NMF to capture semantical relationships among words and to perform document clustering [43]. Aroused with manifold learning and Principal Component Analysis (PCA), one NMF model with robust-graph regularization [13] was developed, which expressly resumed the low-rank clean data with a factorization by utilizing the nonlinear structures of a manifold. Furthermore, a new NMF model with structured incoherence and low-rank representation was presented in the work [29]. From these existing works, it is worth noting that NMF is of service to process complex data and abstract the latent structure. It is meaningful to combine the NMF to address multi-view clustering problems. It has absolutely been proven that NMF is hard to handle when the number of basis vectors is larger [50]. In addition, the conditions about the solution of NMF matter are shown in Arora et al. [3].

Given a nonnegative matrix $X = [x_1, x_2, \ldots, x_n] \in R_+^{m\times n}$, which includes n samples with m dimensions, NMF aims to learn two nonnegative matrices, called basis matrix $W \in R_+^{m\times r}$ and representation matrix $H \in R_+^{r\times n}$ to represent the original data matrix X, where r denotes the reduced dimension. Such a way can be formulated as $X \approx WH$. Formally, the problem of NMF is solved as follows:

$$\arg\min_{W \geq 0, H \geq 0} \quad \|X - WH\|_F^2. \tag{1}$$

Then a multiplicative optimization algorithm was proposed to find the results of W and H [17]. We first take into account NMF to extract the latent features from initial data matrix, learn the frequent representation in multiple data views and then put forward a consensus clustering model for screening AD.

3.1. Relational matrix

In general, data samples/instances in different views may have similar relationship patterns. For example, the diagnosis results of the same AD patients from MRI and CT images are usually similar even though they concern different aspects of brain and use different measure standards. Therefore, we consider to utilize neighbor relationship as shared knowledge among views and define the relational matrix to assist in integration of different views.

We first construct relational matrix for each view. The relational matrix S^k of the kth view is defined as follows:

$$S_{ij}^k = \begin{cases} 1 \\ 0 \end{cases} \tag{2}$$

where $S^k \in R^{N \times N}$. This is a binary relationship. That is, if the ith sample belongs to the neighbor set of the jth sample, then $S_{ij}^k = 1$; otherwise, $S_{ij}^k = 0$.

3.2. Consensus multi-view nonnegative matrix factorization clustering algorithm

In this work, we exploit the property of NMF to obtain the effective features of each single view, and combine the similar neighbor patterns to fuse knowledge from multiple views. It is capable of fusing shared and complementary information among views, avoiding manually parameter setting for each view and directly learning a concensus representation matrix containing sufficient and effective features.

Suppose there are V views corresponding to different sets of features, and $\mathbb{X} = \{X^{(1)}, \ldots, X^{(V)}\}$ represents the original multi-view data matrices, the proposed objective function $\mathcal{L}$ is formally defined as bellow:

$$\begin{aligned} \mathcal{L} = \quad & \sum_{k=1}^{V} \|X^{(k)} - P^{(k)}U^{(k)}\|_F^2 \\ & + \lambda_1 \|P^{(k)}\|_1 + \lambda_2 \|P^{(k)} - S^k P^*\|_F^2 \\ & s.t.\ U^{(k)T}U^{(k)} = I \end{aligned} \tag{3}$$

where $X^{(k)} \in R^{N \times M_k}$ (N is the number of samples; and M_k is the dimension of features of the kth view); $P^{(k)} \in R^{N \times d}$ is the representation matrix of the kth view (d is the dimension after factorization for all views); similarly, $U^{(k)} \in R^{d \times M_k}$ denotes the basis matrix; and P^* is defined as the concensus representation matrix; λ_1 and λ_2 are trade-off parameters. As shown in Eq. (3), there are three terms in our objective function. The first item is to factorize the original data into a feature matrix and a basis matrix by using the property of NMF; the second is a L_1 norm for each feature view; the last one is to learn a concensus representation matrix P^*, so that each feature view can be expressed as the multiplication of specific similarity pattern and the consistent feature representation.

In this work, the motivation that we introduce the L_1 and L_2 norms is to keep the sparsity of solutions and to avoid the drawback of overfitting. In addition, we exploit the neighbor pattern and relational matrix to connect multiple views and to capture the consistent and complementary information from different data views to obtain a unified multi-view representation. In such a way, it can directly fuse the features from multiple views and do not need to manually set the weight of each view.

4. Optimization algorithm

The difficulties of settling the proposed model mainly lies in computing the $P^{(k)}$, $U^{(k)}$ and P^*. Given the model proposed in the previous section, the optimization processes of approaching and solving $P^{(k)}$, $U^{(k)}$ and P^* using the objective function $\mathcal{L}$ are presented in this section. We utilize an altering scheme to solve each problem respectively.

4.1. Optimization process for solving $P^{(k)}$

To derive the algorithm for solving $P^{(k)}$, we rewrite the objective function (i.e., Eq. (3)) as

$$\begin{aligned} \mathcal{L} \iff \quad & tr(X^{(k)T}X^k) + tr(P^{(k)T}P^{(k)}) - 2tr(X^{(k)T}P^{(k)}U^{(k)}) \\ + \quad & \lambda_1 \|P^{(k)}\|_1 + \lambda_2 \|P^{(k)} - S^k P^*\|_F^2. \end{aligned} \tag{4}$$

Then removing the terms unrelated to $P^{(k)}$, Eq. (4) is simpled as below (here we use $\mathcal{L}'$ to denote the result):

$$\begin{aligned} \mathcal{L}' = \quad & tr(P^{(k)T}P^{(k)}) - 2tr(X^{(k)T}P^{(k)}U^{(k)}) \\ & + \lambda_1 \|P^{(k)}\|_1 + \lambda_2 \|P^{(k)} - S^k P^*\|_F^2. \end{aligned} \tag{5}$$

In order to update the $P^{(k)}$ conveniently, we introduce two variables denoted as $E^{(k)}$ and $F^{(k)}$, where $E^{(k)}$ is defined as:

$$E^k = (1 + \lambda_2)I. \tag{6}$$

$F^{(k)}$ is defined as:

$$F^k = X^{(k)}U^{(k)T} + \lambda_2 S^k P^*. \tag{7}$$

Then, the $\mathcal{L}'$ is reformulated as:

$$\begin{aligned} \mathcal{L}' \quad & = tr(P^{(k)T}E^k P^{(k)}) - 2tr(F^{(k)T}P^{(k)}) + \lambda_1 \|P^{(k)}\|_1 \\ & = (1 + \lambda_2) tr(P^{(k)T}P^{(k)}) - 2tr(F^{(k)T}P^{(k)}) + \lambda_1 \|P^{(k)}\|_1. \end{aligned} \tag{8}$$

The derivation of Eq. (8) with respect to $P^{(k)}$ is shown as below

$$\frac{\partial \mathcal{L}'}{\partial P^{(k)}} = 2(1 + \lambda_2)P^{(k)} - 2F^{(k)T}. \tag{9}$$

Let $\frac{\partial \mathcal{L}'}{\partial P^{(k)}} = 0$, we have

$$P^{(k)} = \frac{F^{(k)T}}{1 + \lambda_2}. \tag{10}$$

4.2. Optimization process for solving $U^{(k)}$

To derive the algorithm for solving $U^{(k)}$, the terms unrelated to $U^{(k)}$ are also removed in the objective function $\mathcal{L}$, and the resulting formula is derived as follows:

$$\mathcal{L} = -2tr(X^{(k)T}P^{(k)}U^{(k)}). \tag{11}$$

Then we have

$$\min(\mathcal{L}) \iff \max(tr(X^{(k)}P^{(k)T}U^{(k)})). \tag{12}$$

Eq. (12) is an orthogonal procrustes problem according to the work [34]. Suppose that

$$SVD(X^{(k)T}P^{(k)}) = A\Sigma B^T \tag{13}$$

where SVD is the singular value decomposition, Σ denotes the positive semi-definite diagonal matrix, and A and B are the left and right singular arrays of $X^{(k)T}P^{(k)}$, respectively.

Then, $U^{(k)}$ can be solved with

$$U^{(k)} = AB^T. \tag{14}$$

4.3. Optimization process for solving $P^{(*)}$

To solve $P^{(*)}$, the terms unrelated to $P^{(*)}$ are omitted in the objective function $\mathcal{L}$. The resulting formula is shown as below

$$\mathcal{L} = \sum_{k=1}^{V} \|P^{(k)} - S^k P^*\|_F^2. \tag{15}$$

where S^k is the self-existent prior knowledge of data objects, which represents the relationship between data samples in the kth view dataset, that is, similarity. This prior relationship is used as weight to learn the consensus view in the algorithm of our CMC model. It can better consider the structure of each view itself, so as to obtain a more effective consensus view in the learning process.

Then the derivation of Eq. (15) with respect of ∂P^* is

$$\frac{\partial \mathcal{L}}{\partial P^*} = 2\sum_{k=1}^{V} S^{(k)T}(S^k P^* - P^{(k)}). \tag{16}$$

Let

$$\frac{\partial \mathcal{L}}{\partial P^*} = 0, \tag{17}$$

then we have

$$2\sum_{k=1}^{V} S^{(k)T} S^k P^* = 2\sum_{k=1}^{V} S^{(k)T} P^{(k)}. \tag{18}$$

Hence, the $P^{(*)}$ is optimized as follows:

$$P^* = (\sum_{k=1}^{V} S^{(k)T} S^k)^{-1} (\sum_{k=1}^{V} S^{(k)T} P^{(k)}). \tag{19}$$

In summary, the overall optimization algorithm is shown in Algorithm 1.

Algorithm 1: The optimizing process of our proposed *CMC* model.

Input: The original multi-view data matrices $\mathbb{X} = [X^1, X^2, \ldots, X^V]$, the dimension of the consistent representation d, the maximum iteration number *maxiter*, and the stop coefficient $stop_a$.

Output: $\mathbb{P} = [P^1, P^2, \ldots, P^V]$, $\mathbb{U} = [U^1, U^2, \ldots, U^V]$, and $P*$.

Initial representation matrices $\mathbb{P} = [P^1, P^2, \ldots, P^V]$;
Initial basis matrices $\mathbb{U} = [U^1, U^2, \ldots, U^T]$;
Initial concensus representation matrix $P*$;
Initial objective function $\mathcal{L}$ (i.e., Eq.3);
for $i = 1 : maxiter$ **do**
  for $k = 1 : V$ **do**
    Update P^k according to Eq.~10;
    Update U^k according to Eq.~14;
    Update $P*$ according to Eq.~19;
  end for
  Normalize the elements of $P*$;
  if $loss \leq stop_a$ **then**
    break;
  end if
end for
return $\mathbb{P}$, $\mathbb{U}$, and $P*$.

5. Experiments and results

In this section, we introduce our experiments and results with five parts, i.e., datasets, experiment settings, multi-view data construction and preprocessing, and experimental results and analysis.

Table 1
ADNI2: two-view MRI data with six classes corresponding to the multi-stages of AD progression.

Subjects	# Samples	# View 1	# View 2	Multi-stages
CN	287	2449	2480	stage 1
SMC	110	709	567	stage 2
EMCI	307	2753	2469	stage 3
MCI	79	296	406	stage 4
LMCI	176	1808	1658	stage 5
AD	159	1150	1016	stage 6

Table 2
ADNI3: two-view MRI data with six classes corresponding to the multi-stages of AD progression.

Subjects	# Samples	# View 1	# View 2	Multi-stages
CN	408	471	998	stage 1
SMC	48	64	122	stage 2
EMCI	74	107	229	stage 3
MCI	143	173	362	stage 4
LMCI	35	54	105	stage 5
AD	44	52	117	stage 6

5.1. Datasets

We use brain MRI datasets from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database[2] to evaluate the proposed method. ADNI is an international, large-scale and longitudinal multicenter study, which unites most researchers with study data as they work to determine the progression of AD [54].

We downloaded the MRI datasets of ADNI2 and ADNI3 database from the ADNI platform upon our vising time (6th September 2019). The number of subjects/classes in our MRI data is six in total, i.e., CN, SMC, EMCI, MCI, LMCI and AD. That is, there are normal samples, 4 different developmental states of AD patients and confirmed AD samples. Each sample of all the six classes is represented with two views (i.e., AXI view as View 1 and SAG view as View 2). The statistics of the two datasets are summarized in Tables 1 and 2.

In Tables 1 and 2, the total number of samples is 1118 and 752, respectively. The total MRI number in view 1 and view 2 for each class is also presented in Tables 1 and 2, respectively. However, the number of MRI data from View 1 and View 2 is different. To solve the problem of inconsistent data volumes, we select the largest one MRI data from View 1 and View 2 for each sample, which ensures the consistency of the data quantity of the two views. Specifically, each sample has two MRI images from View 1 and View 2. Hence, the ADNI2 datasets of View 1 and View 2 have 1118 original images. Meanwhile, the ADNI3 datasets have 752 original MRIs for View 1 and View 2, respectively. Some examples of the AXI and SAG MRI data of the six subjects from ADNI2 and ADNI3 database are shown in Fig. 2. In addition, in order to use samples as many as possible to train our model, we construct twelve views datasets in the experiments. The constructing method will be clear shortly recommended.

5.2. Experiment settings

All experiments are conducted on a PC Server (Intel(R) Xeon (R) CPU E5-2620 v4 @2.10 GHZ, 256 GB RAM, GPU NVIDIA TITAN Xp with 16 GB).

For the data in our experiments, we first randomly sample the initial brain MRI datasets from ADNI platform with two views. Second, we use the feature extraction methods called Scale Invariant

[2] http://adni.loni.usc.edu/

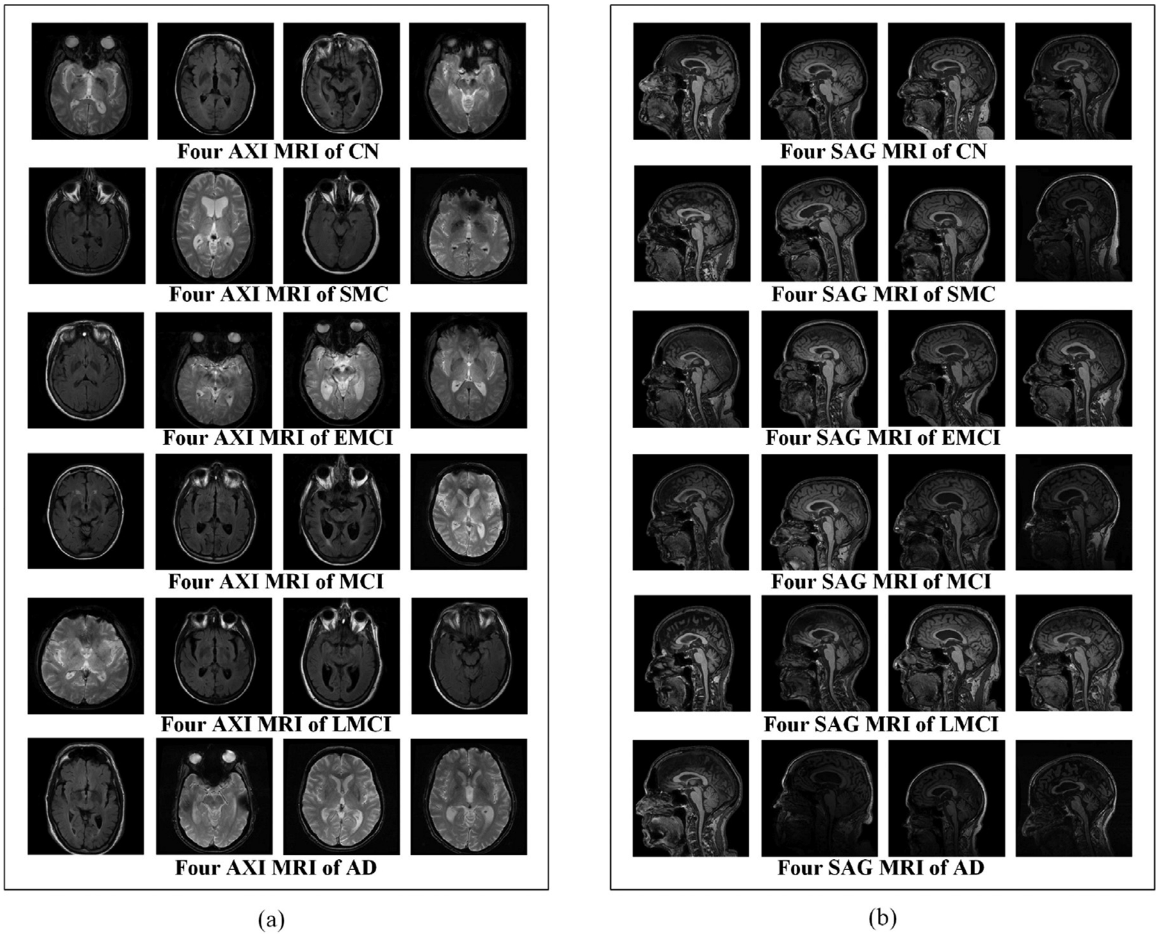


Fig. 2. (a) Four images of the six subjects for AXI MRI view. (b) Four images of the six subjects with SAG MRI view.

Table 3
The setting value for each parameter of CMC.

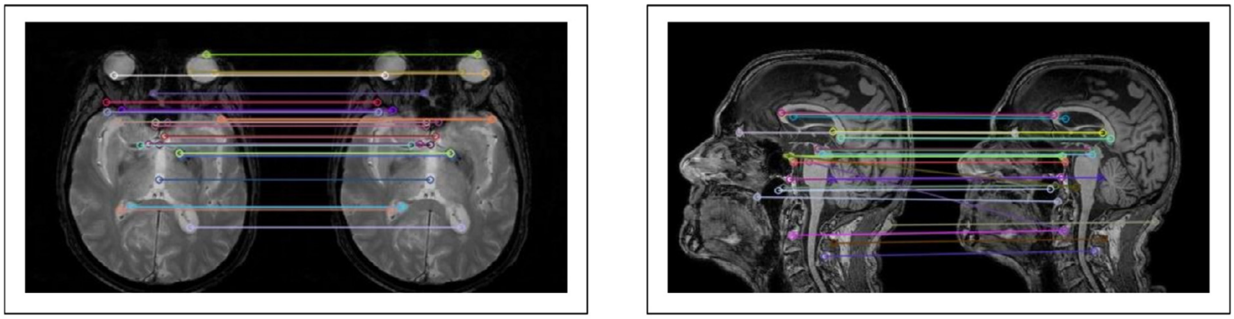
Parameters	# Descriptions	# Values
V	Number of views in the datasets	12
N	Number of objects for per view in ADNI2, ADNI3	1118, 752
D	Matrix Dimensions Constructed by Intermediate Processes	500
k	Data dimension after dimensionality reduction with PCA	800
lamp1	The trade-off parameter	0.5
lamp2	The trade-off parameter	0.4
maxiter	Number of matrices produced in the intermediate processes	200
$stop_a$	The stop coefficient	0.1

Feature Transform (SIFT) [24] and KAZE [2] to construct two views' features from each original view data. Then, the newly constructed four views are changed and expanded to eight views with Gabor filter technology [47], which will be illustrated in the next subsection. Meanwhile, all the twelve views are preprocessed using PCA and normalized processing method.

For the baselines, we compare our CMC with eight solid multi-view clustering methods, i.e., Multiview Spectral Embedding (MSE) [57], Weighted Robust Multi-view K-means (WRMK) [5], Multi-view clustering via joint Nonnegative Matrix Factorization (MultiNMF) [26], Robust Multi-View Spectral Clustering (RMSC) [56], Parameter-free auto-weighted Multiple Graph Learning (AMGL) [38], Multi-View Clustering via Concept factorization (MVCC) [51], Multi-view Learning with Adaptive Neighbors (MLAN) [37] and Graph-based Multi-view clustering (GMC) [52]. We also compare two classic methods in single view clustering, i.e., K-Means [44] and Affinity Propagation (AP) [9]. The clustering performance is measured using four metrics, i.e., the Accuracy (ACC), Normalized Mutual Information (NMI), Precision Recall F1 (F1) and Adjusted rand index (ARI) values [33]. In addition, the setting values for each parameter of our CMC are shown in Table 3.

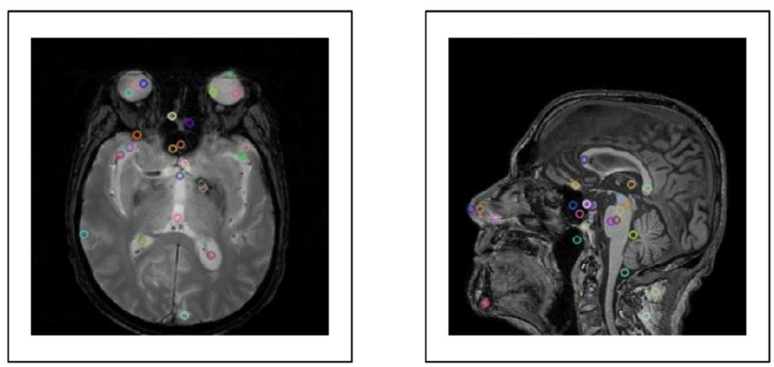
5.3. Multi-view data construction and preprocessing

Since the sampled set of medical data is usually small, we need to fully learn the feature information in such limited data samples. Therefore, to further mine the hidden knowledge and obtain the sufficient information of samples, the methods of multi-view data construction are performed on the original brain MRI datasets. In this subsection, we first introduce the detailed process of constructing multi-view datasets with the techniques of SIFT, KAZE and Gabor filter. Then, all the multi-view datasets are preprocessed with the PCA and the normalized processing method to obtain abundant features.



(a) (b)

Fig. 3. (a) The image feature map for one sample of AXI MRI with SIFT. Left: The original AXI MRI. Right: The extracted feature map from original MRI. (b) The image feature map for one sample of SAG MRI with SIFT. Left: The original SAG MRI. Right: The extracted feature map from SAG MRI.



(a) (b)

Fig. 4. (a) The original AXI MRI sample and its feature map with KAZE. (b) The original SAG MRI sample and its feature map with KAZE.

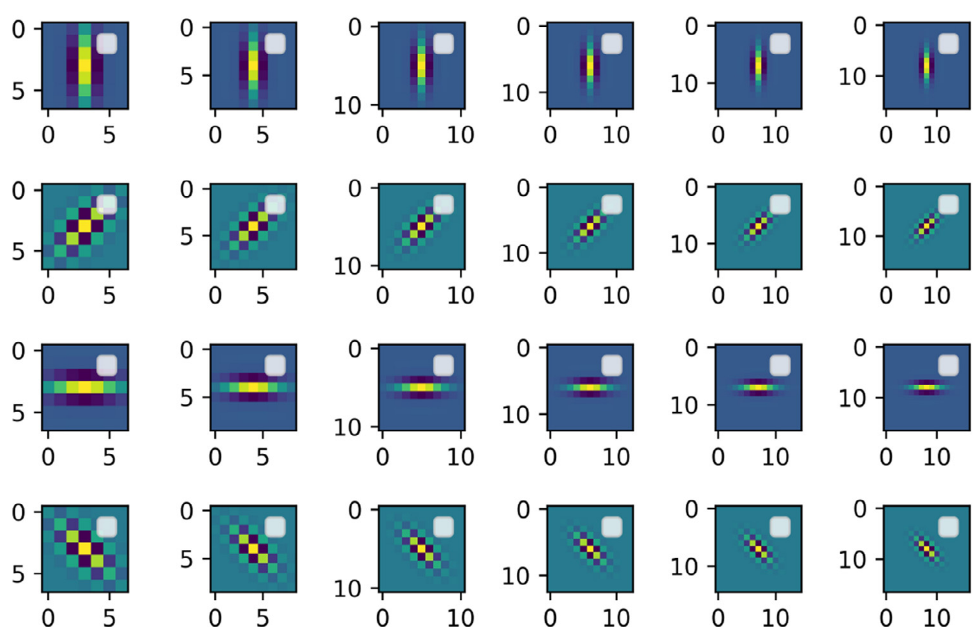


Fig. 5. The effect map of the Gabor filters with six scales and four directions.

SIFT feature extraction As SIFT is a feature descriptor with scale invariance and illumination invariance, which can effectively guarantee the quality information of the MRI [24]. Hence, the method of SIFT feature extraction is chosen to extract local features for the original datasets. And the image extracted features with SIFT for one sample of AXI MRI i.e. View 1 is shown in the Fig 3(a). Besides, the image processed with SIFT for one sample of SAG MRI i.e. View 2 is presented in Fig 3(b).

The intuitive analyse of Fig. 3 shows that the SIFT feature matching among different brain MRI images provides the knowledge of feature consistency of different objects.

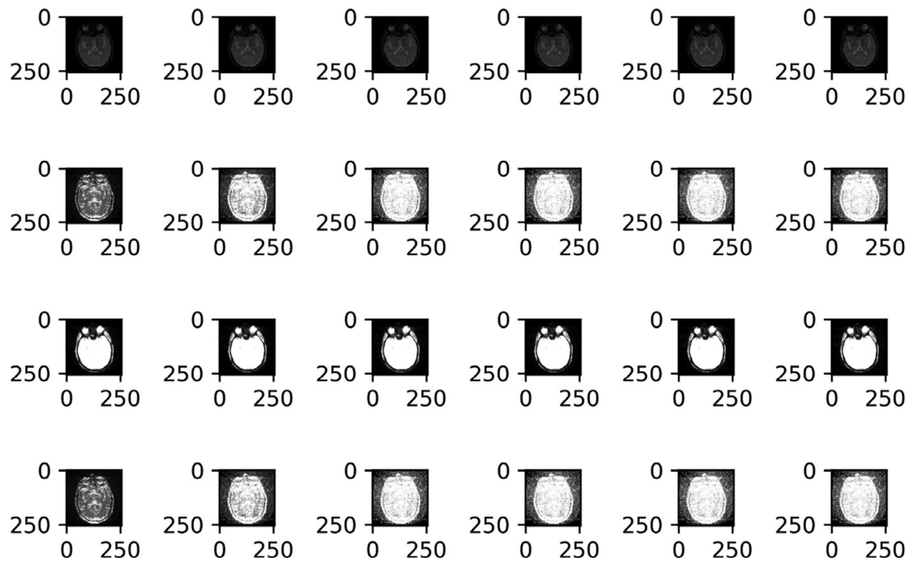


Fig. 6. The examples of filtered images from the original brain AXI MRI datasets by the Gabor filters. Note that the display drawing of SAG MRI examples through the Gabor filters is the same with it.

KAZE feature extraction In addition, KAZE is a feature point detection method based on a non-linear scale space. This non-linear scale space shows that the amount of information loss at the edge of the image in the scale change is small, thereby greatly maintaining image detail information [2]. The KAZE is also exploited for extracting features of the original data. The images extracted KAZE features for one sample of AXI and SAG data views are shown in Fig 4.

An intuitive analysis of Fig. 4 shows that the left is for the AXI MRI map and the right is for the SAG MRI map precessed by the KAZE technology. The locations of key points are also displayed in the image, which effectively captures the details of the brain MRI.

Gabor filter feature extration Besides above, Gabor filter is a windowed Fourier transform, which can extract related features at different scales and directions in the frequency domain, and is particularly suitable for texture feature extraction of biological images [47]. In order to further achieve more views and collect more effective information of MRI images, we adopt Gabor filters with six scales of [7, 9, 11, 13, 15, 17] to extract spatial local frequency features of the original two views datasets in the four directions of [0°, 45°, 90°, 135°]. The effect map of the Gabor filters with these six scales and four directions is in Fig. 5.

A comparative analysis of Fig. 5 shows that the Gabor filters perform superiorly when the scale is set as 15. Hence, we utilize the Gabor filters with one scale of 15 and four directions of [0°, 45°, 90°, 135°] to construct 8 texture feature views. And the examples of filtered images from the original brain AXI MRI datasets by the Gabor filters are shown in Fig. 6.

An intuitive analysis of Fig. 6 shows the filtered brain MRI effects processed by the Gabor filters with one scale of 15 and four directions of [0°, 45°, 90°, 135°]. And the display graph of the filtered brain SAG MRI effects is very similar to Fig. 6, so it is not presented here.

Multi-view datasets preprocessing Based on the above experiments, the original 2 views MRI datasets are expanded into 12 views for each dataset, separately. Also, each MRI has 1024 pixel features. Therefore, there are 1118 objects, 1024 features and 6 subjects for each view from ADNI2 database. In addition, the 752 objects, 1024 features and 6 subjects for each view are obtained from ADNI3 database. Considering the large amount of data and operating efficiency, the useful information of 1024 features need to be selected and processed. In practice, we utilize PCA method to make the dimensions reduction of image data while persisting most of the changes in the dataset [46]. The 800 features of each view from ADNI2 database and the 500 features of each view from ADNI3 database are selected by the PCA method, which are normalized finally.

5.4. Experimental results and analysis

In this section, we exhibit the results and analyses of the baselines (i.e., K-Means [44], AP [9], MSE [57], WRMK [5], MultiNMF [26], RMSC [56], AMGL [38], MVCC [51], MLAN [37], GMC [52]) and our CMC model using two datasets. Note that for the two single-view clustering methods (i.e. K-Means and AP), we record the performance on each view and finally show the average performance across all the 12 views.

5.4.1. Performance evaluation

The performance results of each method in terms of ACC, NMI, F1 and ARI on ADNI2 and ADNI3 datasets are shown in Table 4, separately.

According to the experimental results from Table 4, we have the following observations.

- The best clustering performance results on each datasets are highlighted with bold font in Table 4. Also, the clustering results of each method on the two data sets with ACC, NMI, F1 and ARI are different. Almost most multi-view clustering methods achieve superior results than the two single-view approaches.
- For ADNI2 data set, it is noted that our CMC achieves the best ACC and NMI values for predicting AD progression with the 12 brain MRI views. In term of ACC performance, the best ACC value of CMC is 34.07%, which is higher than the other better method i.e. WRMK by almost 5 percentage points. And in term of NMI, our CMC captures 2.13%, which is higher than the other better method i.e. RMSC by nearly 1 percent.
- Based on ADNI2 data set, the performance of CMC is not very outstanding in terms of F1 and ARI. The F1 value of CMC is 32.16%, which is lower than the best F1 value i.e. 40.07% from MultiNMF method by almost 8 percent. And for the NMI performance, there are two methods (RMSC and GMC) capturing

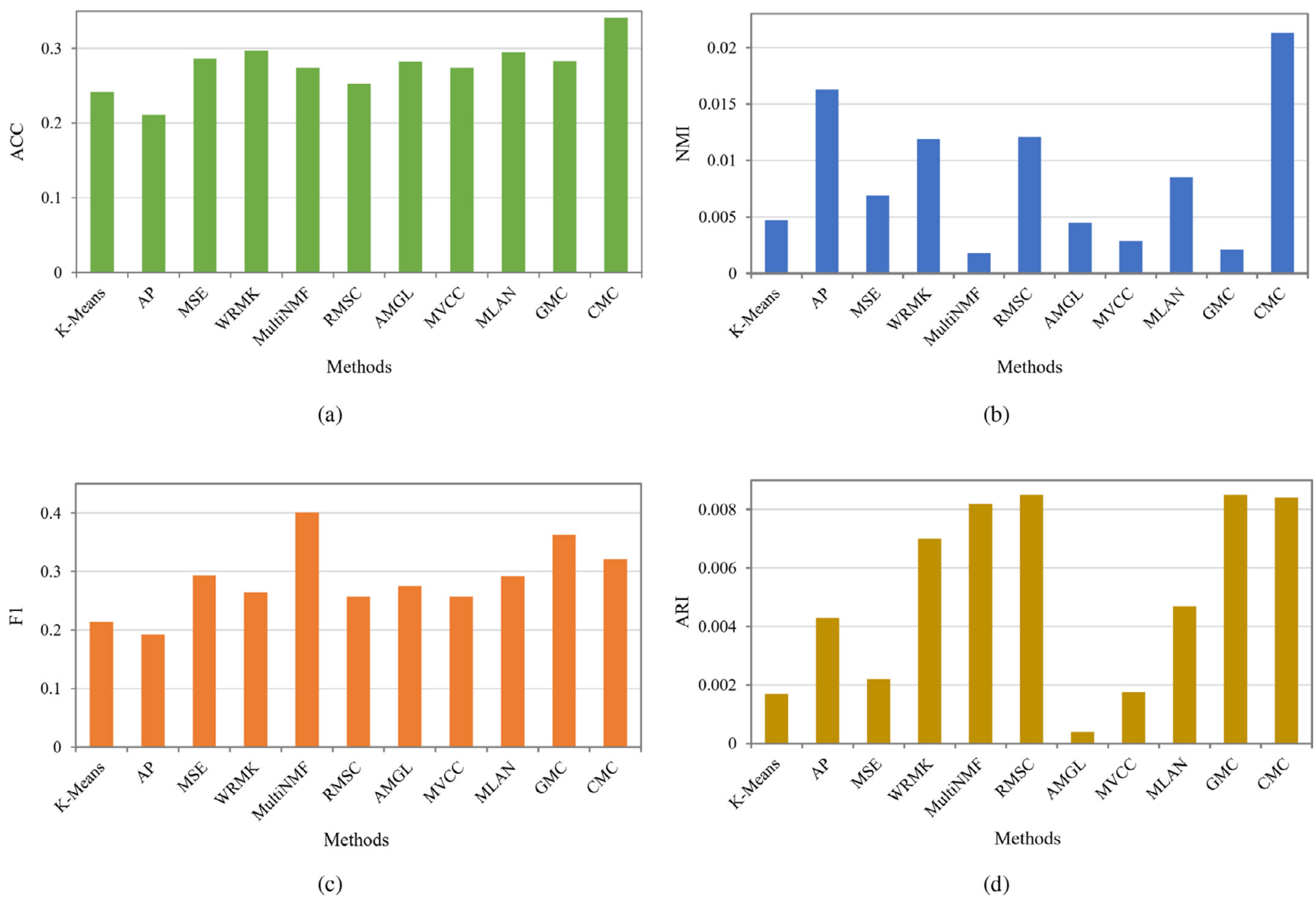


Fig. 7. The comparison results of each clustering performance with the 12 views on the ADNI2 dataset.

Table 4
The performances of each clustering method on the two of ADNI2 and ADNI3 datasets.

Datasets	ADNI2				ADNI3			
Evaluations	ACC	NMI	F1	ARI	ACC	NMI	F1	ARI
K-Means	0.2415	0.0047	0.2141	0.0017	0.3105	0.0052	0.3344	0.0009
AP	0.2114	0.0163	0.1920	0.0043	0.3487	0.0050	0.3644	0.0057
MSE	0.2865	0.0069	0.2933	0.0022	0.3513	0.0096	0.3464	0.0151
WRMK	0.2971	0.0119	0.2647	0.0070	0.3096	0.0103	0.3086	0.0058
MultiNMF	0.2746	0.0018	**0.4007**	0.0082	0.5426	0.0071	0.5169	0.0105
RMSC	0.2526	0.0121	0.2565	**0.0085**	0.3065	0.0119	0.2996	0.0028
AMGL	0.2823	0.0045	0.2749	0.0004	0.3793	0.0049	0.3953	0.0017
MVCC	0.2745	0.0029	0.2565	0.0018	0.2980	0.0046	0.3093	0.0067
MLAN	0.2951	0.0085	0.2921	0.0047	0.5452	0.0119	0.5367	0.0050
GMC	0.2835	0.0021	0.3629	**0.0085**	0.5226	0.0064	0.5008	**0.0201**
CMC	**0.3407**	**0.0213**	0.3216	0.0084	**0.5726**	**0.0260**	**0.5968**	0.0126

the best same value i.e. 0.85%, while the NMI of our CMC gets 0.84%, of which exist 0.01 percent gap.

- For ADNI3 data set, it can be seen that our CMC obtains the best ACC, NMI and F1 values with 12 brain MRI views. The best ACC of CMC is 57.26%, which is higher than the other better method i.e. MultiNMF by 3 percent. And the NMI of CMC is 2.60%, which is higher than the other better NMI from RMSC method by almost 1.5 percent. At the same time, the F1 of CMC achieves 59.28%, that is higher than the other better F1 value of MultiNMF about 8 percent. Besides, in term of ARI, our CMC result is 1.26%, which is lower than MSE and GMC methods.
- For ADNI2 and ADNI3 data sets, the performances of our CMC model are not very optimistic in terms of ARI metric. The reason may be that there are six states of AD, and the distance and discrepancy among them are usually small. So, it may need to improve our method by depicting and defining more specific features for each state of AD.
- In addition, it's real that NMI and ARI scores on the two data sets are so small in Table 4. The main reason is that the features of medical MRI image data are very sparse and the difference between different classes of original sample MRI data is little, which makes some difficulties to machine learning methods.
- In general, the clustering performances of our CMC model are much better than other baselines, which shows that our CMC is capable of capturing effective features and detailed information from MRI data.

More intuitively, in case of the 12 views are from ADNI2, the comparison results of experiments for each clustering method are

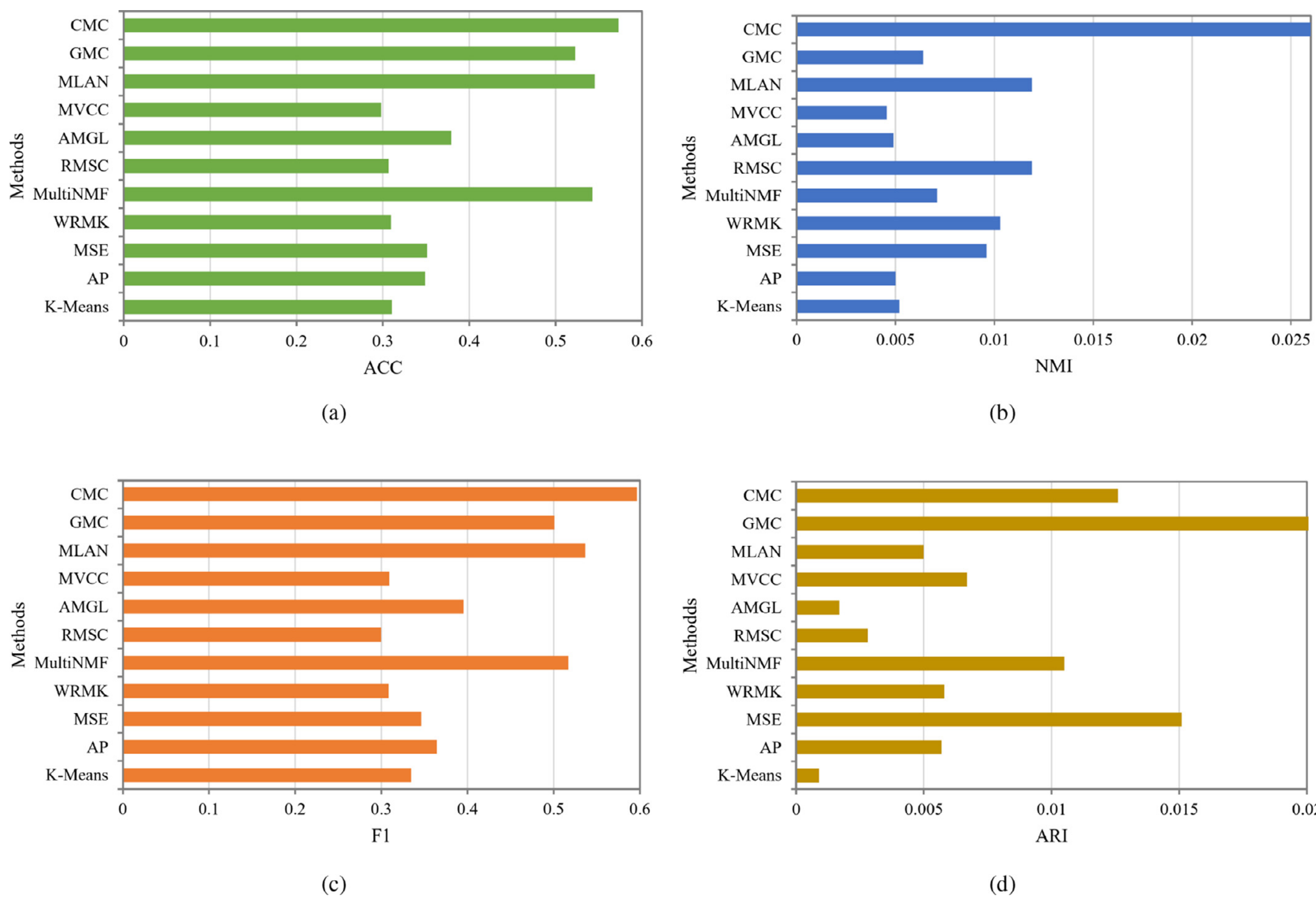


Fig. 8. The comparison results of each clustering performance with the 12 views on the ADNI3 dataset.

shown in Fig. 7, while the comparison results are presented in Fig. 8 when the datasets come from ADNI3.

A comparative analysis of Figs. 7 and 8 respectively shows that our CMC method achieves the highest values in terms of ACC, NMI and F1 metrics. Figs. 7 and 8 also show that the performance of multi-view clustering methods is superior to the performance of single-view clustering methods. This due to that multi-view clustering is able to capture more effective results for screening AD than single-view clustering.

In addition, to verify the usefulness and difference of other data views by above feature extraction tools, we also implement our CMC model on ADNI2 and ADNI3 datasets with initial 2 views (i.e. AXI and SAG views) separately. And the comparable experiment results with 2 intial views and 12 views by feature extraction tools (i.e. SIFT [24], KAZE [2] and Gabor filter methods [47]) are shown in Fig. 9.

From Fig. 9, we can observe that the ACC, NMI, F1 and ARI values of CMC model based on 12 views of ADNI2 and ADNI3 data sets are obviously higher than that of 2 views, separately. It is demonstrated that the 12 views data by those feature extraction tools are more useful and effective than the initial 2 views data for clustering.

5.4.2. Visualization results

To be more intuitive, we make the learned latent representation visualization using t-Distributed Stochasitc Neighbor Embedding (t-SNE) [16] on each dataset. The visualization results are shown in Fig. 10. It is obviously to see the performances in Fig. 10 are well consistent with the outcomes in Table 4. The results from Fig. 10 also shows that our model (i.e. CMC) can mine the underlying cluster structure and achieve a compact latent representation.

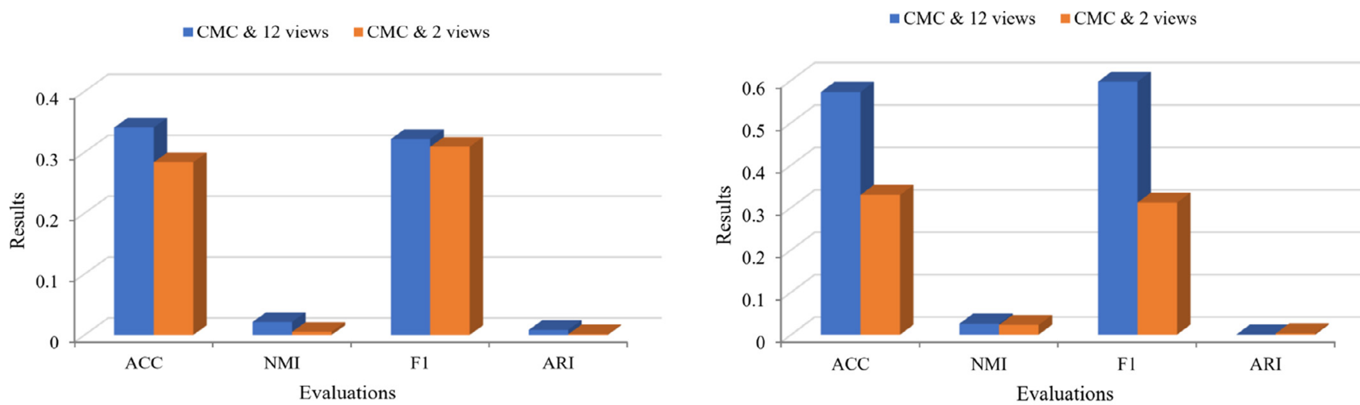
5.4.3. Parameters selection

In the experiments, the trade-off parameters λ_1 and λ_2 of CMC are chosen by searching a wide range of grids $\{0.1, 0.2, 0.3, \ldots, 1\}$. Figs. 11 and 12 show how the average clustering performance of our CMC model changes by different parameters λ_1 and λ_2, where the horizontal axis indicates the variation of λ_1 and the vertical axis expresses the variation of λ_2. Considering the instable results in Figs. 11 and 12, λ_1 and λ_2 are set to moderate values, such as λ_1=0.5 and λ_2 =0.4. Also, each grid with different shades of colors reflects the clustering effects of CMC method.

5.4.4. Convergence analysis

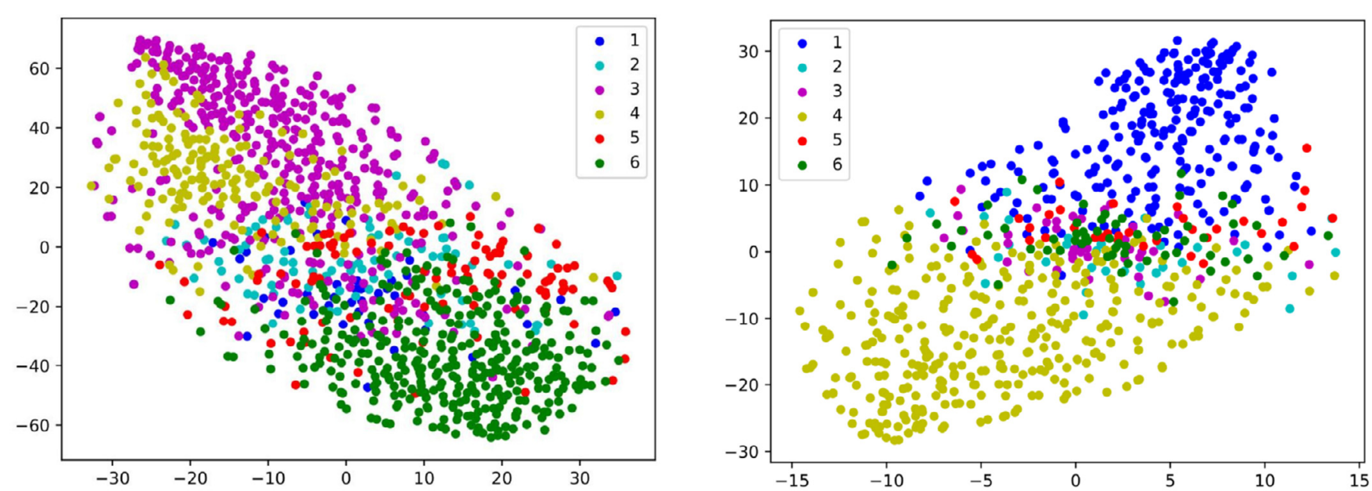
Next, we experiment the convergence anaslysis on the two different datasets. The convergence curves of the loss function on the datasets ADNI2 and ADNI3 are shown in Fig. 13. From the results, we can see that our algorithm CMC converges quickly. Specifically speaking, it converges within 90 iterations on the dataset ADNI2 and within 75 iterations on the dataset ADNI3.

In summary, a comparative analysis from the results in Fig. 13 shows that our CMC method is able to converge around 100 iterations. Besides, our CMC model performs the best in predicting the six classes (i.e., CN, SMC, EMCI, MCI, LMCI and AD). Although the highest accuracy of screening AD is only 34.07% for ADNI2 and 57.26% for ADNI3, our CMC model only uses the original data of MRI without any other annotation information such as gender, age, gene and others. Also, the datasets with these four Labels (i.e., CN, SMC, MCI and AD) can be achieved better clustering performance



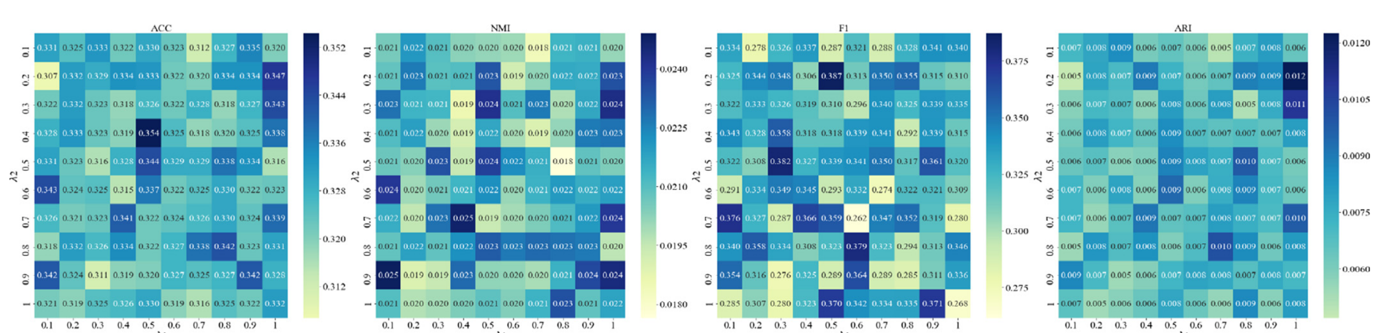
(a) ADNI2 (b) ADNI3

Fig. 9. The comparison results of CMC performance with the initial 2 views (i.e. AXI and SAG views) and 12 views on two datasets.



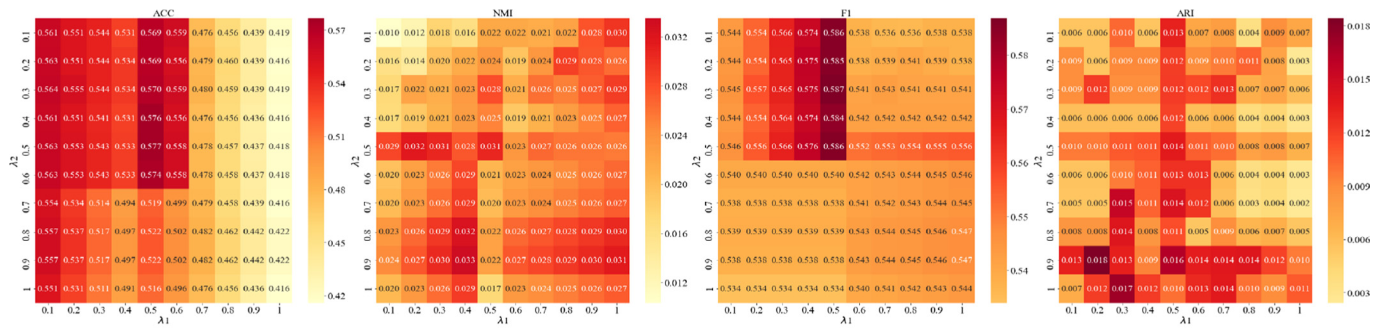
(a) (b)

Fig. 10. (a) The visualization of the latent representation on the dataset ADNI2. (b) The visualization of the latent representation on the dataset ADNI3. Note that the numbers (1, ..., 6) in the legend denote the six class labels.



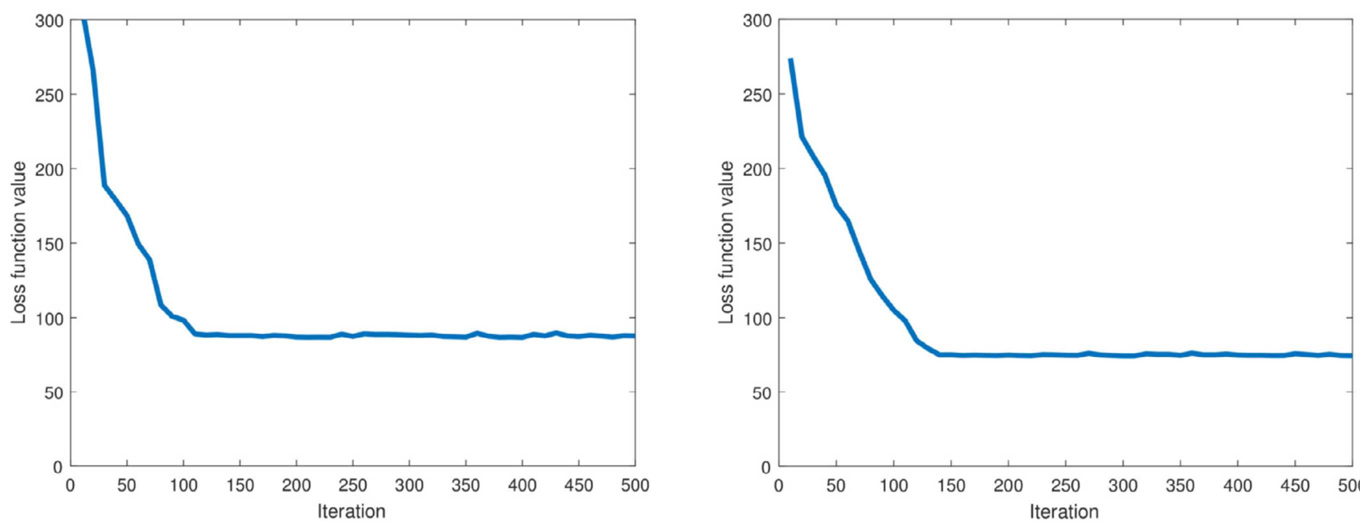
(a) ADNI2-ACC (b) ADNI2-NMI (c) ADNI2-F1 (d) ADNI2-ARI

Fig. 11. Visualization results of the CMC clustering effect with the two trade-off parameters λ_1 and λ_2 on ADNI2 datasets.



(a) ADNI3-ACC (b) ADNI3-NMI (c) ADNI3-F1 (d) ADNI3-ARI

Fig. 12. Visualization results of the CMC clustering effect with the two trade-off parameters λ_1 and λ_2 on ADNI3 datasets.



(a) (b)

Fig. 13. (a) The visualization of the latent representation on the dataset ADNI2. (b) The visualization of the latent representation on the dataset ADNI3. Note that the numbers (1, ..., 6) in the legend denote the six class labels.

with CMC when the subjects of EMCI, MCI and LMCI are integrated into one subject as MCI in practical study. That is, our model is still a promising unsupervised learning method for predicting more detailed AD progression. It provides a theoretical technique for the auxiliary diagnosis of early AD, and has extremely important clinical application value.

6. Conclusions and future work

In this paper, we proposed a novel Consensus Multi-View Clustering model based on Nonnegative Matrix Factorization to group different developmental stages of AD. The CMC integrated the key characteristics of consensus representation, NMF, multi-view clustering techniques. The two real-world datasets were restructured as twelve views data sets for experiments through the SIFT, KAZE and Gabor filter methods. The experimental results demonstrated that our proposed CMC model outperforms the state-of-the-art multi-view clustering and single-view clustering methods.

In this work, we experimented using MRI data. In the future, we will work on collecting more views of other types such as text and voice views, explore the relationship between different modalities and further validate our model on these datasets. In addition, we will improve CMC method with deep learning to achieve better results in terms of predicting the multiple stages in AD progression.

Declaration of Competing Interest

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

Acknowledgments

We would like to thank the Alzheimer's Disease Neuroimaging Initiative (ADNI) platform for the datasets used in our experiments. This work was partially supported by the National Natural Science Foundation of China (Nos. 61976247 and 61572407), and the Key Research and Development Programme in Sichuan Province of China (No. 20ZDYF2837).

References

[1] A. Alberdi, A. Aztiria, A. Basarab, On the early diagnosis of Alzheimer's disease from multimodal signals: a survey, Artif. Intell. Med. 71 (2016) 1–29.
[2] P.F. Alcantarilla, A. Bartoli, A.J. Davison, KAZE features, in: Proceedings of the European Conference on Computer Vision, 2012, pp. 214–227.
[3] S. Arora, R. Ge, R. Kannan, A. Moitra, Computing a nonnegative matrix factorization-provably, in: Proceedings of the Annual ACM Symposium on Theory of Computing, 2012, pp. 145–162.
[4] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, H.M. Arrighi, Forecasting the global burden of Alzheimer's disease, Alzheimer's Dementia 3 (3) (2007) 186–191.
[5] X. Cai, F. Nie, H. Huang, Multi-view K-means clustering on big data, in: Proceedings of the International Joint Conference on Artificial Intelligence, 2013, pp. 2598–2604.

[6] M.A. Ebrahimighahnavieh, S. Luo, R. Chiong, Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review, Comput. Methods Prog. Biomed. 187 (2020) 105242.
[7] V.H. Finder, Alzheimer's disease: a general introduction and pathomechanism, J. Alzheimer's Disease 22 (s3) (2010) S5–S19.
[8] G. Fiscon, E. Weitschek, M.C. De Cola, G. Felici, P. Bertolazzi, An integrated approach based on EEG signals processing combined with supervised methods to classify Alzheimer's disease patients, in: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, 2018, pp. 2750–2752.
[9] B.J. Frey, D. Dueck, Clustering by passing messages between data points, Science 315 (5814) (2007) 972–976.
[10] N. Gessert, T. Sentker, F. Madesta, R. Schmitz, H. Kniep, I. Baltruschat, R. Werner, A. Schlaefer, Skin lesion classification using CNNs with patch-based attention and diagnosis-guided loss weighting, IEEE Trans. Biomed. Eng. 67 (2) (2019) 495–503.
[11] X. Hao, Y. Bao, Y. Guo, M. Yu, D. Zhang, S.L. Risacher, A.J. Saykin, X. Yao, L. Shen, et al., Multi-modal neuroimaging feature selection with consistent metric constraint for diagnosis of Alzheimer's disease, Med. Image Anal. 60 (2020) 101625, doi:10.1016/j.media.2019.101625.
[12] D. Harman, Alzheimer's disease pathogenesis: role of aging, Ann. New York Acad. Sci. 1067 (1) (2006) 454–460.
[13] J. Huang, F. Nie, H. Huang, Robust manifold nonnegative matrix factorization, ACM Trans. Knowl. Discov. Data 83 (2014) 1–21.
[14] L. Huang, Y. Jin, Y. Gao, K.-H. Thung, D. Shen, et al., Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest, Neurobiol. Aging 46 (2016) 180–191.
[15] L. Jing, C. Zhang, M.K. Ng, SNMFCA: supervised NMF-based image classification and annotation, IEEE Trans. Image Process. 21 (11) (2012) 4508–4521.
[16] V.D.M. Laurens, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2605) (2008) 2579–2605.
[17] D.D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, in: Advances in Neural Information Processing Systems, 2001, pp. 556–562.
[18] E. Lee, J.-S. Choi, M. Kim, H.-I. Suk, et al., Toward an interpretable Alzheimer's disease diagnostic model with regional abnormality representation via deep learning, NeuroImage 202 (2019) 116113, doi:10.1016/j.neuroimage.2019.116113.
[19] H.-C. Li, P.-Y. Chen, H.-F. Cheng, Y.-M. Kuo, C.-C. Huang, In vivo visualization of brain vasculature in Alzheimer's disease mice by high-frequency micro–doppler imaging, IEEE Trans. Biomed. Eng. 66 (12) (2019) 3393–3401.
[20] J. Li, Y. Rong, H. Meng, Z. Lu, T. Kwok, H. Cheng, Tatc: predicting Alzheimer's disease with actigraphy data, in: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2018, pp. 509–518.
[21] Q. Li, X. Wu, L. Xu, K. Chen, L. Yao, R. Li, Multi-modal discriminative dictionary learning for Alzheimer's disease and mild cognitive impairment, Comput. Methods Prog. Biomed. 150 (2017) 1–8.
[22] T. Li, C. Ding, Nonnegative matrix factorizations for clustering: asurvey, in: Data Clustering, Chapman and Hall/CRC, 2018, pp. 149–176.
[23] W. Li, X. Lin, X. Chen, Detecting Alzheimer's disease based on 4D fMRI: an exploration under deep learning framework, Neurocomputing 388 (2020) 280–287.
[24] Y. Li, W. Liu, X. Li, Q. Huang, X. Li, GA-SIFT: a new scale invariant feature transform for multispectral image using geometric algebra, Inf. Sci. 281 (2014) 559–572.
[25] N. Lipsman, Y. Meng, A.J. Bethune, Y. Huang, B. Lam, M. Masellis, N. Herrmann, C. Heyn, I. Aubert, A. Boutet, et al., Blood–brain barrier opening in Alzheimer's disease using MR-guided focused ultrasound, Nat. Commun. 9 (1) (2018) 1–8.
[26] J. Liu, C. Wang, J. Gao, J. Han, Multi-view clustering via joint nonnegative matrix factorization, in: Proceedings of the SIAM International Conference on Data Mining, 2013, pp. 252–260.
[27] M. Liu, F. Li, H. Yan, K. Wang, Y. Ma, L. Shen, M. Xu, et al., A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease, NeuroImage 208 (2020) 116459, doi:10.1016/j.neuroimage.2019.116459.
[28] Q. Liu, C. Chen, A. Gao, H.H. Tong, L. Xie, Varifunnet, an integrated multiscale modeling framework to study the effects of rare non-coding variants in genome-wide association studies: applied to Alzheimer's disease, in: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, 2017, pp. 2177–2182.
[29] Y. Lu, C. Yuan, W. Zhu, X. Li, Structurally incoherent low-rank nonnegative matrix factorization for image classification, IEEE Trans. Image Process. 27 (2018) 5248–5260.
[30] L. Luo, X. She, J. Cao, Y. Zhang, Y. Li, P.X. Song, Detection and prediction of ovulation from body temperature measured by an in-ear wearable thermometer, IEEE Trans. Biomed. Eng. 67 (2) (2019) 512–522.
[31] X. Luo, M. Zhou, Y. Xia, Q. Zhu, An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems, IEEE Trans. Ind. Inform. 10 (2) (2014) 1273–1284.
[32] R. Matsuoka, K. Yoshino, E. Watanabe, K. Kiyono, Analysis of multiscale entropy characteristics of heart rate variability in patients with permanent atrial fibrillation for predicting ischemic stroke risk, in: Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine, 2019, p. 2785.
[33] U. Maulik, S. Bandyopadhyay, Performance evaluation of some clustering algorithms and validity indices, IEEE Trans. Pattern Anal. Mach. Intell. 24 (12) (2002) 1650–1654.
[34] D. Modenini, Attitude determination from ellipsoid observations: a modified orthogonal procrustes problem, J. Guid. Control Dynamics 41 (10) (2018) 2324–2326.
[35] S.A. Morris, T.C. Slesnick, Magnetic resonance imaging, Vis. Guide Neonatal Cardiol. 16 (2018) 104–108.
[36] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, D. Shen, Medical image synthesis with deep convolutional adversarial networks, IEEE Trans. Biomed. Eng. 65 (12) (2018) 2720–2730.
[37] F. Nie, G. Cai, X. Li, Multi-view clustering and semi-supervised classification with adaptive neighbours, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2017, pp. 2408–2414.
[38] F. Nie, J. Li, X. Li, Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification, in: Proceedings of the International Joint Conference on Artificial Intelligence, 2016, pp. 1881–1887.
[39] M. Pais, L. Martinez, O. Ribeiro, J. Loureiro, R. Fernandez, L. Valiengo, P. Canineu, F. Stella, L. Talib, M. Radanovic, et al., Early diagnosis and treatment of Alzheimer's disease: new definitions and challenges, Braz. J. Psychiatr. (2020) 1–11, doi:10.1590/1516-4446-2019-0735.
[40] X. Pan, M. Adel, C. Fossati, T. Gaidon, J. Wojak, E. Guedj, A.D.N. Initiative, et al., Multiscale spatial gradient features for 18F-FDG pet image-guided diagnosis of Alzheimer's disease, Comput. Methods Prog. Biomed. 180 (2019) 105027.
[41] F. Previtali, P. Bertolazzi, G. Felici, E. Weitschek, A novel method and software for automatically classifying Alzheimer's disease patients by magnetic resonance imaging analysis, Comput. Methods Prog. Biomed. 143 (2017) 89–95.
[42] B. Richhariya, M. Tanveer, Least squares projection twin support vector clustering (LSPTSVC), Inf. Sci. 533 (2020) 1–23.
[43] A. Salah, M. Ailem, M. Nadif, A way to boost semi-NMF for document clustering, in: Proceedings of the ACM Conference on Information and Knowledge Management, 2017, pp. 2275–2278.
[44] D. Steinley, K-means clustering: a half-century synthesis, Br. J. Math. Stat. Psychol. 59 (1) (2006) 1–34.
[45] R.H. Swerdlow, Pathogenesis of Alzheimer's disease, Clin. Interv. Aging 2 (3) (2007) 347.
[46] A. Tharwat, Principal component analysis: an overview, Pattern Recognit. 3 (3) (2016) 197–240.
[47] L. Tong, W.K. Wong, C. Kwong, Differential evolution-based optimal Gabor filter model for fabric inspection, Neurocomputing 173 (2016) 1386–1401.
[48] T. Tong, K. Gray, Q. Gao, L. Chen, D. Rueckert, et al., Multi-modal classification of Alzheimer's disease using nonlinear graph fusion, Pattern Recognit. 63 (2017) 171–181.
[49] K. Vaithinathan, L. Parthiban, et al., A novel texture extraction technique with T1 weighted MRI for the classification of Alzheimer's disease, J. Neurosci. Methods 318 (2019) 84–99.
[50] S.A. Vavasis, On the complexity of nonnegative matrix factorization, SIAM J. Optim. 20 (2009) 1364–1377.
[51] H. Wang, Y. Yang, T. Li, Multi-view clustering via concept factorization with local manifold regularization, in: 2016 IEEE 16th International Conference on Data Mining (ICDM), IEEE, 2016, pp. 1245–1250.
[52] H. Wang, Y. Yang, B. Liu, GMC: graph-based multi-view clustering, IEEE Trans. Knowl. Data Eng. 32 (6) (2019) 1116–1129.
[53] L. Wang, Y. Liu, X. Zeng, H. Cheng, Z. Wang, Q. Wang, Region-of-interest based sparse feature learning method for Alzheimer's disease identification, Comput. Methods Prog. Biomed. 187 (2020) 105290.
[54] M.W. Weiner, D.P. Veitch, P.S. Aisen, L.A. Beckett, N.J. Cairns, R.C. Green, D. Harvey, C.R. Jack Jr, W. Jagust, J.C. Morris, et al., Recent publications from the Alzheimer's disease neuroimaging initiative: reviewing progress toward improved AD clinical trials, Alzheimer's Dementia 13 (4) (2017) e1–e85.
[55] J. Wen, Z. Zhang, Z. Zhang, L. Fei, M. Wang, Generalized incomplete multiview clustering with flexible locality structure diffusion, IEEE Trans. Cybern. (2020).
[56] R. Xia, Y. Pan, L. Du, J. Yin, Robust multi-view spectral clustering via low-rank and sparse decomposition, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2014, pp. 2149–2155.
[57] T. Xia, D. Tao, T. Mei, Y. Zhang, Multiview spectral embedding, IEEE Trans. Syst. Man. Cybern. Part B (Cybernetics) 40 (6) (2010) 1438–1446.
[58] Y. Yang, H. Wang, Multi-view clustering: a survey, Big Data Min. Anal. 1 (2) (2018) 83–107.
[59] C. Zhang, E. Adeli, T. Zhou, X. Chen, D. Shen, Multi-layer multi-view classification for Alzheimer's disease diagnosis, in: Proceedings of of the AAAI Conference on Artificial Intelligence, 2018, pp. 4406–4413.
[60] F. Zhang, Z. Li, B. Zhang, H. Du, B. Wang, X. Zhang, Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease, Neurocomputing 361 (2019) 185–195.
[61] H. Zhang, X. Huo, X. Guo, X. Su, X. Quan, C. Jin, A disease-related gene mining method based on weakly supervised learning model, BMC Bioinform. 20 (16) (2019) 1–11.
[62] X. Zhang, Y. Yang, T. Li, H. Wang, Z. He, Discovering senile dementia from brain MRI using Ra-DenseNet, in: Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2019, pp. 449–460.
[63] Y. Zhang, H. Zhang, X. Chen, M. Liu, X. Zhu, S.-W. Lee, D. Shen, Strength and similarity guided group-level brain functional network construction for MCI diagnosis, Pattern Recognit. 88 (2019) 421–430.
[64] Z. Zhang, L. Liu, F. Shen, H.T. Shen, L. Shao, Binary multi-view clustering, IEEE Trans. Pattern Anal. Mach. Intell. 41 (7) (2018) 1774–1782.
[65] Z. Zhang, Q. Zhu, G.-S. Xie, Y. Chen, Z. Li, S. Wang, Discriminative margin-sensitive autoencoder for collective multi-view disease analysis, Neural Netw. 123 (2020) 94–107.