A Novel Double-Index-Constrained, Multi-View, Fuzzy-Clustering Algorithm and its Application for Detecting Epilepsy Electroencephalogram Signals

Jiaqi Zhu1,2,7, Kang Li1,2,7, Kaijian Xia3, Xiaqing Gu1, Jing Xue4, Shi Qiu4, Yizhang Jiang1,2,7, (Member, IEEE), Pengjiang Qian1,2,5, (Member, IEEE)

1 School of Digital Media, Jiangnan University, 1800 Lihu Avenue, Wuxi, Jiangsu 214122, P.R. China
2 Jiangsu Key Laboratory of Media Design and Software Technology, 1800 Lihu Avenue, Wuxi, Jiangsu 214122, P.R. China
3 Changshu No.1 People's Hospital, Changshu, Jiangsu 215500, P.R. China
4 School of Information Science and Engineering, Changzhou University, Jiangsu, Changzhou, China
5 Department of Nephrology, the Affiliated Wuxi People’s Hospital of Nanjing Medical University, 299 Qingyang Rd, Wuxi, Jiangsu 214023, P.R. China
6 Key Laboratory of Spectral Imaging Technology CAS, Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an, 710119, P.R. China
7 Jiangsu Jiangda Smart Science and Technology Ltd., No. 18 Shannan Road, Chengbei Park, Jingjiang, Jiangsu 214500, P.R. China

Corresponding author: Yizhang Jiang (E-mail: yzjiang@jiangnan.edu.cn)

ABSTRACT. When processing a multi-view, epilepsy electroencephalogram (EEG) dataset, the traditional single-view clustering algorithms cannot fully mine the correlation information between each view and identify the importance of each view because of the limitations of its own methods. This limitation causes poor clustering performance when using these classic, single-view clustering algorithms. To solve this problem, a novel double-index-constrained, multi-view, fuzzy clustering algorithm (DIC-MV-FCM) is proposed for the automatic detection of epilepsy EEG data. The DIC-MV-FCM algorithm is integrated into the multi-view clustering technology and the view-weighted adaptive learning strategy, which can effectively use the correlation information between each view and control the importance of each view to improve the final clustering performance. The experimental results using several epilepsy EEG datasets show that the proposed DIC-MV-FCM algorithm has better clustering performance than the traditional clustering algorithms for processing multi-view EEG data.

INDEX TERMS: epileptic detecting; multi-view clustering; double-index-constrained fuzzy clustering algorithm

I. INTRODUCTION

Epilepsy is a common and frequent brain disease with transient and repetitive disorders of the central nervous system due to excessive discharge of brain lesions [1]. Electroencephalogram (EEG) signals refer to electrical activity signals inside the brain. When the human body is excited, sleeping, or receives signal stimulation, the EEG signals will show various patterns. Therefore, the use of EEG signals is an important way to analyze brain function. An epilepsy signal is medically defined as being distinctly different from a normal signal. Clinicians usually assess whether there are sharp waves and spikes to determine whether an EEG signal is an epilepsy signal [2]-[8]. In this study, we aim to construct an automated, unsupervised clustering technology for smart recognition of an epilepsy patient. The traditional method of analyzing epilepsy EEG signals is based on single-view learning technology, which
cannot fully mine the correlation information between each view and identify the importance of each view. In [28], Jiang et al found that multi-view learning technology performs better in recognizing epilepsy than single-view learning technology. Thus, we intend to continue exploring how to use multi-view technology to analyze EEG data. In this study, we will propose an unsupervised, multi-view, clustering technology to analyze EEG data instead of the traditional, supervised classification methods [2-8, 28]. As we know, labeling medical data is very time-consuming and expensive. An unsupervised clustering technology does not need data to be labeled when using it to detect EEG data. Thus, it is more appropriate to detect EEG data using an unsupervised clustering technology than a supervised classification method.

Traditional clustering methods, such as fuzzy c-means (FCM)[9],[10], k-means[11,12], possibilistic c-means (PCM)[13], and maximum entropy clustering (MEC)[14,15], all use single-view clustering technology. When handling a multi-view clustering task using one of these single-view clustering algorithms, one common approach is to process the data from each view independently. After the corresponding cluster analysis structure for each view is obtained, an ensemble learning strategy [16,17] is used to integrate the clustering results of each view to obtain the final clustering result. However, we have found this single-view, ensemble learning strategy does not mine the correlation information between each view, and the final recognition performance of these single-view learning-based algorithms is worse than that of a single algorithm that uses multi-view learning technology [28].

In this study, we also found that the traditional clustering method integrated with multi-view learning technology enables each view to achieve collaborative learning in the process of clustering [18-20], which is an effective solution for analyzing multi-view EEG data. Although multi-view clustering technology, especially the multi-view collaborative fuzzy clustering algorithm (Co-FKM) [21] has only been developed recently, it is a very effective multi-view clustering method based on the classic FCM framework. However, this study of the multi-view clustering algorithm found that the existing multi-view clustering algorithms often fail to take into account the importance of each view.

To solve the above problem, we proposed a novel view-weighting mechanism so that each view can effectively adjust the weight coefficient in the clustering process. When this novel view-weighting mechanism is used in our proposed multi-view clustering algorithm, it assigns an appropriate view-weight to each view, according to its clustering performance. Thus, we can obtain a multi-view clustering result using a multi-view, ensemble learning strategy based on the weight of each view.

To achieve this objective, we use a single-view, double-index-constrained, fuzzy clustering algorithm (DI-FCM) and a view-weighting mechanism to develop a novel double-index-constrained, multi-view, fuzzy clustering algorithm (DIC-MV-FCM) for the automated detection of epilepsy EEG data. The proposed DIC-MV-FCM algorithm can not only mine the correlation information between each view but also automatically evaluate the importance of each view for good clustering performance. Compared to traditional single-view clustering algorithms, the main contributions of the proposed DIC-MV-FCM algorithm can be summarized as follows:

1) An unsupervised multi-view clustering technology that does not use labeled data is proposed to analyze EEG data instead of the traditional supervised classification methods.

2) A novel view-weighting mechanism is proposed to automatically evaluate the importance of each view for good clustering performance.

3) A double exponential collaborative learning strategy is used to mine the correlation information between each view.

This article is structured as follows. In section II, we briefly review some related and common EEG data processing methods. In section III, we introduce a single-view, double-index-constrained, fuzzy clustering algorithm (DI-FCM). In section IV, a classical multi-view, collaborative fuzzy clustering algorithm (Co-FKM) [21] is briefly reviewed and analyzed. In section V, we propose a novel double-index-constrained, multi-view, fuzzy clustering algorithm (DIC-MV-FCM) for the automated detection of epilepsy EEG data. In Section VI, the results of the experiments using several epilepsy EEG datasets show that the proposed DIC-MV-FCM algorithm has better clustering performance than the traditional clustering algorithms for processing the multi-view EEG data. Section VII concludes the paper.

II. EEG DATA

This section will briefly introduce the original epilepsy EEG signals collected and how to process these original EEG data [22,28,29]. In addition, we will describe how to construct multi-view EEG data.

A. The Original Epilepsy EEG Signal Data

According to the data distribution characteristics, the epilepsy EEG signal data can be subdivided into five different types. Group A denotes the EEG signal data measured from healthy volunteers when they keep their eyes open. Group B denotes the EEG signal data measured from healthy volunteers when they close their eyes. Group C denotes the EEG signal data in the hippocampus of an epilepsy patient’s brain during epileptic seizures. Group D denotes the EEG signal data in the epileptogenic zone of a patient’s brain during epileptic seizures. Group E denotes the EEG signal data during epileptic seizures. A diagram of the five types of EEG signal data is shown in FIGURE 1.
In the analysis of EEG signals, the short-time Fourier transform (STFT) is often used due to its simplicity and effectiveness. The STFT method can be described as:

$$F_{STFT}(t, f) = \int_{-\infty}^{\infty} x(t) g^*(t-u)e^{-j2\pi ft} dt.$$ 

Here, $F_{STFT}(t, f)$ is the transform function that maps the original EEG signal data into the time-frequency space.

3) KERNEl PRINCIPAL COMPONENT ANALYSIS (KPCA)

If the original EEG signal data is nonlinear, then the kernel principal component analysis (KPCA) [25] is a very typical and successful feature extraction method for extracting the effectiveness features from the original EEG signal data. The KPCA can be used to map the original EEG signal data into a high-dimensional, linear-separable feature space. An example of the extracted features using the KPCA feature extraction method is shown in FIGURE 2.

B. The related feature extraction methods for the original EEG signal data

EEG signals data are often very random; the data are nonstationary signals, and there are many influencing factors. It is difficult to recognize the laws that these signals obey. The usual processing method is to extract the effectiveness feature of EEG signal data by using a quantitative analysis. After several experiments, we found that the time-frequency analysis method is very useful for extracting the effectiveness features of EEG signal data [22,28,29]. The following is a brief introduction to classic feature extraction methods for EEG signal data.

1) WAVELET PACKET DECOMPOSITION (WPD)

In recent years, a variety of wavelet transform methods [23] have been applied to the analysis of epilepsy EEG signal data. Because the wavelet transform methods have better time-frequency localization, they are very suitable for the analysis of these EEG signal data with the time-varying and nonstationary characteristics. Moreover, it is not difficult to see from the waveform of an epilepsy EEG signal that the distribution range of the characteristic wave used to identify epilepsy includes a sharp wave of 5 Hz~12.5 Hz; a spike wave of 13.5 Hz~50 Hz; and a slow wave of 1 Hz~2.5 Hz. Thus, we can use the typical wavelet transform method to process the original EEG signal data according to these three frequency bands.

2) SHORT TIME FOURIER TRANSFORM (STFT)

The short-time Fourier transform [24] is also a feature extraction method that is often used for time-frequency analysis. This method is also widely used in EEG signal analysis due to its simplicity and effectiveness. In the feature extraction of the initial EEG signal data, the data are first divided into different time slices by using the moving time window function; then, the Fourier transform is performed. For a given continuous EEG signal data, the STFT method can be described as:

$$F_{STFT}(t, f) = \int_{-\infty}^{\infty} x(t) g^*(t-u)e^{-j2\pi ft} dt.$$
\[ J_{m,r}(U, V) = \sum_{k=1}^{m} \sum_{i=1}^{n} \mu_{ik}^m \| x_i - v_k \|^2, \quad m > 0 \]  

where \( U = [\mu_{ik}]_{m \times n} \) is the fuzzy partition matrix, \( V = \{v_1, v_2, \ldots, v_m\} \) is the center set, \( v_i \) is the center of the \( i \) th class. \( m > 0 \) is a fuzzy indicator, \( r > 0 \) is a power index of the constraint condition, and for any integer \( c \), there is \( 2 \leq c \leq n \). The element \( \mu_{ik} \) satisfies the following three constraints:

\[ \mu_{ik} \in [0,1], k = 1,2,\ldots,n, i = 1,2,\ldots,c \tag{2} \]

\[ \sum_{i=1}^{c} \mu_{ik}^m = 1, m > r > 0, k = 1,2,\ldots,n \tag{3} \]

\[ 0 < \sum_{k=1}^{m} \mu_{ik} < n, i = 1,2,\ldots,c \tag{4} \]

According to the Lagrange multiplier, it is easy to obtain the necessary conditions for \( J_{m,r} \) to obtain the local minimum under the constraints (2)~(4) as follows:

\[ v_i = \frac{\sum_{j=1}^{n} \mu_{ij}^m x_j}{\sum_{j=1}^{n} \mu_{ij}^m}, m > 0, i = 1,2,\ldots,c \tag{5} \]

\[ \mu_{ik} = \left( \frac{\| x_k - v_i \|^2}{\sum_{j=1}^{c} \| x_k - v_j \|^2} \right)^{\frac{1}{r}}, \quad \text{s.t.} \quad m > r > 0, i = 1,2,\ldots,c, k = 1,2,\ldots,n \tag{6} \]

Obviously, when \( r = 1 \), the DI-FCM algorithm degenerates into the FCM algorithm. When \( r \neq 1 \), according to the relevant theory of information theory, the \( r \) th order \( \beta \) entropy of the \( k \) th sample in the \( x \) data set is:

\[ H_{\beta}(x_k) = \frac{1}{1-2^{-r}} \left( 1 - \sum_{i=1}^{c} \mu_{ik}^m \right), r > 0 \]  

Obviously, when \( r = 1 \) and the constraint term in Eq.(3) is satisfied, the value of \( H_{\beta}(x_k) \) is 0, which means that the sample \( x_k \) in the data set \( X \) is subordinate to each class with the least uncertainty. Therefore, the essence of the DI-FCM algorithm is to solve for the minimum value of the objective function \( J_{m,r} \) when the \( r \) th order \( \beta \) entropy of the fuzzy membership \( \mu_{ik} \) of each sample \( x_k \) in the data set \( X \) is zero.

With the above Lagrange multiplier optimization strategy, we can finally obtain the fuzzy partition matrix \( U \). As we know, the DI-FCM is still a single-view clustering algorithm, and it only can run on a single-view dataset. If we want to use DI-FCM to process a multi-view clustering task, one effective way is to divide the multi-view dataset into several single-view datasets, and then use the DI-FCM algorithm (Eq.(1)) to cluster these single-view datasets. Finally, we can use an ensemble learning strategy to integrate the \([U_1,\ldots,U_N]\) of each view into a solution \( \hat{U} \) with a global description ability. The working principle of this method is shown in FIGURE 3. After performing this single-view processing of the working principle of the clustering task of multi-view data, it is not difficult to find that this clustering technology cannot fully mine the correlation information between each view and identify the importance of each view. In addition, even though the ensemble learning strategy was introduced at the end to obtain a global clustering result, the interaction between each view (such as the similarity) was not used well because the interactive learning between each view was neglected. This ensemble learning strategy will also cause the final clustering performance of DI-FCM to suffer from the impact of a certain view with a bad clustering result.

IV. MULTI-VIEW CO-FKM ALGORITHM

To solve the difficulties encountered by the single-view algorithm in dealing with a multi-view clustering task, a multi-view collaborative fuzzy clustering algorithm (Co-FKM) [21] is proposed based on the classical FCM algorithm. This algorithm first introduces the following collaborative membership constraint item of each view:

\[ \Delta_k = \eta \sum_{k=1}^{K} \sum_{i=1}^{c} \left( \mu_{ik}^m - \mu_{ik}^* \right) \| x_k - v_i \|^2 \]  

Then, we can obtain the objective function of the Co-FKM algorithm as follows:

\[ J_{Co-FKM}(U, V) = \sum_{k=1}^{K} \sum_{i=1}^{c} \left[ \mu_{ik}^m \| x_k - v_i \|^2 + \eta \Delta_k \right] \tag{8} \]
\[ \Delta_i = \eta \frac{1}{K-1} \sum_k \sum_{i=1}^K \sum_{j=1}^N (\mu_{ij,k}^m - \mu_{ij,k}^m) \| x_{j,k} - v_{i,k} \|^2 \] . \quad (9)

\text{st.} \quad \mu_{ij,k} \in [0,1] \quad \text{and} \quad \sum_{i=1}^C \sum_{j=1}^N \mu_{ij,k} = 1 \quad 1 \leq j \leq N \quad 1 \leq k \leq K

Substituting Eq. (9) into Eq. (8), and after simplification, the following objective function can be expressed as:

\[ J_{Co-FKM} (U, V) = \sum_k \sum_{i=1}^K \sum_{j=1}^N \left[ \mu_{ij,k}^m \| x_{j,k} - v_{i,k} \|^2 \right] . \quad (10) \]

where \( \mu_{ij,k}^m = (1-\eta)\mu_{ij,k}^m + \eta \sum_{k=1}^K \mu_{ij,k}^m \); the parameter \( \eta \) is used to adjust the weighting coefficient of the division of membership of each view, and \( \hat{\mu}_{ij,k,\eta} \) represents the weighted average between the fuzzy membership \( \mu_{ij,k}^m \) at the current view and the fuzzy membership \( \mu_{ij,k}^m \) of other views.

The iterative function of fuzzy membership \( \mu_{ij,k}^m \) and the center \( v_{i,k} \) can also be obtained by the Lagrange multiplier as follows:

\[ v_{i,k} = \frac{\sum_{j=1}^N \sum_{i=1}^C \mu_{ij,k,\eta} x_{j,k}}{\sum_{j=1}^N \sum_{i=1}^C \mu_{ij,k,\eta}} \quad i = 1, 2, \ldots, C \] \quad (11)

\[ \mu_{ij,k} = \frac{1}{\sum_{j=1}^N \sum_{i=1}^C \mu_{ij,k,\eta} \left[ (1-\eta)d_{ij,k}^2 + \eta \sum_{k=1}^K \sum_{i=1}^C d_{ij,k}^2 \right]^{\frac{1}{m-1}}} \]

\[ \left[ (1-\eta)d_{ij,k}^2 + \eta \sum_{k=1}^K \sum_{i=1}^C d_{ij,k}^2 \right] \quad i = 1, 2, \ldots, C; \quad j = 1, 2, \ldots, N; \quad k = 1, 2, \ldots, K . \] \quad (12)

Based on the above iterative strategy, the corresponding fuzzy membership partitioning matrix of each view can be obtained. To obtain the global decision, the algorithm utilizes the final result of the geometric average partitioning of the fuzzy membership obtained from each view:

\[ \hat{\mu}_{ij} = K \prod_{k=1}^K \mu_{ij,k} . \]

The clustering results that express the whole thing or object can be obtained after defuzzification. The specific working mechanisms of the multi-view fuzzy clustering algorithm is shown in FIGURE 4.

Although this algorithm implements interactive learning of each view when dealing with multi-view clustering tasks and has better clustering performance than the previous single-view ensemble learning strategy, there are still some improvements to be made, such as the disadvantages we outlined in the introduction. For this purpose, a new multi-view clustering algorithm is proposed in the following section.

V. DIC-MV-FCM ALGORITHM

To address the shortcomings in the processing of multi-view data with the single-view method proposed in the introduction, this paper first adopts a multi-view clustering algorithm based on the DI-FCM algorithm, so the data of each view can be learned collaboratively. When the algorithm converges, the results of the spatial division of each view are as similar as possible, so each view can be interconnected and interactively learned in the multi-view clustering task. Thus, better clustering performance can be obtained than when using the single-view clustering algorithm. In addition, when using most of the classical single-view algorithms or the multi-view algorithms, it should always be assumed that the importance of each view is equal when clustering multi-view data. In fact, the importance of each view is different; for example, there is indivisibility in some views because of spatial overlaps. Obviously, it is unreasonable to give these views the same degree of importance as the views that have good segmentation in the whole clustering process. This study considers that the relationship between each view should be complementary and that each view should be addressed differently. Here, we propose a novel view-weighting mechanism to solve the problem of giving the same weight to each view, thereby causing the decreasing clustering effect. Finally, we also adopted a double exponential collaborative learning strategy to effectively extend the value range of the fuzzy indicator \( m \), and extend the value of the original fuzzy indicator \( m \) from the original \( m > 1 \) to \( m > r > 1 \), in order to mine the correlation information between each view.

Combined with these two multi-view learning strategies (the view-weighting mechanism and the double exponential collaborative learning strategy), a novel double-index-constrained, multi-view, fuzzy clustering algorithm (DIC-MV-FCM) is proposed for the automated detection of epilepsy EEG data. The objective function of DIC-MV-FCM algorithm is reconstructed in the framework of the DI-FCM algorithm as follows:

\[ J(U,V,W) = \sum_{k=1}^K w_{ik}^m \sum_{j=1}^C \sum_{i=1}^N \mu_{ij,k}^m \| x_{j,k} - v_{i,k} \|^2 , \quad (13) \]
where \( \mathbf{U}_k = [u_{ij,k}]_{N \times K} \) is the fuzzy membership partitioning matrix, and the element \( \mu_{ij,k} \) is the fuzzy membership of the \( j \) th sample data in the \( i \) th category in the \( k \) th view. \( \mathbf{W} = [w^m_{k}]_{i \times K} \) represents the view weight matrix, element \( w^m_{k} \) represents the weight of the \( k \) th view, and they satisfy the constraints as \( \mu_{ij,k} \in [0,1], w^i_k \in [0,1] \), \( \sum_{k=1}^{K} \mu^2_{ij,k} = 1 \) and \( \sum_{i=1}^{N} \mu^2_{ij,k} = 1 \).

The constraints as \( \mu_{ij,k} \in [0,1], w^i_k \in [0,1] \), \( \sum_{k=1}^{K} \mu^2_{ij,k} = 1 \) and \( \sum_{i=1}^{N} \mu^2_{ij,k} = 1 \).

Of course, in order to obtain the clustering result with the final global decision, this study defines a new multi-view ensemble learning strategy based on the resulting view-weighted matrix \( \mathbf{W} = [w^m_{k}]_{i \times K} \); its specific form is as follows:

\[
\hat{\mathbf{U}} = \sum_{k=1}^{K} w^m_{k} \mathbf{U}_k
\]

In this paper, the classical optimization theory is combined with the Lagrange multiplier to optimize the objective function Eq.(13). To obtain the optimal solution of the objective function, the following theorems are given:

**Theorem 1**: When the values of the fuzzy membership partitioning matrix \( \mathbf{U}_k \) of the \( k \) th view and the weight value \( w^m_{k} \) of the \( k \) th view are given, the objective function Eq.(13) needs to satisfy the following condition to obtain the extremum:

\[
\nabla \mathbf{J}(\mathbf{V}_k) = \sum_{k=1}^{K} \sum_{i=1}^{N} \mu_{ij,k} x_j \nabla \mathbf{J}(\mathbf{V}_k) = \sum_{k=1}^{K} \sum_{i=1}^{N} \mu_{ij,k} x_j \frac{\partial \mathbf{J}}{\partial \mathbf{V}_k} = 0
\]

**Proof**: Using the obtained fuzzy membership partitioning matrix \( \mathbf{U}_k \) of the given \( k \) th view and the weight \( w^m_{k} \) of the \( k \) th view, the partial derivative is obtained for the objective function \( \mathbf{J}(\mathbf{V}_k) \), and letting \( \frac{\partial \mathbf{J}}{\partial \mathbf{V}_k} = 0 \), then Theorem 1 can be proved.

**Theorem 2**: When the center matrix \( \mathbf{V}_k \) of the \( k \) th view and the weight value \( w^m_{k} \) of the \( k \) th view are given, the objective function (13) needs to satisfy the following necessary condition to obtain the extremum:

\[
\mu_{ij,k} = \left( \frac{1}{\sum_{k=1}^{K} w^{m}_{k} \left\| x_j - v_{ij,k} \right\|^{2m_{ij,k}} + \lambda} \right)^{\frac{1}{m_{ij,k}}}
\]

**Proof**: Using the obtained cluster central matrix \( \mathbf{V}_k \) of the \( k \) th view and the weight \( w^m_{k} \) of the \( k \) th view, according to the

constraint condition \( \sum_{k=1}^{K} \mu^2_{ij,k} = 1 \), the corresponding optimization objective function is obtained by using Lagrange multiplier:

\[
\mathbf{J}(\mu_{ij,k}, \lambda) = \sum_{k=1}^{K} w^m_{k} \sum_{i=1}^{N} \sum_{j=1}^{K} \mu_{ij,k} \left\| x_j - v_{ij,k} \right\|^2 + \lambda \left( \sum_{k=1}^{K} \mu^2_{ij,k} - 1 \right)
\]

Calculate the partial derivative for \( \mu_{ij,k} \) and \( \lambda \), and let the partial derivative be zero. According to \( \frac{\partial \mathbf{J}}{\partial \mu_{ij,k}} = 0 \), we conclude that:

\[
\mu_{ij,k} = \left( -\frac{\lambda r_{ij,k}^{1 - \frac{1}{m_{ij,k}}} \left\| x_j - v_{ij,k} \right\|^2}{m_{ij,k} \sum_{k=1}^{K} w^{m}_{k} \left\| x_j - v_{ij,k} \right\|^{2m_{ij,k}}} \right)^{\frac{1}{m_{ij,k}}}
\]

Applying the constraint condition \( \sum_{k=1}^{K} \mu^2_{ij,k} = 1 \), we conclude that:

\[
\mu_{ij,k} = \left( \frac{1}{\sum_{k=1}^{K} \left( \sum_{i=1}^{N} \sum_{j=1}^{K} \mu_{ij,k} \left\| x_j - v_{ij,k} \right\|^{2m_{ij,k}} \right)^{\frac{1}{m_{ij,k}}}} \right)^{\frac{1}{m_{ij,k}}}
\]

**Theorem 3**: When the cluster central matrix \( \mathbf{V}_k \) of \( k \) th view and the values of the fuzzy membership partitioning matrix \( \mathbf{U}_k \) of \( k \) th view are given, the objective function (13) should satisfy the following necessary condition to obtain the extremum:

\[
\mathbf{w}_k = \left( \frac{1}{\sum_{k=1}^{K} \left( \sum_{i=1}^{N} \sum_{j=1}^{K} \mu_{ij,k} \left\| x_j - v_{ij,k} \right\|^{2m_{ij,k}} \right)^{\frac{1}{m_{ij,k}}}} \right)^{\frac{1}{m_{ij,k}}}
\]

**Proof**: According to the proof for Theorem 2, Theorem 3 can be similar obtained by the Lagrange multiplier and the constraint condition \( \sum_{k=1}^{K} w^m_{k} = 1 \).

According to this related optimization theory and the parameter learning rules derived from the iterative formula, the specific steps of this DIC-MV-FCM algorithm are as follows:
DIC-MV-FCM algorithm

**INPUT:** There are $K$ views ($1 \leq k \leq K$) in the multi-view data sample $\text{View} = \{\text{View}_1, \ldots, \text{View}_K\}$, the sample set corresponding to each view is $\text{view}_k = \{x_1, \ldots, x_n\}$, number of categories of cluster $C(2 \leq C < N)$, iterative threshold $\varepsilon$, weight index of the view $m_1$, fuzzy index $m_2$, power index of the constraint condition $t_1, t_2$, number of iterations $f$, maximum number of iterations $L$. 

**OUTPUT:** The weight of each view $w_k$, the global fuzzy partition matrix $U$ and the cluster central point of each view.

- **Step 1:** Initialize a randomly generated central point set $v_{ik}$, randomly generate a normalized fuzzy membership $\mu_{ik}$, and randomly generate a normalized view weight $w_k$.
- **Step 2:** Update the cluster central point $v_{ik}$ under each view according to the Eq. (15);
- **Step 3:** Update the fuzzy membership of each view $\mu_{ik}$ according to Eq. (16);
- **Step 4:** Update the weight of each view $w_k$ according to Eq. (20);
- **Step 5:** if $|g_{ik} - f_{ik}| \leq \varepsilon$ or $f > L$, then the algorithm ends, otherwise, jump back to **Step 2**;
- **Step 6:** After the algorithm converges, output the cluster center $V$ of each view, the fuzzy membership partitioning matrix $U$ and the weights $W$ of all views;
- **Step 7:** According to the weights $W$ and the fuzzy membership partitioning matrix of each view, we finally obtain the global fuzzy membership partitioning matrix $U$ by using the results in **Step 6** and the Eq.(14). In addition, the best clustering view can also be determined by using $\max(w_k)$.

The specific working mechanism of our proposed DIC-MV-FCM algorithm is shown in **FIGURE 5**.

![FIGURE 5. The principle of processing the multi-view clustering task by using our proposed DIC-MV-FCM algorithm](image)

**VI. EXPERIMENTAL STUDY**

To verify the effectiveness of the proposed algorithm for dealing with multi-view clustering tasks, the DIC-MV-FCM was evaluated using real-world EEG data. In this section, the multi-view EEG datasets that were built and the experiment settings are described first. Then, the clustering performance of the DIC-MV-FCM is compared with the performance of four algorithms and discussed.

**A. Datasets and Experimental Settings**

1) Datasets

The original EEG data used in the experiment were downloaded from the website [http://epileptologiebonn.de/cms/front_content.php?idcat=193&lang=3&changelange=3](http://epileptologiebonn.de/cms/front_content.php?idcat=193&lang=3&changelange=3). The data contain five groups of EEG samples (Groups A-E); each group contains 100 single-channel EEG segments of 23.6 s duration. The sampling rate of all the EEG singles is 173.6 Hz. **TABLE 1** shows a detailed description of the five groups of EEG signals. The EEG signals of Groups A and B are from healthy subjects; the signals for Groups C, D and E are from epilepsy patients. In this study, we extract features of EEG signals using three feature extraction methods: wavelet packet decomposition (WPD), short-time Fourier transform (STFT), and kernel principal component analysis (KPCA). Then, we set different combinations of multiple feature extractions as a view and construct eight feature-level, multi-view EEG datasets, as shown in **TABLE 2**.

2) Comparison Algorithms

Four algorithms are used in our experiments for comparison purposes: the fuzzy C-means (FCM), the K-means algorithm, the co-clustering algorithm based on feature and sample space co-clustering, and the multi-view fuzzy clustering Co-FKM [21]. The optimal parameters for all the algorithms are selected by the five cross-sectional and grid search strategies: (a) the number of clusters is set according to the composition of the datasets, as shown in **TABLE 2**; and (b) the fuzzy index for all fuzzy clustering algorithms is selected within the grid $\{0.5, 1, 1.5, 2, 2.5\}$.

3) Performance index

Two evaluation indicators are used to evaluate the performance of all algorithms: normalized mutual information (NMI) and the rand index (RI). NMI is defined as follows:

$$NMI = \frac{\sum_{i=1}^{C} \sum_{j=1}^{C} N_{ij} \log N \cdot N_{ij} / N \cdot N_{ij}}{\sqrt{\sum_{i=1}^{C} N_{i} \log N_{i} / N \cdot \sum_{j=1}^{C} N_{j} \log N_{j} / N}}$$

(21)

where $N_{ij}$ represents the number of agreements between the $i$th and $j$th clusters, $N_i$ and $N_j$ represent the number of samples in the $i$th and $j$th clusters, respectively, and $N$ represents the size of the training data.

RI is defined as follows:

$$RI = \frac{f_{00} + f_{11}}{N(N-1)/2}$$

(22)

where $f_{00}$ represents the number of samples that belong to different clusters but have different cluster labels, and $f_{11}$ represents the number of matching pairs of samples that have the same cluster label and belong to the same cluster.

NMI and RI can efficiently evaluate the degree of agreement between the known clusters and the estimated data structure. Both NMI and RI have a value within the interval $[0, 1]$, and a larger value means better cluster performance.
4) Experimental environment
The hardware platform for the experiment is an Intel Core i5-4590 CPU, with clock speed of 3.30 GHz, and 4 GB memory. The programming environment is MATLAB R2018a.

### TABLE 1. Description of the original EEG data

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Groups</th>
<th>Size of groups</th>
<th>Descriptions of datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>A</td>
<td>100</td>
<td>EEG signals measured from healthy people with eyes open.</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>100</td>
<td>EEG signals measured from healthy people with eyes closed.</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>C</td>
<td>100</td>
<td>EEG signals obtained in the hippocampal formation of the opposite hemisphere of the brain during seizure-free intervals.</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>100</td>
<td>EEG signals obtained from within the epileptogenic zone during seizure-free intervals.</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>100</td>
<td>EEG signals measured during seizures.</td>
</tr>
</tbody>
</table>

### TABLE 2. Details of the eight multi-view EEG datasets used

<table>
<thead>
<tr>
<th>Datasets</th>
<th>View</th>
<th>Training dataset</th>
<th>Size</th>
<th>Dimension</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>WPD</td>
<td>Groups A and B</td>
<td>200</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>STFT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KPCA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td>STFT</td>
<td>Groups C and D</td>
<td>200</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>D3</td>
<td>STFT</td>
<td>Groups D and E</td>
<td>200</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>D4</td>
<td>STFT</td>
<td>Groups C and E</td>
<td>200</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>D5</td>
<td>STFT</td>
<td>Groups C, D and E</td>
<td>300</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>D6</td>
<td>STFT</td>
<td>Groups A, B, C</td>
<td>300</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>KPCA</td>
<td>and D</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D7</td>
<td>STFT</td>
<td>Groups A, B, D</td>
<td>300</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>KPCA</td>
<td>and E</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D8</td>
<td>STFT</td>
<td>Groups A, B, C, D</td>
<td>500</td>
<td>30</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>KPCA</td>
<td>and E</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### FIGURE 6. The weight of each view for D1 dataset in the DIC-MV-FCM algorithm

An analysis of the results in **TABLE 3** is presented in the following paragraphs.

a) When compared with the three non-multi-view clustering algorithms (FCM, K-means, and Co-clustering), the two multi-view clustering algorithms achieved better NMI and RI. The reason is that the FCM and K-means algorithms perform a separate cluster analysis on the sample data of each view and cannot exploit the links between the samples of each view. Second, the co-clustering algorithm is a multi-task clustering algorithm. This algorithm simply combines the multi-view sample into a set to train the model, and the samples of each view become a feature subset in the training set. Thus, as a single-view clustering algorithm, co-clustering cannot achieve satisfactory clustering performance.

b) The two multi-view clustering algorithms (Co-FKM and DIC-MV-FCM) can adopt more knowledge from multi-view datasets than the single-view clustering algorithms. Due to collaborative learning and view-weighting mechanisms, the proposed DIC-MV-FCM achieves the best performance. This result indicates that multi-view cooperation with view

### TABLE 3. Comparison of NMI & RI performance indices of several algorithms on the synthetic datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Index</th>
<th>FCM</th>
<th>K-means</th>
<th>Co-clustering</th>
<th>Co-FKM</th>
<th>DIC-MV-FCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>NMI</td>
<td>0.3081</td>
<td>0.1751</td>
<td>0.2855</td>
<td>0.2981</td>
<td>0.3182</td>
</tr>
<tr>
<td></td>
<td>RI</td>
<td>0.5830</td>
<td>0.5768</td>
<td>0.6495</td>
<td>0.6501</td>
<td>0.6651</td>
</tr>
<tr>
<td>D2</td>
<td>NMI</td>
<td>0.0012</td>
<td>0.0011</td>
<td>0.0020</td>
<td>0.0025</td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>RI</td>
<td>0.4912</td>
<td>0.4918</td>
<td>0.4987</td>
<td>0.4991</td>
<td>0.4993</td>
</tr>
<tr>
<td>D3</td>
<td>NMI</td>
<td>0.3201</td>
<td>0.2288</td>
<td>0.3760</td>
<td>0.3791</td>
<td>0.3861</td>
</tr>
<tr>
<td></td>
<td>RI</td>
<td>0.7038</td>
<td>0.6934</td>
<td>0.6733</td>
<td>0.7180</td>
<td>0.7211</td>
</tr>
<tr>
<td>D4</td>
<td>NMI</td>
<td>0.0688</td>
<td>0.6096</td>
<td>0.6070</td>
<td>0.6060</td>
<td>0.6069</td>
</tr>
<tr>
<td></td>
<td>RI</td>
<td>0.5182</td>
<td>0.7622</td>
<td>0.8111</td>
<td>0.8014</td>
<td>0.8111</td>
</tr>
<tr>
<td>D5</td>
<td>NMI</td>
<td>0.3283</td>
<td>0.2702</td>
<td>0.4127</td>
<td>0.4155</td>
<td>0.4175</td>
</tr>
<tr>
<td></td>
<td>RI</td>
<td>0.6191</td>
<td>0.6089</td>
<td>0.6402</td>
<td>0.6747</td>
<td>0.6804</td>
</tr>
<tr>
<td>D6</td>
<td>NMI</td>
<td>0.2577</td>
<td>0.2702</td>
<td>0.2514</td>
<td>0.2634</td>
<td>0.2753</td>
</tr>
<tr>
<td></td>
<td>RI</td>
<td>0.6282</td>
<td>0.6089</td>
<td>0.6606</td>
<td>0.6747</td>
<td>0.6911</td>
</tr>
<tr>
<td>D7</td>
<td>NMI</td>
<td>0.2577</td>
<td>0.2272</td>
<td>0.3523</td>
<td>0.2616</td>
<td>0.3535</td>
</tr>
<tr>
<td></td>
<td>RI</td>
<td>0.6282</td>
<td>0.6347</td>
<td>0.6985</td>
<td>0.6844</td>
<td>0.6998</td>
</tr>
<tr>
<td>D8</td>
<td>NMI</td>
<td>0.3201</td>
<td>0.2288</td>
<td>0.3760</td>
<td>0.3791</td>
<td>0.3861</td>
</tr>
<tr>
<td></td>
<td>RI</td>
<td>0.7038</td>
<td>0.6934</td>
<td>0.6733</td>
<td>0.7180</td>
<td>0.7211</td>
</tr>
</tbody>
</table>

**B. Performance Comparison**
The clustering results for the five algorithms with optimal parameters are shown in Table 3 (in terms of average NMI and RI). The best results are indicated in bold.
weighting is effective for detecting epilepsy electroencephalogram signals.

To further observe its ability to identify the importance of each view, the weight of each view for D1 dataset in DIC-MV-FCM is displayed in FIGURE 6. It can be easily seen that each view obtains a different weight according to its contribution to the clustering result. View 1 has a high weight value, so it has more excellent spatial division characteristics. View 2 and View 3 have low weight values, so they may cause serious overlapping in the clustering. Considering these weights, DIC-MV-FCM can identify the importance of each view, thereby achieving an accurate fuzzy partition. Co-FKM considers all views as equally important, and will therefore produce incorrect clustering results.

VII. CONCLUSION

In this paper, a novel double-index-constrained, multi-view, fuzzy clustering algorithm (DIC-MV-FCM) is proposed for detecting epilepsy EEG signals. Unlike the classical multi-view clustering that treats all views as equally important, the DIC-MV-FCM weights the degree of importance of each view by using a multi-view adaptive weighting strategy. Automatically evaluating the importance of each view, DIC-MV-FCM then weights each view and performs a weighted, multi-view, fuzzy clustering based on the FCM framework. The experiments conducted using real-world EEG signal datasets show that the proposed algorithm has achieved a satisfactory clustering performance. In the future, more research can focus on the proposed algorithm. The number of clusters is an important issue in clustering, so how to effectively automate the number of clusters has to be considered. FCM, as the baseline algorithm of DIC-MV-FCM, is sensitive to noise. Therefore, the application of our proposed algorithm to noisy EEG signals is also a challenge. Furthermore, developing a fast learning strategy for DIC-MV-FCM applied to large-scale EEG signals is an important study to be conducted in the future.

REFERENCES


**Jiaqi Zhu** is currently pursuing his M.D. degree at the School of Digital Media, Jiangnan University, China. His research area includes machine learning and its applications on medicine.

**Kang Li** is currently pursuing his M.D. degree at the School of Digital Media, Jiangnan University, China. His research area includes computational intelligence and its applications on medicine.

**Kaijian Xia** received his M.D. degree at Jiangnan University, Wuxi, China, in 2010, and now, he’s pursuing his Ph.D. degree at the China University of Mining and Technology. His research direction is medical information and medical image proceeding and so on. Now he worked in Department of Computer, Changshu No. 1 People’s Hospital. He has published several papers in international journals including Journal of medical systems, Journal of medical imaging and health informatics.

**Xiaqing Gu** received her Ph.D. degree at Jiangnan University, China, Wuxi, China, in 2017. She is currently a Lecturer in the School of Information Science and Engineering, Changzhou University, Changzhou, China. She has published over twenty papers in international/national authoritative journals. Her current research interests include pattern recognition and machine learning.

**Jing Xue** received her M.D (Doctor of Medicine) degree at the Department of Nephrology, Nanjing Medical University, Nanjing, China, in 2019. She received her M.M. degree from the Department of Nephrology, Nanjing Medical University, China. Her research area includes machine learning and its applications on medicine. She is a Doctor at the Affiliated Wuxi People’s Hospital of Nanjing Medical University, China.

**Shi Qiu** received the M.S. degree at Xidian University, Xi’an, P. R. China, in 2012, the Ph.D. degree at University of Chinese Academy of Sciences, Xi’an, P. R. China, in 2016. He is associate research fellow with Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an, P. R. China. His research interests include image restoration, sparse representation, and object detection.

**Yizhang Jiang** (M’2012) received his Ph.D. degree at Jiangnan University, Wuxi, China, in 2015. He is currently an Associate Professor with the School of Digital Media, Jiangnan University. He has also been a research assistant in the Department of Computing, Hong Kong Polytechnic University, for two years. His research interests include pattern recognition, intelligent computation and their applications. He is the author/co-author of more than 50 research papers in international/national journals, including IEEE Trans. on Fuzzy Systems, IEEE Trans. on Neural Networks and Learning Systems, IEEE Trans. on Cybernetics, ACM Trans. on Multimedia Computing, Communications, and Applications and Information Sciences. He has served as a reviewer or co-reviewer of several international conferences and journals, such as ICDM, TKDE, TFS, TNNLS, Pattern Recognition, Neurocomputing, and Neural Computing & Applications. He has also served as a leader guest editor or guest editor of several international journals, such as journal of Ambient Intelligence and Humanized Computing and journal of cloud computing.
Pengjiang Qian (M’2012) received the Ph.D. degree at Jiangnan University, Wuxi, China, in 2011. He is a Professor with the School of Digital Media, Jiangnan University, and Case Western Reserve University, Cleveland, OH, USA, as a Research Scholar in medical image processing. His current research interests include data mining, pattern recognition, bioinformatics and their applications, such as analysis and processing for medical imaging, intelligent traffic dispatching, and advanced business intelligence in logistics. He has authored/co-authored over 60 papers in international/national journals and conferences.