ShVEEGc: EEG Clustering With Improved Cosine Similarity-Transformed Shapley Value

Guanghui Li, Jiahua Shen, Chenglong Dai, Jia Wu, *Senior Member, IEEE*, and Stefanie I. Becker

***Abstract*—As the number of unlabeled or mislabeled electroencephalogram (EEG) increases dramatically in such applications as cerebral disease diagnosis, rehabilitation, and brain-computer interfaces, the supervised approaches that require labels or markers become inapplicable. Unfortunately, there are few reports on unsupervised studies for unlabeled EEG data, especially for unlabeled EEG clustering. To address the challenging task, we propose an effective approach named ShVEEGc for EEG clustering inspired by an improved Shapley value in cooperative game theory. The idea of ShVEEGc is first utilizing an improved cosine similarity to measure the correlations of EEG data and then calculating the improved Shapley value based on the inherent connection between unlabeled EEG data, which considers both global connections and local relationships potentially hidden in EEG data. Thus, ShVEEGc not only has good anti-interference ability but also can mine potential relationships among unlabeled EEG data. The comparison experiments with fourteen state-of-the-art EEG time series clustering algorithms on eleven real-world EEG datasets with four standard evaluation criteria demonstrate the efficacy and superiority of ShVEEGc for EEG clustering. Besides, the discussion on the impact of several different similarity measures on ShVEEGc also illustrates that the improved cosine similarity proposed in this paper is more suitable for EEG data.**

***Index Terms*—Cosine similarity, cross-correlation, EEG clustering, shapley value.**

I. Introduction

With the development of clinical medicine and brain-computer interface (BCI) technology [1], the study on bioelectrical signals [2] for detecting [3], monitoring [4], diagnosing and treating human diseases becomes significantly essential [5], [6]. As the common bioelectrical signals, electrocardiogram (ECG) [7], EEG [8], [9], electrooculogram (EOG) [10] and electromyogram (EMG) [11] are widely studied for related applications in our daily life, especially the EEG.

Manuscript received 28 November 2021; revised 11 June 2022; accepted 3 July 2022. This work was supported by National Natural Science Foundation of China under Grants 62106087 and 62072216, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20210455, and in part by the Fundamental Research Funds for the Central Universities under Grant JUSRP122033. *(Corresponding author: Chenglong Dai.)*

Guanghui Li, Jiahua Shen, and Chenglong Dai are with the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China (e-mail: ghli@jiangnan.edu.cn; 6201910031@stu.jiangnan.edu.cn; chenglongdai@jiangnan.edu.cn).

Jia Wu is with the Department of Computing, Macquarie University, Sydney, NSW 2109, Australia (e-mail: jia.wu@mq.edu.au).

Stefanie I. Becker is with the School of Psychology, University of Queensland, St Lucia, QLD 4072, Australia (e-mail: s.becker@psy.uq.edu.au).

Digital Object Identifier 10.1109/TETCI.2022.3189385

Electroencephalogram (EEG) is usually collected by inexpensive, low-risk and easy-to-operate non-invasive equipment [12], [13]. Besides, EEG data, which can be regarded as one specific type of time series, has the characteristics of high complexity, high dimensionality, non-stationarity, and low signal-to-noise ratio [14]–[16]. Furthermore, as is known to all, EEG can reflect the functional activities of the human brain and the status of physical health [17], so it is widely used in the driver drowsiness detection [18], [19], cerebral disease diagnosis [20], monitoring, and treatment, including epilepsy (EP) [21], stroke [22], Alzheimer's disease (AD) [23] and Parkinson's disease (PD) [24]. In addition, EEG also plays an essential role in the assistance and rehabilitation of patients with movement disorders to improve their living quality [25]. In detail, the patients can manipulate the wheelchairs or robot arms in their daily lives to access the objectives through motion imagination, and it also can achieve the actions that cannot be completed by themselves by setting specific modes in advance [26]. These applications have driven a surge in the number of multi-trial EEG, however the lack of labels or incorrect labels in practical applications has become the biggest limitation of EEG-based supervised learning techniques. Therefore, to address new problems emerging in EEG research, the unlabeled analysis technology has become a promising solution, i.e., clustering methods [14]–[16], generative models [27], and autoencoders [8]. To the end, this paper explores unsupervised EEG clustering methods and proposes a new solution for it.

A. Motivation

As the number of unlabeled or mislabeled electroencephalogram (EEG) increases dramatically in such applications as cerebral disease diagnosis, rehabilitation, and brain-computer interfaces, traditional supervised learning methods that require high-quality labels are no longer applicable.Therefore, unlabeled data mining technology, that is clustering, has become a powerful analysis method when there is no label or a small number of labels and the study of clustering has become a new direction for unlabeled EEG data, especially how to use a small amount of labeled data to analyze the unlabeled data to explore the internal connection between the data.In addition, the current research on EEG analysis is mainly focused on single-trial [28], so the correlation between EEG data [29] is often ignored, which contains a lot of valuable and interesting information. Multi-trial EEG signals are considered to be a special type of time series because of their high dimensionality, complexity, non-stationarity and low signal-to-noise ratio. Many time series clustering methods have been proposed, which perform well on traditional time series datasets [30], but these methods are difficult to deal with special EEG time series.

As mentioned in [31], most of the previously published algorithms utilized the concept of cluster center or cluster representation to group unlabeled data, which is usually based on global optimization or local optimization. On most time series datasets, this category of methods is effective. However, the clustering center will be severely disturbed by the noise, high complexity, and extreme instability of EEG data. Moreover, the clustering center only vaguely reflects the inherent concept of clustering, not to mention that the clustering center has been deteriorated by noise. Moreover, in the process of clustering, only the distance between trials and the nearest cluster center is considered, which seriously underestimates the importance of other trials. Therefore, we need a clustering method that can suppress noise interference and consider the inherent relationship between EEG data. Besides, in order to solve the problem of EEG data clustering, we propose a clustering method based on the improved Shapley value [32] that is an important concept in cooperative game theory [33]. In detail, the proposed method comprehensively considers the global connections and local relationships of clusters, which can effectively reduce the degradation of noise.

In summary, we utilized the concept of Shapley value in cooperative games to address EEG clustering and mapped it to a convex cooperative game [34]. Cooperative games require participants to cooperate to establish a collective rational coalition. Shapley value is one of the classic solutions of cooperative games, which can be understood as the contribution of participants to the coalition. For a given cluster (coalition), we calculate the improved similarity-transformed Shapley value of un-clustered samples and then cluster them according to the greedy principle. Remarkably, the proposed method emphasizes the collective cooperation and internal connection among EEG data.

B. Contributions and Outline

For the emerging unlabeled or mislabeled EEG data in aforementioned applications, EEG clustering becomes a challenging but valuable task. However, only a few related studies have been reported so far [14]. In this regard, we proposed an EEG clustering method based on an improved similarity-transformed Shapley value in cooperative game theory and mainly addressed three issues in EEG clustering: (1) the convexity of EEG clustering-mapped cooperative game; (2) the calculation of the Shapley value during clustering; and (3) the full consideration of collective rationality and individual desires of the participants (EEG data). The contribution of this paper is highlighted as follows.

- We, inspired by cooperative game theory, propose an improved Shapley value-transformed approach named ShVEEGc to quantify the contribution of every EEG signal in coalitions (i.e., clusters), which fully considers the collective rationality and individual desire of each EEG signal. In detail, ShVEEGc clusters EEG data based on the Shapley value of un-clustered EEG data, which can guarantee the inherent relationship among EEG data without complex computation for cluster centers.
- We propose a modified similarity measure, which exploits the cross-correlation theory to integrate the displacement information into the cosine distance. This method can more effectively measure the similarity of EEG data, thereby improving the clustering ability of various clustering algorithms.
- We propose a method to quantify the performance of similarity measures on EEG datasets via three self-designed evaluation criteria and carried out related experiments. The experimental results show that cosine distance, correlation and normalized cross-correlation have better characterization capabilities.
- We also show the efficacy of ShVEEGc via a detailed experimentation on eleven EEG datasets with respect to several standard evaluation criteria such as *rand index*, *F-score*, *Fleiss' kappa*, and *normalized mutual information*. Besides, the comparisons with fourteen state-of-the-art EEG time series clustering algorithms further demonstrate the superiority of ShVEEGc on EEG clustering.

The rest of the paper is organized as follows: The related works on EEG time series clustering algorithms and time series similarity measures are summarized in Section II. Then, the background knowledge of cosine distance, cross-correlation theory and Shapley value is introduced in Section III. Section IV presents in detail the proposed EEG clustering method, i.e., ShVEEGc, based on improved Shapley value, along with the time complexity analysis. Subsequently, Section V shows the experimental results and discussions on the efficacy of ShVEEGc. Finally, Section VI concludes the work of this paper and makes a prospect for future work.

II. Related Works

The choice of clustering algorithm and similarity measure is critical for clustering. In this part, we introduce clustering algorithms and similarity measures for EEG time series.

A. Time Series Clustering

So far, most of the studies on EEG data focused on supervision methods, especially classification [35]. There are few clustering studies related to unlabeled EEG data [14]–[16]. Fortunately, EEG data can be processed as a special type of time series by reshaping EEG data to one-dimensional. Therefore, the time series clustering methods provide a potential solution for EEG clustering. In recent years, many time series analysis methods have emerged, especially those based on cluster center searching, such as k-means, k-Shape [36] and so on. Unfortunately, this type of algorithms is subject to the clustering center optimization, which probably raises the problems of center initialization and convergence. Therefore, this type of methods may not be practical for EEG data. According to the characteristics of time series clustering algorithms, it can be roughly divided into six categories.

1) Classic Time Series Clustering: Among the classic clustering algorithms, k-means is commonly considered the most robust and efficient one for time series since it can easily apply to various types of time series datasets. In detail, the k-means allocates samples by calculating the Euclidean distance between samples and the cluster center. However, on some datasets, e.g., EEG, the accuracy of this algorithm is relatively low, and it is easy to fall into a local minimum and cannot handle the unbalanced datasets well. To address its drawbacks, many variant algorithms of k-means have been proposed, including km++ [37], RSFKM [38] and KMM [39]. To a large extent, km++ solves the problem of high convergence uncertainty of the k-means algorithm with probability distribution, which requires the mutual distance between the initial cluster centers to be as far as possible. This method inherits the shortcomings of k-means that are very sensitive to outliers and tend to converge locally, although it can cluster time series more efficiently. To solve these problem, RSFKM was subsequently proposed, which integrated the robustness and sparsity of clustering. Further, KMM forms a bipartite graph with candidate data and multiple sub-cluster centers, and applies rank constraints to cluster, so it can handle particular distribution datasets, such as crescent-shaped datasets that k-means cannot deal with.

2) Feature Selection-Based Time Series Clustering: This category of clustering methods mainly includes UDFS [40], NDFS [41] and RUFS [42], which first extracts distinguishable features from the data and then performs standard k-means for clustering. This type of clustering algorithms has many drawbacks, such as high computational complexity of feature extraction, too many parameters that need tuning, too much human supervision, and feature loss. Furthermore, due to the high dimension and the low signal-to-noise ratio of EEG data, it is challenging to apply feature selection-based clustering algorithms to cluster EEG data, although they have achieved a great of progress on traditional time series clustering.

3) Similarity Adjacency Matrix-Based Time Series Clustering: This type of clustering algorithms mainly includes HC [43], SC [44] and mwcEEGC [15]. HC defines the similarity or distance between network nodes through a given network topology and then uses single-connection hierarchical clustering or fully-connected hierarchical clustering to form a tree-like hierarchy of network nodes. Unfortunately, this tree structure is easily degraded by outliers. SC is a clustering method based on the spectrogram theory, which has the advantage that it can perform on the sample space of any shape and converge to the global optimal solution, compared to the traditional clustering methods. However, SC is very dependent on dimension reduction techniques and high-quality similarity matrix, which are challenging to satisfy, especially on the EEG datasets. Besides, the newly proposed mwcEEGC is a method for specifically clustering EEG data. Although mwcEEGc can well overcome the above shortcomings, its similarity measure is highly complicated and requires artificial thresholds to cut the EEG-mapped vertices in a weighted graph.

4) Distance-Based Time Series Clustering: This category of clustering algorithms mainly includes DBA [45] and K-SC [46]. These methods have three procedures for clustering: 1) the choice of distance measures; 2) the definition of cluster centers; 3) the definition of optimization function. In detail, DBA clusters time series by calculating dynamic time warping (DTW) and DTW barycenter averaging (DBA). K-SC uses a scale-and-shift-invariant similarity measure and seeks cluster centers according to the spectral norm of the similarity matrix. Unfortunately, these methods require many iterative operations, and they are sensitive to outliers and cannot well handle the non-stationary data well.

5) Shape-Based Time Series Clustering: As Paparrizos [36] mentioned, shape/shapelet-based clustering algorithms, such as k-Shape [36], MTEEGc [14] and USSL [47], cluster time series by capturing shape patterns or shapelets of time series with the scale-and-shift-invariant similarity. In detail, this type of methods first extracts the shape/shapelet information from the time series data as a reference feature and then performs clustering based on the shape/shapelet feature. Both k-Shape and MTEEGc are optimized by extracting time series features and then converting the optimization function into the Rayleigh quotient. These two methods usually take much time to calculate the shape/shapelet features and the clustering results are unstable and approximate. Furthermore, USSL scans the entire time series to extract local shapelet features and has good interpretability. However, this method sets many parameters that require experts' knowledge, so it seems not to be effectively applied to EEG datasets.

6) Density-Based Clustering: This kind of methods mainly includes DBSCAN [48], OPTICS [49] and SNN [50], which mainly utilizes certain density estimation methods to divide high-density areas into different clusters. DBSCAN needs to preset the radius (ϵ) and threshold (MinPts) based on experts' experience. However, this method is difficult to deal with clusters of different densities, such as EEG dataset for healthy subjects and patients, and it is sensitive to input parameters. As an improved DBSCAN, OPTICS is not sensitive to the input parameters and able to identify outliers. For another density-based clustering, SNN relies on the KNN strategy and is very sensitive to the value of k. In short, density-based methods have poor adaptability for density-imbalance datasets, and their parameter tuning is also a tough task. To address these issues, a density ratio strategy is proposed in [51], which includes R_DBSCAN, R_OPTICS and R_SNN, and the methods mentioned above are modified to use density estimator to compute density-ratio.

B. Similarity Measures

It is generally believed that the study of similarity measures is more important than the study of clustering algorithms because similarity measures can help algorithms characterize the correlation and irrelevance between data [36], [52]. There are many similarity measures, but as far as we know, no studies have reported which similarity measure is most suitable for EEG data. In Section V-E, we conducted a detailed experiment on the suitability of similarity measures for EEG data. The widely used similarity measures include Euclidean distance (ED), dynamic time warping (DTW) [45], cosine distance (COS), Minkowski distance (MK), CityBlock distance

(CB), correlation (CL), spearman (SPM) and normalized cross-correlation (NCCc) [53]. Euclidean distance is considered a kind of similarity measure with high robustness, good effectiveness and high efficiency, but unfortunately, on the EEG datasets, its performance is not good because it cannot capture the flexible change of EEG time series. Dynamic time warping is considered a very effective similarity measure, minimizing sequence similarity by distorting time series, but it has high time complexity and is over-optimized. In most cases, DTW is better than ED, but on the EEG datasets, the performance of DTW is unsatisfactory. As a modification of DTW, constrained dynamic time warping (cDTW) [54] shows better measuring efficiency and accuracy than DTW, but its constraints such as the window range and slope constraint need adjusting repeatedly, so it is not suitable for high-dimensional EEG data. Cosine distance measures the cosine value of the angle between pairwise time series, which can reflect the similarity in the sample space, but ignore the magnitude of the vector. MK is defined as the distance of Minkowski space, which is a space-time composed of one time dimension and three space dimensions, but its parameter needs to be determined by iteration. CityBlock, also called block distance, is a special case of Minkowski distance and it can be used to calculate the distance between intersections. Correlation represents the degree of correlation between two random variables, and SPM evaluates the monotone relationship between two random variables. Further, normalized cross-correlation can reflect the maximum connectivity of time series, and it is competitive to DTW. However, it only considers the maximum connectivity between time series.

III. Preliminaries

This section mainly introduces the preliminaries that we used in our method, including cosine distance, cross-correlation and Shapley value in cooperative game theory.

A. Cosine Distance

Cosine distance [55] uses the cosine value of the angle between two vectors in a vector space to measure the difference between two individuals. A vector is a directional line segment in a multidimensional space. To determine whether the directions of the two vectors are the same or not, the law of cosines is used to calculate the vector's angle. Given two vectors $\boldsymbol{x} = (x_1, \ldots, x_m)$ and $\boldsymbol{y} = (y_1, \ldots, y_m)$, the traditional cosine distance (TCD) is defined as

$$TCD(\boldsymbol{x}, \boldsymbol{y}) = 1 - \frac{\sum_{i=1}^{m} x_i y_i}{\sqrt{\sum_{i=1}^{m} x_i^2}\sqrt{\sum_{i=1}^{m} y_i^2}}, \tag{1}$$

where $TCD(\boldsymbol{x}, \boldsymbol{y}) \in [0, 2]$, and the smaller the TCD is, the higher the similarity is. Traditional cosine similarity (TCS) can be obtained by TCD. Namely,

$$TCS(\boldsymbol{x}, \boldsymbol{y}) = 1 - TCD(\boldsymbol{x}, \boldsymbol{y}), \tag{2}$$

where $TCS(\boldsymbol{x}, \boldsymbol{y}) \in [-1, 1]$, and the larger the TCS is, the higher the similarity is.

B. Cross-Correlation

As a widely used signal processing method, traditional cross-correlation (TCC) [56] generates a sequence containing displacement information through shifting or sliding. This displacement sequence represents the connectivity of different segments of two time series. TCC can process time series of different lengths, but we just need to use it to process EEG sequences of equal length in this paper. Given two vectors $\boldsymbol{x} = (x_1, \ldots, x_m)$ and $\boldsymbol{y} = (y_1, \ldots, y_m)$, TCC can be simplified to

$$TCC_\delta(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{m-|\delta|} (x_{i+\delta} \cdot y_i), \tag{3}$$

where δ is shift factor of vector $\boldsymbol{x}$ whose range is $(-m, m)$. As all δ are considered, the cross-correlation sequence (i.e., TCC) contains $2m - 1$ values.

In order to clarify the calculation of TCC simply, inspired by [14], we fix $\boldsymbol{y}$ and then slide $\boldsymbol{x}$ over $\boldsymbol{y}$ to calculate the point-to-point inner product for each shift δ of $\boldsymbol{x}$. Then, the part of $\boldsymbol{x}$ that slides away from $\boldsymbol{y}$ can be replaced by 0. To calculate the cross-correlation, we only need to take the inner product of the fixed $\boldsymbol{y}$ with the sliding sequence $\boldsymbol{x}$. The sliding sequence $\boldsymbol{x}$ can be expressed by (4).

$$\boldsymbol{x}_\delta = \begin{cases} (\overbrace{0, \ldots, 0}^{|\delta|}, x_1, x_2, \ldots, x_{m-\delta}), & \delta \geq 0 \\ (x_{1-\delta}, \ldots, x_{m-1}, x_m, \underbrace{0, \ldots, 0}_{|\delta|}), & \delta < 0 \end{cases}, \tag{4}$$

where $\boldsymbol{x}_\delta$ has $2m - 1$ expressions, corresponding to $2m - 1$ values of TCC . For the convenience of calculation, the (3) can also be equivalently expressed as

$$TCC_\delta(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} \sum_{i=1}^{m-\delta} x_{i+\delta} \cdot y_i, & \delta \geq 0 \\ TCC_{-\delta}(\boldsymbol{y}, \boldsymbol{x}), & \delta < 0 \end{cases}. \tag{5}$$

C. Shapley Value

The Shapley value [57] is one of the solutions to the static cooperative game [33]. Before introducing the Shapley value, we first make a brief introduction to the cooperative game.

1) Cooperative Game: Cooperative game is a game in which the interests of both coalitions can be increased, or at least the interest of either coalition is increased, while that of the other coalition is not degraded. In a cooperative game, players can be united into a binding and enforceable game coalition. Further, the cooperative game emphasizes collective rationality, fairness and justice. Cooperative games can be represented by set algebra $(\boldsymbol{N}, v)$, where $N = \{1, 2, 3, \ldots, n\}$ denotes the set of players, and $v : 2^n \rightarrow \mathbb{R}$ represents the value function. Payoff allocation is the most important concept of cooperative games,and we use $\boldsymbol{p} = \{p_1, p_2, \ldots, p_n\}$ to represent it. It is important to note that each player's share of payoff in the coalition should not be less than the payoff earned by operating individually, that is, $p_i \geq v(i)$ where $i \in \boldsymbol{N}$. Therefore, a cooperative game has many solutions, and of course, most of them are undesirable. The desired solutions to static cooperative games are mainly located in the core [58], stable sets, Shapley value, negotiation sets, kernel and nucleolus, and they probably obtain different payoff assignments for players.

2) Shapley Value: Although the cooperative game has many solutions, many unique solutions have emerged with its development, the most important of which is the Shapley value. The important breakthroughs and subsequent developments of cooperative games are largely derived from the proposal of the Shapley value, which portrays an intuitive and fair distribution scale. The formula for calculating the Shapley value of player i is defined as

$$\varphi_i(\boldsymbol{N},v)=\sum_{\boldsymbol{S}\subseteq\boldsymbol{N}\setminus\{i\}}\frac{|\boldsymbol{S}|!(n-|\boldsymbol{S}|-1)!}{n!}\{v(\boldsymbol{S}\cup\{i\})-v(\boldsymbol{S})\},\tag{6}$$

where $i\in\boldsymbol{N}$ and $\boldsymbol{S}$ is subset of $\boldsymbol{N}$, which reflects the average marginal contribution of players to the coalition. It should be noted that $|\boldsymbol{S}|$ represents the number of players in the game. The Shapley value can be expressed as $\varphi(\boldsymbol{N},v)=(\varphi_1(\boldsymbol{N},v),\varphi_2(\boldsymbol{N},v),\ldots,\varphi_n(\boldsymbol{N},v))$.

In order to simplify the calculation of Shapley value, we refer to [58]. For the $|\boldsymbol{N}|!$ permutations in the cooperative game, the permutation orders ($\boldsymbol{\Omega}$) can be expressed by (7).

$$\boldsymbol{\Omega}_{\epsilon,s}=\{i\in\boldsymbol{N}:\epsilon(i)\le s\}\quad(s\in\{1,2,\ldots,|\boldsymbol{N}|\}),\tag{7}$$

where ϵ represents the permutation of the players and $\epsilon(i)$ represents the position of player i in the permutation ϵ, so the payoff vector ($\boldsymbol{p}^\epsilon$) can be calculated by (8).

$$p_i^\epsilon=v\left(\boldsymbol{\Omega}_{\epsilon,\epsilon(i)}\right)-v\left(\boldsymbol{\Omega}_{\epsilon,\epsilon(i)-1}\right)\quad(\forall i=1,2,\ldots,|\boldsymbol{N}|).\tag{8}$$

According to Shapley's conclusion [59], the cooperative game is convex if its Shapley value (φ) is the center of the payoff vector (p^ϵ), that is

$$\varphi_i(\boldsymbol{N},v)=\frac{1}{|\boldsymbol{N}|!}\sum_{\epsilon\in\boldsymbol{\Psi}}p_i^\epsilon,\tag{9}$$

where $\boldsymbol{\Psi}$ is the set of all permutations of $\boldsymbol{N}$. The convexity of the cooperative game is proved in Section IV-C.

IV. SHVEEGC: SHAPLEY VALUE INSPIRED EEG CLUSTERING

This paper proposes a novel clustering method, that we call ShVEEGc, for EEG based on normalized cosine-based similarity measure, and improved Shapley value.

A. Improved Cosine-Based Similarity

In order to better measure the similarity of EEG data, we designed an improved cosine-based distance that fuses the spatial angle information and sequence displacement information, that is denoted by $dist$. Specifically, given two EEG signals $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$, the improved cosine-based distance is defined as (10), which consists of two items: traditional cosine distance (TCD) and cross-correlation-mapped displacement information ($Disp$).

$$dist(\boldsymbol{e}_i,\boldsymbol{e}_j)=\frac{\alpha}{2}\cdot TCD(\boldsymbol{e}_i,\boldsymbol{e}_j)+(1-\alpha)\cdot Disp(\boldsymbol{e}_i,\boldsymbol{e}_j),\tag{10}$$

where $Disp\in[0,1]$ and $TCD\in[0,2]$; α is the weight coefficient of TCD and $Disp$, and $\alpha\in[0,1]$.

Although many studies use the maximized cross-correlation value as the similarity, such as [14], [16], [36], [53], this measure only considers the maximum connectivity of two sequences and the largest connected segment may also be a coincidence (i.e., EOG noise). Consequently, it ignores the overall connectivity of sequences. Therefore, we consider the global displacement information $Disp$ to modify cosine distance for EEG data. Specifically, the displacement information $Disp$ is defined as

$$Disp(\boldsymbol{e}_i,\boldsymbol{e}_j)=\frac{1}{2m-1}\sum_{\delta\in[-m,m]}TCC_\delta(\boldsymbol{e}_i,\boldsymbol{e}_j).\tag{11}$$

Obviously, similar EEG series have similar displacement information. In other words, the smaller the $dist$, the more similar the two EEG series.

B. Cooperative Game-Mapped EEG Clustering

Given an n-trial EEG dataset $\boldsymbol{E}$ with length of m, we need to redefine $dist$ as a similar function $ICOS=f(dist):[0,1]\to(0,1]$, where $ICOS$ and $dist$ denote the improved cosine-based similarity and the degree of dissimilarity (also known as distance) between two EEG series, respectively. More specifically, f is a monotonic non-increasing similarity function defined on $dist:\boldsymbol{E}\times\boldsymbol{E}\to[0,1]$. For scale invariance, the similarity function f is defined as:

$$ICOS=f\left(dist\left(\boldsymbol{e}_i,\boldsymbol{e}_j\right)\right)=1-\frac{dist\left(\boldsymbol{e}_i,\boldsymbol{e}_j\right)}{dist_{\max}+1},\tag{12}$$

where $dist_{\max}$ denotes the maximum distance of all pairwise EEG series.

In our method, each EEG signal $\boldsymbol{e}$ is mapped as a player in a cooperative game and the number of players is the size of the EEG dataset: $|\boldsymbol{E}|=n$. Furthermore, we define a value function v, which is directly related to the establishment of cooperative games. Players in a cooperative game need to cooperate with each other to maximize the overall value of the coalition expressed by v and each of them is forbidden to build an individual coalition (i.e., only one player in the coalition), namely $v(\boldsymbol{e}_i)=0$. Mathematically, the value of coalition $\boldsymbol{C}$ can be calculated by (13).

$$v(\boldsymbol{C})=\frac{1}{2}\sum_{\substack{\boldsymbol{e}_i,\boldsymbol{e}_j\in\boldsymbol{C}\\ \boldsymbol{e}_i\neq\boldsymbol{e}_j}}f\left(dist\left(\boldsymbol{e}_i,\boldsymbol{e}_j\right)\right).\tag{13}$$

C. Convexity of EEG Clustering Cooperative Game

According to Shapley's theory [59], the convexity of the EEG clustering-transformed cooperative game $(\boldsymbol{N},v)$ is equivalent to proving the value function v is convex. In other words, for any two coalitions $\boldsymbol{C}$ and $\boldsymbol{D}$ (satisfying $\boldsymbol{C}\subseteq\boldsymbol{D}\subseteq\boldsymbol{E}$), it's to prove

$$v(\boldsymbol{C}\cup\{\boldsymbol{e}_i\})-v(\boldsymbol{C})\le v(\boldsymbol{D}\cup\{\boldsymbol{e}_i\})-v(\boldsymbol{D}),\tag{14}$$

where the value of $v(\boldsymbol{C}\cup\{\boldsymbol{e}_i\})-v(\boldsymbol{C})$ represents the marginal distribution of i in coalition $\boldsymbol{C}$. We can see from the (14) that in convex cooperative games $(\boldsymbol{N},v)$, the marginal contribution of players increases as the size of coalition increases.

Theorem 1: The EEG clustering-mapped cooperative game with the value function defined in (13) is convex.

Proof: According to [59], the marginal contribution of EEG clustering-transformed cooperative game can be calculated by (14), i.e.,

$$\begin{aligned}
&v(\boldsymbol{D}\cup\{\boldsymbol{e}_i\})-v(\boldsymbol{C}\cup\{\boldsymbol{e}_i\})\\
&=\frac{1}{2}\sum_{\substack{\boldsymbol{e}_j,\boldsymbol{e}_k\in\boldsymbol{D}\\ \boldsymbol{e}_j\neq\boldsymbol{e}_k}} f\left(dist\left(\boldsymbol{e}_j,\boldsymbol{e}_k\right)\right)+\sum_{\boldsymbol{e}_j\in\boldsymbol{D}} f\left(dist\left(\boldsymbol{e}_j,\boldsymbol{e}_k\right)\right)\\
&\quad-\frac{1}{2}\sum_{\substack{\boldsymbol{e}_j,\boldsymbol{e}_k\in\boldsymbol{C}\\ \boldsymbol{e}_j\neq\boldsymbol{e}_k}} f\left(dist\left(\boldsymbol{e}_j,\boldsymbol{e}_k\right)\right)-\sum_{\boldsymbol{e}_j\in\boldsymbol{C}} f\left(dist\left(\boldsymbol{e}_j,\boldsymbol{e}_k\right)\right)\\
&=\frac{1}{2}\sum_{\substack{\boldsymbol{e}_j,\boldsymbol{e}_k\in\boldsymbol{D}\backslash\boldsymbol{C}\\ \boldsymbol{e}_j\neq\boldsymbol{e}_k}} f\left(dist\left(\boldsymbol{e}_j,\boldsymbol{e}_k\right)\right)+\sum_{\substack{\boldsymbol{e}_j\in\boldsymbol{D}\backslash\boldsymbol{C}\\ \boldsymbol{e}_k\in\boldsymbol{C}}} f\left(dist\left(\boldsymbol{e}_j,\boldsymbol{e}_k\right)\right)\\
&\quad+\sum_{\boldsymbol{e}_j\in\boldsymbol{D}\backslash\boldsymbol{C}} f\left(dist\left(\boldsymbol{e}_j,\boldsymbol{e}_i\right)\right)\\
&=v(\boldsymbol{D})-v(\boldsymbol{C})+\sum_{\boldsymbol{e}_j\in\boldsymbol{D}\backslash\boldsymbol{C}} f\left(dist\left(\boldsymbol{e}_j,\boldsymbol{e}_i\right)\right)\\
&\geq v(\boldsymbol{D})-v(\boldsymbol{C})\Longrightarrow Eq.(14).
\end{aligned}$$

Consequently, the EEG cooperative game established by $v(\boldsymbol{C})=\frac{1}{2}\sum_{\substack{\boldsymbol{e}_i,\boldsymbol{e}_j\in\boldsymbol{C}\\ \boldsymbol{e}_i\neq\boldsymbol{e}_j}} f(dist(\boldsymbol{e}_i,\boldsymbol{e}_j))$ is convex. ■

D. Shapley Value in EEG Clustering

As Theorem 1 indicates, the EEG clustering-transformed cooperative game is convex, so the Shapley value φ can be equivalently calculated according to (8) and (13).

Theorem 2: The Shapley value of our EEG clustering-transformed cooperative game can be calculated with (13).

Proof: According to (8) and (13), the Shapley value of our EEG clustering-transformed cooperative game can be equivalently written as

$$\begin{aligned}
\varphi_i(\boldsymbol{N},v)&=\frac{1}{|\boldsymbol{N}|!}\sum_{\epsilon\in\boldsymbol{\Psi}} v\left(\boldsymbol{\Omega}_{\epsilon,\epsilon(i)}\right)-v\left(\boldsymbol{\Omega}_{\epsilon,\epsilon(i)-1}\right)\\
&=\frac{1}{|\boldsymbol{N}|!}\sum_{\epsilon\in\boldsymbol{\Psi}}\left[\sum_{\epsilon(j)<\epsilon(m)\leq\epsilon(i)} f\left(dist\left(\boldsymbol{e}_i,\boldsymbol{e}_j\right)\right)\right.\\
&\qquad\left.-\sum_{\epsilon(j)<\epsilon(m)\leq\epsilon(i)-1} f\left(dist\left(\boldsymbol{e}_i,\boldsymbol{e}_j\right)\right)\right]\\
&=\frac{1}{|\boldsymbol{N}|!}\sum_{\epsilon\in\boldsymbol{\Psi}}\sum_{\epsilon(j)<\epsilon(i)} f\left(dist\left(\boldsymbol{e}_i,\boldsymbol{e}_j\right)\right).
\end{aligned}$$

Since there are $(|\boldsymbol{N}|-1)!$ permutation orders for any EEG data $\boldsymbol{e}_i\in\boldsymbol{E}$, there are $(|\boldsymbol{N}|-2)!$ permutation orders for each of other EEG data $\boldsymbol{e}_j\in\boldsymbol{E}\backslash\{\boldsymbol{e}_i\}$. To the end, the Shapley value can be simplified to (15).

$$\begin{aligned}
\varphi_i&=\frac{(|\boldsymbol{N}|-2)!}{|\boldsymbol{N}|!}\left[\sum(|\boldsymbol{N}|-1)\sum_{\substack{\boldsymbol{e}_j\in\boldsymbol{E}\\ j\neq i}} f\left(dist\left(\boldsymbol{e}_i,\boldsymbol{e}_j\right)\right)\right]\\
&=\frac{1}{2}\sum_{\substack{\boldsymbol{e}_j\in\boldsymbol{E}\\ j\neq i}} f\left(dist\left(\boldsymbol{e}_i,\boldsymbol{e}_j\right)\right).
\end{aligned} \tag{15}$$

■

Considering that the EEG dataset itself is noisy and cooperative games pursue collective rationality, we put forward a combination of collective rationality and individual desire in our method, to improve Shapley value, which is defined as

$$\varphi_i=\frac{1}{2}\beta\sum_{\substack{\boldsymbol{e}_j\in\boldsymbol{E}\\ j\neq i}} f\left(dist\left(\boldsymbol{e}_i,\boldsymbol{e}_j\right)\right)+(1-\beta)\cdot max(f(dist(\boldsymbol{e}_i))), \tag{16}$$

where $\beta\in[0,1]$ is the weight coefficient, and $max(f(dist(\boldsymbol{e}_i)))$ denotes the similarity between $\boldsymbol{e}_i$ and the trial with the highest degree of cooperation among the EEG data that cooperate with $\boldsymbol{e}_i$.

E. ShVEEGc Algorithm

Algorithm 1 describes the primary process of ShVEEGc. First of all, the algorithm requires $|\boldsymbol{K}|$ initial clusters, and it can be generated by k-means or directly set according to the trials of original EEG datasets. In practice, the clusters are initialized by using the clusters near cluster centers generated by k-means, and we set the size of each cluster to be around 10 (see Line 1).

The proposed algorithm needs to repeatedly calculate $f(dist)$, so we can calculate the similarity matrix (i.e., sim) in advance for subsequent queries, see Lines 2-6. First, we calculate the Shapley value for each cluster of un-clustered trials, see Lines 9-13. Then, according to the greedy principle, we globally search for the trial with the largest contribution (i.e., largest Shapley value) and assign it to the corresponding cluster, and meanwhile update the Shapley value matrix. The above process of searching and updating is repeated until set $\boldsymbol{G}$ is empty (see Lines 14-19). It should be noted that when updating the Shapley value matrix, only the links related to the newly added trial need to be updated. Subsequently, according to the degree vector of each cluster, we select the trials with the largest degree as centers and regenerate the initial clusters at the same time. The size of the initial clusters remains unchanged, and the trial clusters with the greatest correlation to the central trials are selected as the initial clusters, see Lines 20-24. Finally, we count the sum of the average values of multiple coalitions (i.e., clusters) as the iteration termination condition, see Lines 25-29. If the total contribution drops, the iteration will terminate.

F. Time Complexity of ShVEEGc

ShVEEGc uses an improved cosine similarity and improved Shapley value for EEG clustering. The most fundamental is the

Algorithm 1: ShVEEGc.

Input: $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$: Weight coefficient;
$\boldsymbol{E}_{n\times m} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n\}$: EEG dataset with n EEG signals; num_inter: Number of iterations;
$\boldsymbol{K} = \{1, 2, 3, \dots\}$: Set of cluster serial numbers;
Output: $\boldsymbol{C}_1, \boldsymbol{C}_2, \dots, \boldsymbol{C}_{|\boldsymbol{K}|}$: Cluster sets of EEG;
1: Initialize $\boldsymbol{C}_k (k \in \boldsymbol{K})$, $\boldsymbol{G} = \boldsymbol{E}$; $Coalition\_Value = 0$;
2: **for** $i = 1$ to n **do**
3:   **for** $j = i$ to n **do**
4:     $sim(\boldsymbol{e}_i, \boldsymbol{e}_j) \leftarrow f(dist(\boldsymbol{e}_i, \boldsymbol{e}_j))$ with (10)–(12);
5:   **end for**
6: **end for**
7: **for** $inter = 1$ to num_inter **do**
8:   $\boldsymbol{G} = \boldsymbol{G} \backslash \{\boldsymbol{C}_1 \cup \boldsymbol{C}_2 \cup \dots \cup \boldsymbol{C}_{|\boldsymbol{K}|}\}$;
9:   **for** $i = 1$ to $|\boldsymbol{K}|$ **do**
10:     **for** $j = 1$ to $size(\boldsymbol{G})$ **do**
11:       $\varphi_{ij} \leftarrow \frac{1}{2}\beta \sum_{\substack{\boldsymbol{e}_k \in \boldsymbol{C}_i \\ k \neq j}} sim(\boldsymbol{e}_j, \boldsymbol{e}_k) + (1-\beta) \cdot max(sim(\boldsymbol{e}_j))$ with (15) and (16);
12:     **end for**
13:   **end for**
14:   **repeat**
15:     $[i, j] \leftarrow max(\varphi_{ij})$;
16:     $\boldsymbol{C}_i = \boldsymbol{C}_i \cup \{\boldsymbol{e}_j\}$;
17:     $\boldsymbol{G} = \boldsymbol{G} \backslash \{\boldsymbol{e}_j\}$;
18:     Update φ_{ij} with (15) and (16);
19:   **until** $\boldsymbol{G} = \varnothing$;
20:   **for** $k = 1$ to $|\boldsymbol{K}|$ **do**
21:     $\boldsymbol{C}_k\_sim = sim(\boldsymbol{C}_k, \boldsymbol{C}_k)$;
22:     $\boldsymbol{C}_k\_D = sum(\boldsymbol{C}_k\_sim)$;
23:     Re-initialize $\boldsymbol{C}_k (k \in \boldsymbol{K})$ with $\boldsymbol{C}_k\_D$;
24:   **end for**
25:   $new\_Coalition\_Value \leftarrow \sum_{k \in \boldsymbol{K}} (\frac{2v(\boldsymbol{C}_k)}{size(\boldsymbol{C}_k))(size(\boldsymbol{C}_k)-1)})$ with (12) and (13);
26:   **if** $new\_Coalition\_Value < Coalition\_Value$ **then**
27:     $break$;
28:   **end if**
29:   $Coalition\_Value = new\_Coalition\_Value$;
30:   $\boldsymbol{G} = \boldsymbol{E}$;
31: **end for**

improved cosine similarity calculation, which needs repeatedly computing during the clustering process. In practice, we calculate the improved cosine similarity in advance and store them for the following usage. When calculating Shapley value, we only need to query the similarity matrix without extra computing, so it greatly reduces the calculation time.

1) Time Complexity of Improved Cosine Similarity: The improved cosine similarity can be transformed into calculating $Disp$ and TCD separately. However, to calculate $Disp$, we must first calculate cross-correlation curve whose complexity derived from (3) is $O(m^2)$, where m represents the length of EEG data. Refer to [60], Fast Fourier Transform (FFT) reduces the time complexity of TCC to $O(m \log m)$. Therefore, the time complexity of calculating $Disp$ is $O(m \log m + m)$. To the end, given n EEG data, the time complexity is $O(\max\{n^2 m \log m + n^2\, m, m^2 n^2\}) = O(m^2 n^2)$.

2) Time Complexity of Improved Shapley Value: As (16) indicates, the time complexity of Shapley value computation for one EEG signal is $O(n)$. Hence, the time for n EEG data is required $O(n^2)$, see Lines 9-13 in Algorithm 1. Besides, the time complexity to update and maintain Shapley values is $O(\frac{n^3}{|\boldsymbol{K}|})$, where $\boldsymbol{K}$ denotes the set of cluster numbers (see Lines 14-19).

3) Overall Time Complexity: To sum up, the exact time complexity of ShVEEGc to cluster n-trial EEG data is $O(|\boldsymbol{K}|\gamma nm + n^2\, m^2 + \gamma[n^2 + \frac{n^3}{|\boldsymbol{K}|} + |\boldsymbol{K}|(n^2 + n)])$, where γ denotes the number of iterations. In fact, $\gamma \approx |\boldsymbol{K}| \ll n \lll m$ commonly, so the time complexity of ShVEEGc can be expressed as $O(n^3 + n^2\, m^2)$, which indicates that the clustering efficiency seems to be low on large-scale EEG datasets.

V. Experiments

This section introduces the details of EEG datasets used in our study, evaluation criteria and baseline algorithms. Subsequently, we present the experimental results of ShVEEGC compared to fourteen state-of-the-art EEG time series clustering algorithms on eleven real-world EEG datasets, along with sensitivity analysis and similarity measure impact analysis.

A. EEG Datasets

In order to propose a general clustering algorithm for EEG data, we have to adopt multiple paradigms for EEG datasets and use some subjects' data from them. As shown in Table I, four types of EEG datasets are applied to evaluate the algorithms' performance, including (1) Two-cluster slow cortical potentials (SCPs[1]) datasets, i.e., #1 bciII_I_a and #2 bciII_I_b, recorded from a healthy subject and an artificially respirated ALS patient. It is worth noting that dataset #3, i.e., bci_I_a&I_b, is a mixture of dataset #1 and dataset #2 employing downsampling and deleting channels; (2) Three-cluster mental imagery EEG datasets,[2] i.e., bciIII_V_s1$\sim$bciIII_V_s3, spatially filtered through a surface Laplacian from 3 healthy subjects; (3) Four-cluster motor imagery EEG datasets,[3] i.e., bciIV_2a_s1$\sim$bciIV_2a_s3, including imagination of movements of left hand, right hand, both feet and tongue, from 3 healthy subjects; (4) Two-cluster motor imagery EEG datasets[3], i.e., bciIV_2b_s $\sim$ bciIV_2b_s2, including imagination of movements of left hand and right hand, from 2 healthy subjects. Importantly, the aforementioned EEG datasets are originally labeled, so we removed those labels for clustering in this paper and used all EEG channels, except for the dataset #2, where we kept the vEOG channel (i.e., vertical eye movements channel). In addition, all the EEG datasets are

[1]The SCPs datasets are publicly available online at https://www.bbci.de/competition/ii/.

[2]The mental imagery datasets are publicly available online at https://www.bbci.de/competition/iii/ for free.

[3]The motor imagery datasets are also publicly available online at https://www.bbci.de/competition/iv/.

TABLE I
EEG DATASETS DETAILS

S/N	Dataset	Detailed Desciption	Size of EEG Dataset (Trials×Length)	# of Channels	# of Clusters	# of Trial Proportions
# 1	bciII_I_a	Self-regulation of SCPs from a healthy subject	268×5376	6	2	135:133
# 2	bciII_I_b	Self-regulation of SCPs from an artificially respirated ALS patient	200×8064	7	2	100:100
# 3	bciII_I_a&I_b	Mixed dataset # 1 (one healthy subject) and dataset # 2 (an ALS patient)	468×5376	6	4	135:133:100:100
# 4	bciIII_V_s1	Mental imagery data spatially filtered by means of a surface Laplacian from 3 healthy subjects (s1,s2,s3)	3488×96	8	3	1440:928:1120
# 5	bciIII_V_s2		3472×96			1472:880:1120
# 6	bciIII_V_s3		3424×96			1104:1185:1135
# 7	bciIV_2a_s1	Multi-class motor imagery dataset (left hand, right hand,feet and tongue) from 3 healthy subjects (s1,s2,s3)	288×6886	22	4	72:72:72:72
# 8	bciIV_2a_s2					
# 9	bciIV_2a_s3					
# 10	bciIV_2b_s1	Motor imagery dataset (left hand and right hand) from 2 healthy subjects (s1,s2)	120×939	3	2	60:60
# 11	bciIV_2b_s2					

preprocessed by z-normalization before clustering and reshaped to one-dimension, which is used in [14]. Besides, we do not adopt any preprocessing techniques (e.g., band-pass filtering, time-frequency variation and feature extraction) because specific preprocessing techniques cannot cope with datasets with multiple paradigms. The datasets used in this paper are available at **https://github.com/Jackie-Day/EEG-data-and-descriptions** and the code of our work is publicly available at **https://github.com/shenjiahua36/shveegc**.

B. Evaluation Criteria

Seven evaluation criteria are used in this paper, including *rand index* (RI), *F-score*, *Fleiss' kappa* (κ), *normalized mutual information* (*NMI*), *discrimination ability* (*DA*), *category difference* (*CD*) and *composite indicator* (*CI*). Particularly, the first four are used to evaluate the performance of clustering algorithms, which are widely used to evaluate the performance of clustering methods [14]–[16], and the last three are used to evaluate the ability of similarity measures on EEG data. The last three criteria are designed in this paper and comprehensively reflect the prior distribution of intra-cluster compactness and inter-cluster scatter, which is based on similarity matrix.

- *Rand index* (*RI*) is a common and naive measure of clustering quality, which is calculated based on the confusion matrix. Mathematically, $RI = \frac{TP+TN}{TP+TN+FP+FN}$, where *TP,TN,FP,FN* stand for the number of true positives, true negatives, false positives and false negatives, respectively. The value range of *RI* is $[0,1]$, and a larger value means that the clustering results are consistent with the real situation.
- *F-score* is a indicator derived from *RI* and comprehensively weighs the precision (p) and recall (r) by setting a scale parameter $\beta \geq 0$. β is usually set according to the actual demand for p and r, and $\beta = 1$ in this paper. Mathematically, *F-score* $= (1+\beta^2)\cdot\frac{p\cdot r}{\beta^2 p + r}$, where $p = \frac{TP}{TP+FP}$ and $r = \frac{TP}{TP+FN}$.
- *Fleiss' kappa* (κ) is a statistic used to evaluate the consistency of multiple clusters. Consistency ratio ($\overline{P}$) and overall random agreement ($\overline{P_e}$) are required to calculate *Fleiss' kappa*. In detail, $\kappa = \frac{\overline{P}-\overline{P}_e}{1-\overline{P}_e}$, where $\overline{P} = \frac{1}{Nn(n-1)}(\sum_{i=1}^{N}\sum_{j=1}^{k} n_{ij}^2 - Nn)$ and $\overline{P}_e = \sum_{j=1}^{k}(\frac{1}{Nn}\sum_{i=1}^{N} n_{ij})^2$.
- *Normalized mutual information* (*NMI*) is an information measure that measures the degree of agreement between two data distributions. Specifically, it is used to calculate the correlation between two sets of events (i.e., clusters). Mathematically, *NMI* $= \frac{-2\sum_{i=1}^{C_A}\sum_{i=1}^{C_B}\boldsymbol{C}_{ij}\cdot\log(\frac{\boldsymbol{C}_{ij}\cdot\boldsymbol{N}}{\boldsymbol{C}_i\cdot\boldsymbol{C}_j})}{\sum_{i=1}^{C_A}\boldsymbol{C}_i\cdot\log(\frac{\boldsymbol{C}_i}{\boldsymbol{N}})+\sum_{i=1}^{C_B}\boldsymbol{C}_{ij}\cdot\log(\frac{\boldsymbol{C}_j}{\boldsymbol{N}})}$, where $\boldsymbol{N}$, $\boldsymbol{C}$, $\boldsymbol{C}_A(\boldsymbol{C}_B)$ and $\boldsymbol{C}_i(\boldsymbol{C}_j)$ denote the number of samples, a confusion matrix, the number of set $\boldsymbol{A}(\boldsymbol{B})$ and the sum of elements in matrix $\boldsymbol{C}$, respectively.
- *Discrimination Ability* (DA) measures the ability of a similarity measure to distinguish EEG data, which is defined as

$$DA = \frac{\frac{1}{|K|}\sum_{k\in K} DS_k}{\frac{1}{2\sum(|K|-1)}\sum_{k_i,k_j\in K} DS_{k_i k_j}} \tag{17}$$

where $DS_{k_i k_j}$ represents the degree of similarity between clusters; DS_k denotes the degree of similarity within clusters; $DS_{k_i k_j} = \frac{1}{|C_{k_i}||C_{k_j}|}\sum_{\substack{e_m\in C_{k_i}\\ e_n\in C_{k_j}}} similarity(e_m, e_n)$ and DS_k is a special case of $DS_{k_i k_j}$ when $i = j$. The bigger the DA, the better the discrimination ability of the similarity measure.
- *Class Difference* (CD), different from DA that ignores the dimension, measures the differences between the same

TABLE II
EEG CLUSTERING RESULTS ON ELEVEN EEG DATASETS WITH RESPECT TO *RI*,*F-SCORE*,κ AND *NMI*

Dataset	Measure	km++	KMM	UDFS	NDFS	RUFS	SC	HC	DBA	K-SC	MTEEGc	k-Shape	R_SNN	R_DB SCAN	RBM	ShVEEGc
#1	*RI*	0.5256	0.4982	0.4982	0.5324	0.5498	0.5297	0.4982	0.5051	0.4983	**0.5978**	0.5160	0.4982	0.4982	0.5451	0.5954
	F-score	0.5250	0.5056	0.4970	0.6306	0.5112	0.5254	0.4925	0.5460	0.4989	0.5728	0.4675	0.5075	0.5037	0.6530	**0.7201**
	κ	0.2299	0.0131	0.0188	0.2618	0.3207	0.2512	0.0076	0.0937	0.0185	0.2158	0.1853	0.0166	0.0000	0.3060	**0.4386**
	NMI	0.0411	0.0002	0.0004	0.0509	0.0757	0.0466	0.0201	0.0118	0.0215	0.0302	0.0260	0.0002	0.0000	0.0686	**0.1903**
#2	*RI*	0.4994	0.4978	0.4978	0.4999	0.5073	0.4990	0.4975	0.4982	0.4980	**0.5397**	0.4975	0.4975	0.4975	0.4977	0.5088
	F-score	0.5070	0.4915	0.5005	0.5350	0.5650	0.5170	0.5050	0.5150	0.4990	0.5732	0.4985	0.5000	0.5000	0.4900	**0.5750**
	κ	0.0560	0.0170	0.0250	0.0700	0.1400	0.0540	0.0100	0.0300	0.0260	0.1000	0.0070	0.0000	0.0000	0.0200	**0.1500**
	NMI	0.0029	0.0124	0.0005	0.0036	**0.0452**	0.0022	0.0235	0.0044	0.0008	0.0011	0.0001	0.0000	0.0000	0.0003	0.0313
#3	*RI*	0.7313	0.7392	0.6987	0.7419	0.7401	0.7356	0.2599	0.6707	0.6147	0.6855	0.6915	0.5941	0.2537	0.5783	**0.7686**
	F-score	0.2641	0.4857	0.2011	0.3440	0.2970	0.3109	0.2799	0.2650	0.2449	**0.6987**	0.2536	0.2543	0.2885	0.3590	0.6154
	κ	0.3612	0.3174	0.3620	0.3133	0.3212	0.3822	0.0065	0.2401	0.1688	0.3716	0.3476	0.1535	0.0000	0.2250	**0.4833**
	NMI	0.4119	0.5262	0.2987	0.4272	0.4221	0.4288	0.0346	0.1806	0.1546	0.2875	0.2689	0.0810	0.0000	0.1641	**0.6119**
#4	*RI*	0.5744	0.4150	0.5695	0.5747	0.6058	0.5930	0.3456	0.5650	0.5913	0.5892	0.5922	0.3896	0.4554	0.5655	**0.6352**
	F-score	0.3208	0.3756	0.3296	0.3071	0.4128	0.3040	0.2658	0.3928	0.3019	0.5519	0.3354	0.4128	0.3446	0.1775	**0.6178**
	κ	0.3304	0.0603	0.3266	0.3309	0.2901	0.3453	-0.0048	0.2667	0.3112	0.3327	0.2739	0.0145	0.1067	0.1966	**0.4345**
	NMI	0.1653	0.0805	0.1637	0.1656	0.1374	0.1560	0.0206	0.0999	0.1088	0.1171	0.0964	0.0050	0.1147	0.0935	**0.2220**
#5	*RI*	0.5300	0.3536	0.5313	0.5224	0.4868	0.5522	0.3516	0.5531	0.5358	0.5712	0.5653	0.3896	0.4554	0.5496	**0.6359**
	F-score	0.3008	0.4307	0.3514	0.4026	0.4240	0.3618	0.2497	0.2695	0.3699	**0.7325**	0.3524	0.4128	0.3446	0.2535	0.5732
	κ	0.1880	0.0133	0.0185	0.1541	0.1434	0.1403	0.0071	0.1724	0.1851	0.2988	0.1843	0.0145	0.1067	0.1664	**0.3443**
	NMI	0.1252	0.0385	0.0033	0.1241	0.0368	0.0915	0.0310	0.0770	0.0823	0.1101	0.0679	0.0050	0.1147	0.0362	**0.1709**
#6	*RI*	0.5301	0.3404	0.5356	0.5479	0.4659	0.5498	0.3361	0.5573	0.5543	0.5524	0.5274	0.3575	0.3398	0.4888	**0.6374**
	F-score	0.3341	0.3137	0.3215	0.4267	0.3665	0.3339	0.3321	0.3049	0.3591	0.4876	0.3164	0.3195	0.3137	0.3487	**0.6145**
	κ	0.1296	0.0089	0.0569	0.1365	0.0973	0.1443	0.0055	0.1013	0.1425	0.2207	0.0906	0.0131	0.0137	0.0312	**0.4221**
	NMI	0.0266	0.0408	0.0099	0.0219	0.0353	0.0236	0.0255	0.0181	0.0219	0.0421	0.0173	0.0041	0.0426	0.0039	**0.1613**
#7	*RI*	0.6328	0.2840	0.6250	0.6296	0.6264	0.6365	0.2579	0.2486	0.6283	0.6376	0.6270	0.5336	0.2474	0.6370	**0.6526**
	F-score	0.2573	0.2566	0.2476	0.2882	0.2917	0.2580	0.2500	0.2500	0.2448	**0.4672**	0.2427	0.2778	0.2500	0.2361	0.4444
	κ	0.1542	0.0185	0.0745	0.0972	0.1389	0.1185	0.0093	0.0015	0.0764	0.1884	0.0676	0.0741	0.0000	0.1481	**0.2593**
	NMI	0.0558	0.0372	0.0161	0.0349	0.0635	0.0468	0.0468	0.0090	0.0175	0.0414	0.0149	0.0370	0.0000	0.0527	**0.1003**
#8	*RI*	0.5915	0.2706	0.6255	0.6264	0.6256	0.6317	0.2579	0.2474	0.6262	0.6284	0.6249	0.3870	0.2474	0.6283	**0.6406**
	F-score	0.2563	0.2472	0.2528	0.1806	0.3160	0.2403	0.2535	0.2500	0.2490	**0.4577**	0.2514	0.2535	0.2500	0.2569	0.4063
	κ	0.0662	0.0181	0.0602	0.0787	0.1019	0.0931	0.0093	0.0000	0.0653	0.2038	0.0731	0.0417	0.0000	0.0787	**0.2083**
	NMI	0.0169	0.0502	0.0108	0.0203	0.0468	0.0279	0.0468	0.0000	0.0105	0.0341	0.0123	0.0250	0.0000	0.0147	**0.0529**
#9	*RI*	0.5276	0.2727	0.6301	0.6298	0.6284	0.6318	0.2578	0.2474	0.6300	0.6302	0.6248	0.2918	0.2474	0.6262	**0.6448**
	F-score	0.2507	0.2507	0.2576	0.2986	0.2882	0.2583	0.2465	0.2500	0.2646	0.4471	0.2556	0.2361	0.2500	0.2396	**0.4653**
	κ	0.0667	0.0167	0.1005	0.1296	0.0741	0.1130	0.0139	0.0000	0.1014	0.2822	0.0806	0.0231	0.0000	0.0787	**0.2870**
	NMI	0.0210	0.0390	0.0271	0.0354	0.0468	0.0266	0.0466	0.0000	0.0240	0.0607	0.0173	0.0392	0.0000	0.0197	**0.0942**
#10	*RI*	0.4972	0.4959	0.4977	0.4964	0.5048	0.4968	0.4959	0.4963	0.5053	**0.5652**	0.5049	0.4958	0.4958	0.5098	0.5518
	F-score	0.4975	0.5033	0.5150	0.5167	0.5667	0.5050	0.5083	0.4944	0.5025	0.5648	0.4825	0.5000	0.5000	0.4167	**0.6667**
	κ	0.0417	0.0100	0.0500	0.0333	0.1333	0.0400	0.0167	0.0222	0.1183	0.1122	0.1050	0.0000	0.0000	0.1667	**0.3333**
	NMI	0.0021	0.0162	0.0028	0.0008	0.0318	0.0016	0.0318	0.0007	0.0139	0.0154	0.0144	0.0000	0.0000	0.0209	**0.0887**
#11	*RI*	0.4973	0.4972	0.5037	0.4958	0.5195	0.4959	0.4959	0.4960	0.4989	0.5541	0.4982	0.4958	0.4958	0.4959	**0.5636**
	F-score	0.5075	0.5217	0.5017	0.5000	0.5583	0.4958	0.4917	0.5028	0.4725	0.5512	0.5050	0.5000	0.5000	0.4917	**0.6833**
	κ	0.0417	0.0467	0.1233	0.0000	0.2167	0.0117	0.0167	0.0167	0.0617	0.1107	0.0567	0.0000	0.0000	0.0167	**0.3667**
	NMI	0.0104	0.0448	0.0117	0.0000	0.0937	0.0001	0.0318	0.0165	0.0044	0.0085	0.0044	0.0000	0.0000	0.0002	**0.1008**
RI Avg. Rank		7.5455	11.3636	7.6364	6.7273	5.9091	5.3636	13.1818	9.5455	6.4545	3.0909	7.3636	12.6364	14.3636	7.5455	**1.2727**
# Best *RI*		0	0	0	0	0	0	0	0	0	3	0	0	0	0	**8**
F-score Avg. Rank		8.3636	8.3636	9.6364	6.0909	3.9091	8.1818	10.5455	9.5455	10.6364	2.0000	11.0000	8.9091	10.7273	10.7273	**1.3636**
# Best *F-score*		0	0	0	0	0	0	0	0	0	4	0	0	0	0	**7**
κ Avg. Rank		6.0909	11.7273	7.8182	6.5455	5.2727	5.7273	13.1818	10.1818	7.0000	3.1818	7.9091	12.6364	14.0909	7.6364	**1.0000**
# Best κ		0	0	0	0	0	0	0	0	0	0	0	0	0	0	**11**
NMI Avg. Rank		6.3636	6.4545	9.9091	7.0909	3.8182	7.1818	7.3636	10.4545	9.8182	6.0909	10.5455	12.0000	12.1818	9.6364	**1.0909**
# Best *NMI*		0	0	0	0	1	0	0	0	0	0	0	0	0	0	**10**

class and the different class, which is defined as

$$CD = \frac{1}{|\boldsymbol{K}|}\sum_{k\in\boldsymbol{K}} DS_k - \frac{1}{2\sum(|\boldsymbol{K}|-1)}\sum_{k_i,k_j\in\boldsymbol{K}} DS_{k_ik_j} \tag{18}$$

Further, the larger the CD is, the easier it is for the similarity measure to distinguish EEG data.

- *Composite Indicator* (CI) simultaneously considers the self-recognition ability and class difference. Mathematically, $CI = \frac{CD}{|\boldsymbol{K}|}\sum_{k\in\boldsymbol{K}} DS_k$. A higher CI reflects an qualitatively outstanding similarity measure.

In a word, the higher the *RI*, κ, *F-score* and *NMI* are, the better the quality of clustering methods achieve. Likewise, the higher the *DA*, *CD* and *CI* are, the better the ability of similarity measure achieve.

C. Baseline Methods and Selected Similarity Measures

In order to illustrate the efficacy of ShVEEGc on EEG clustering, we compared it with fourteen state-of-the-art EEG time series clustering algorithms introduced in Section II. These baseline algorithms are mainly divided into six categories: (1) k-means-derived variants: km++ [37] and KMM [39]; (2) feature selection-based algorithms: UDFS [40], NDFS [41] and RUFS [42]; (3) methods based on similarity adjacency matrix: SC [44] and HC [43]; (4) distance-based algorithms: DBA [45]

TABLE III
REPRESENTATION ABILITY OF SIMILARITY MEASURES ON ELEVEN EEG DATASETS WITH RESPECT TO *DA*,*CD* AND *CI*

Dataset	Measure	ED	DTW	COS	MK	CB	CL	SPM	NCCc
#1	*DA*	1.04833	1.02245	1.09309	1.04833	1.06285	1.06945	1.05783	**1.13031**
	CD	0.03249	0.01718	0.06268	0.03249	0.03696	0.04876	0.03706	**0.06796**
	CI	0.02289	0.01344	**0.04613**	0.02289	0.02309	0.03661	0.02512	0.04006
#2	*DA*	1.00319	1.00164	1.00181	1.00319	1.00269	1.00334	**1.00418**	1.00264
	CD	0.00233	0.00137	0.00144	0.00233	0.00205	0.00233	**0.00271**	0.00187
	CI	0.00171	0.00115	0.00114	0.00171	0.00156	0.00163	**0.00176**	0.00133
#3	*DA*	1.15952	1.08451	1.47250	1.15952	1.14242	1.35750	1.30825	**1.50836**
	CD	0.10804	0.06825	**0.25960**	0.10804	0.10103	0.21258	0.17974	0.22684
	CI	0.08485	0.05977	**0.21003**	0.08485	0.08187	0.17160	0.13711	0.15268
#4	*DA*	1.07845	1.04589	1.12575	1.07845	1.10535	**1.14729**	1.02936	1.11274
	CD	0.04897	0.03080	0.07839	0.04897	0.04762	**0.08317**	0.01668	0.06426
	CI	0.03297	0.02162	**0.05502**	0.03297	0.02379	0.05388	0.00975	0.04076
#5	*DA*	1.01926	1.01369	1.03616	1.01926	1.02676	**1.06436**	1.01626	1.05325
	CD	0.01321	0.01020	0.02429	0.01321	0.01461	**0.03312**	0.00905	0.02872
	CI	0.00924	0.00770	0.01691	0.00924	0.00818	**0.01814**	0.00512	0.01631
#6	*DA*	1.00878	1.00398	1.01265	1.00878	1.01113	**1.02040**	1.01072	1.01874
	CD	0.00568	0.00294	0.00871	0.00568	0.00626	**0.01174**	0.00640	0.01100
	CI	0.00370	0.00218	0.00608	0.00370	0.00356	**0.00690**	0.00386	0.00658
#7	*DA*	1.01534	1.00499	1.05138	1.01534	1.01710	1.05062	**1.05234**	1.03779
	CD	0.00961	0.00379	0.02164	0.00961	0.01034	0.02159	**0.02171**	0.01199
	CI	0.00611	0.00289	0.00958	0.00611	0.00636	**0.00968**	0.00948	0.00394
#8	*DA*	1.04145	1.01835	1.03585	1.04145	1.03773	1.03578	1.03273	**1.05097**
	CD	0.00395	0.00351	**0.00543**	0.00395	0.00419	0.00542	0.00542	0.00353
	CI	0.01255	0.00796	**0.01370**	0.01255	0.01235	0.01368	0.01310	0.01308
#9	*DA*	1.04257	1.03862	1.05064	1.04257	1.04242	1.04867	1.04846	**1.07173**
	CD	0.01260	0.01156	**0.01503**	0.01260	0.01271	0.01476	0.01432	0.01441
	CI	0.00389	0.00359	0.00469	0.00389	0.00397	**0.00470**	0.00444	0.00310
#10	*DA*	1.02550	1.01714	1.03122	1.02550	1.02312	1.03095	1.03033	**1.05087**
	CD	0.00998	0.00835	0.01088	0.00998	0.00972	0.01036	0.00993	**0.01344**
	CI	0.00400	0.00414	0.00391	0.00400	**0.00418**	0.00358	0.00335	0.00373
#11	*DA*	1.02393	1.01638	1.03170	1.02393	1.02441	1.03180	1.03247	**1.06046**
	CD	0.01106	0.00856	0.01382	0.01106	0.01120	0.01388	0.01382	**0.01604**
	CI	0.00523	0.00455	0.00622	0.00523	0.00526	**0.00625**	0.00607	0.00451
DA Avg. Rank		4.90909	7.90909	3.09091	5.90909	5.09091	2.63636	4.36364	2.09091
# Best *DA*		0	0	0	0	0	3	2	6
CD Avg. Rank		5.09091	7.81818	2.45455	6.09091	5.54545	2.09091	4.27273	2.63636
# Best *CD*		0	0	3	0	0	3	2	3
CI Avg. Rank		4.45455	7.00000	2.54545	5.45455	4.81818	2.27273	4.45455	5.00000
# Best *CI*		0	0	4	0	1	5	1	0

and K-SC [46]; (5) shape/shapelet-based algorithms: k-Shape [36] and MTEEGc [14]; (6) density-based algorithms: R_SNN [51] and R_DBSCAN [51]. In addition, we also compared an unsupervised learning-based generative model RBM [27]. To reduce the accidental errors, all baseline algorithms are run ten times, and the clustering results are averaged. Besides, the parameters of the above fourteen algorithms are set as the same as the original papers, except that for RBM, we increase the training epochs to make it converge, and the number of clusters is preset according to the class number of original EEG datasets.

As an essential role in EEG clustering, similarity measures probably affect the performance of clustering algorithms, so we also analyzed eight different similarity measures in the experiments, including ED, DTW, COS, MK, CB, CL, SPM and NCCc. All experiments are conducted with Matlab 2019b on Ubuntu machines with 3.5 GHz CPU and 64 GB memory.

D. Clustering Results and Discussion

We analyze cluster performance of ShVEEGc from the view of accuracy, consistency and correlation by using *RI*, *F-score*, κ and *NMI*, and compare it with fourteen baseline methods for EEG time series clustering. The comparison results are shown in detail in Table II (The best performing values are highlighted in bold), which clearly demonstrates the efficacy and superiority of ShVEEGc on EEG clustering, as it yields better *RI*, *F-score*, κ and *NMI* than those fourteen state-of-the-art clustering algorithms. Besides, the average statistics of *RI*, *F-score*, κ and *NMI* also demonstrate that ShVEEGc outperforms the state-of-the-art clustering algorithms.

Most of the fourteen methods are based on EEG-to-center relationships, local density ratio, local relationship or global segmentation, so they just consider the similarities between EEG data and centers or high-density EEG clusters, while ignore the connections to other EEG data in the same cluster or the low-density EEG clusters. Moreover, although those methods based on global segmentation can take into account the connections of global EEG data, they ignore the value of the local relationship. As a result, their clustering accuracy is degraded. On the contrary, ShVEEGc can make full use of the global connections between EEG data and pays extra attention to the value of local relationships, so it can yield a better clustering result on complex EEG datasets.

E. Similarity Measure Ability Quantification and Discussion

It is generally believed that similarity measure is more important than clustering algorithms [36], [52], but selecting suitable similarity measure for EEG data is a crucial task. Unfortunately, although many studies evaluate the quality of similarity measures from the level of the similarity matrix, such as [15], no relevant experiments have been carried out to illustrate the adaptability of similarity measures to EEG datasets. In this paper, we quantify the level of similarity matrix and analyze the ability of various similarity measures from a numerical point of view, i.e., *DA*, *CD* and *CI*. The results are shown in detail in Table III, which demonstrates that COS, CL and NCCc have better characterization capabilities, but it is hard to discern the best similarity measure for EEG data since no similarity measure can get the best average ranking on all EEG datasets.

DTW is a kind of similarity measures that is more suitable for traditional time series, but its performance on the EEG datasets is unacceptable. Besides, it's very time consuming, especially for high dimensional data, such as EEG data. NCCc is experimentally regarded as a better similarity measure for EEG than DTW thanks to its scale-and-shift invariance and low computational complexity. However, although the time-consuming similarity measures such as DTW and its variant cDTW perform well on some EEG datasets, they cannot guarantee the good performance on all EEG datasets. Actually, this kind of similarity measures get extremely poor performance on most of EEG datasets, see Table III. In contrast, those similarity measures with low computational complexity such as ED, CL and COS perform well on almost of EEG datasets.

F. Sensitivity Analysis

1) Impact of α and β: We exploited an improved cosine similarity modified by cross-correlation mean to measure EEG data in the paper, which contains the displacement information of time series in the whole time period, as (10) defined, where α weighs cosine similarity measure and displacement information.

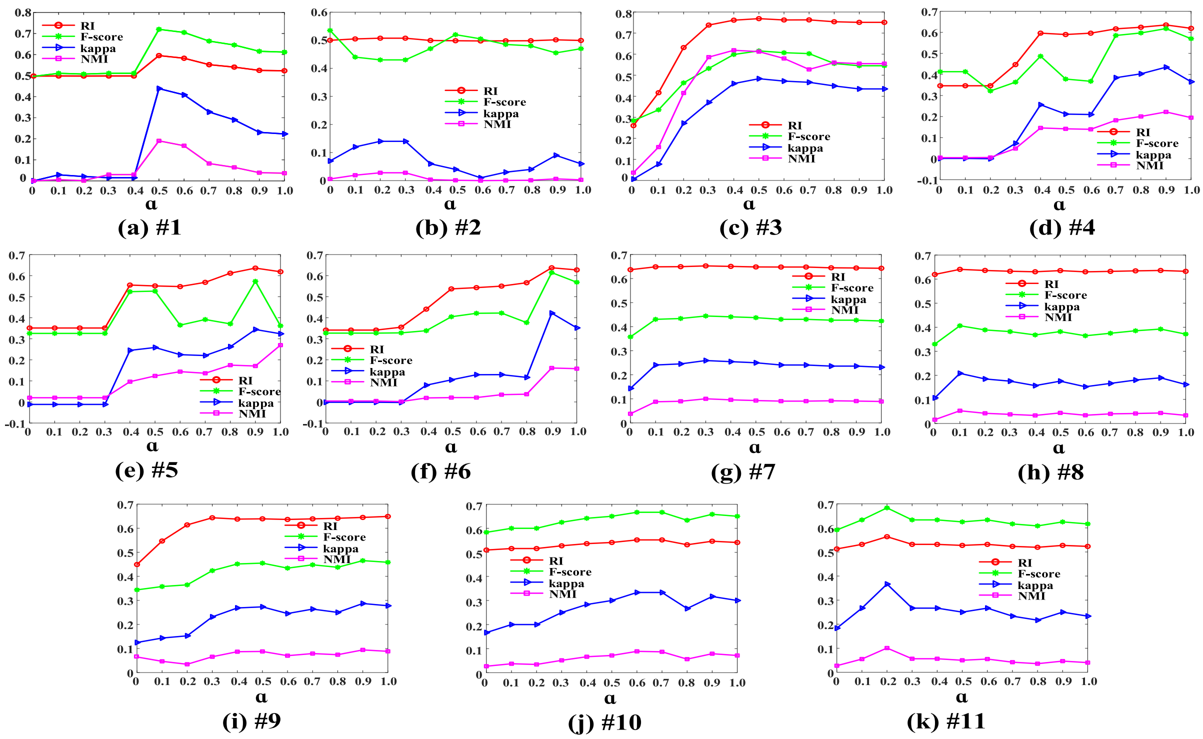


Fig. 1. The impact of α on ShVEEGc. A uniform sampling of $[0, 1]$ is used to measure the performance of ShVEEGc.

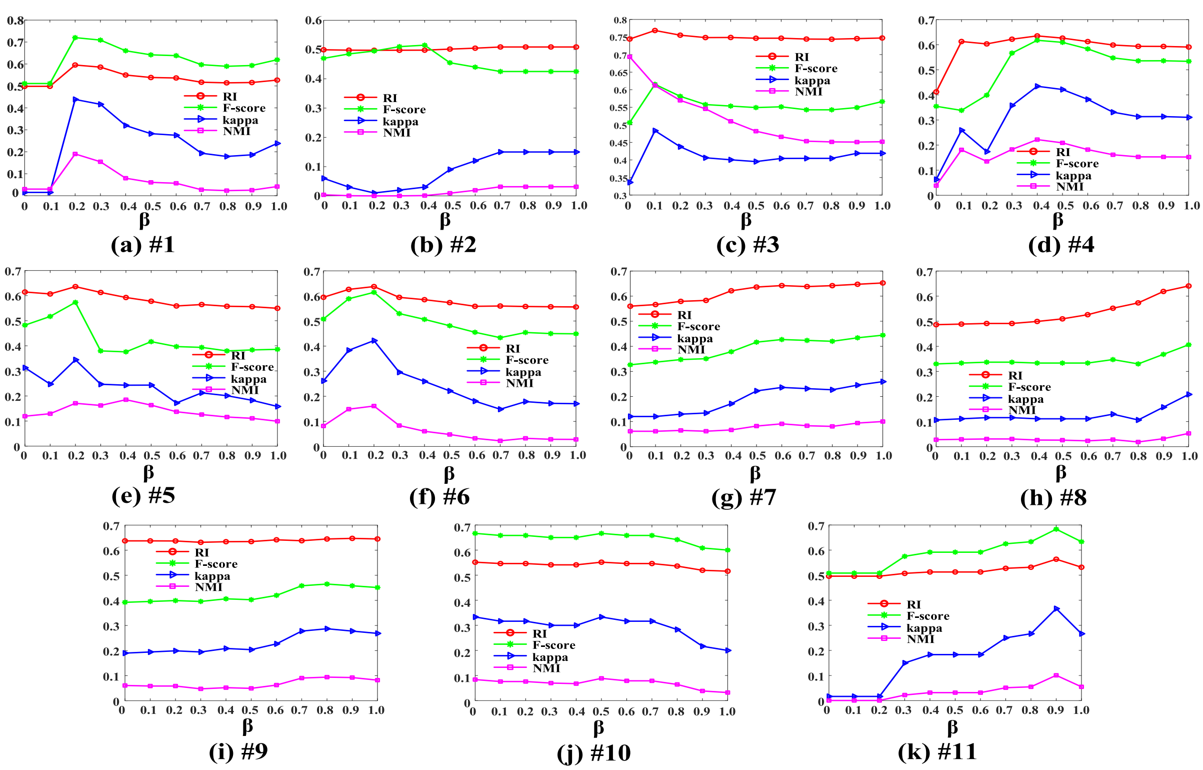


Fig. 2. The impact of β on ShVEEGc. A uniform sampling of $[0, 1]$ is used to measure the performance of ShVEEGc.

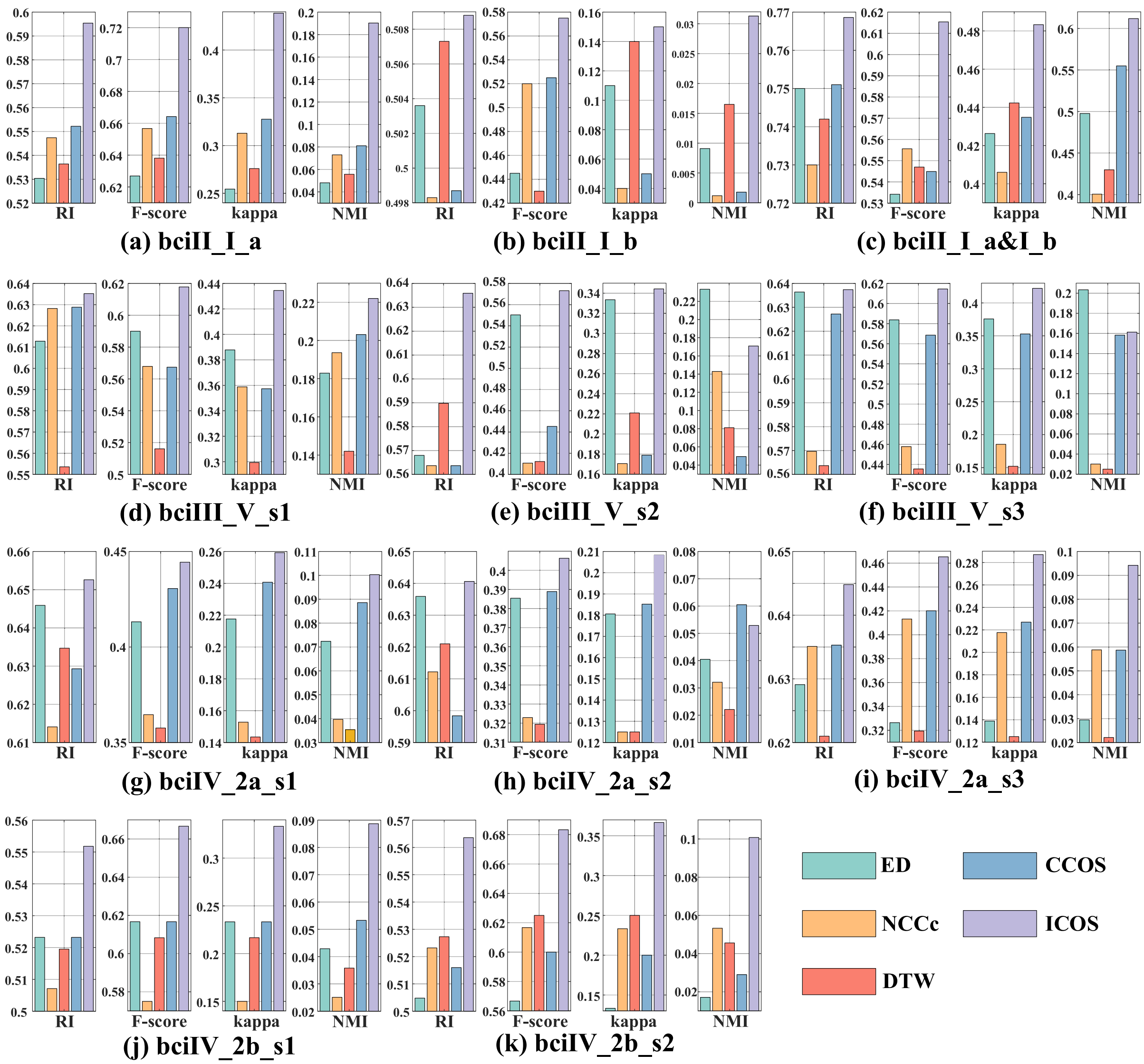


Fig. 3. The impact of similarity measures on ShVEEGc. All the results are achieved with optimal parameters. (ED: Euclidean similarity measure; NCCc: Normalized cross correlation; DTW: Dynamic time warping; CCOS: Conventional cosine without modification; ICOS: Improved cosine.

In this section, we discuss the impact of α on ShVEEGc for EEG clustering. It should be noted that α is determined by grid search because the computational cost is very low. The results are shown in Fig. 1, which indicates that different types of datasets or experiments on different tasks require different α when clustering, but the datasets of the same experiment usually share similar α (i.e., α for dataset bciIII_V_s1, bciIII_V_s2 and bciIII_V_s3 are set to around 0.9). As Fig. 1 illustrates, we can see that when only the cosine similarity measure is considered (i.e., $\alpha = 1$) or only the cross-correlation mean is considered (i.e., $\alpha = 0$), the performance of the ShVEEGc is not as good as that with the fusion of cosine similarity and displacement information.

As introduced before, clustering using global connections or local relationships of EEG series alone cannot achieve good results, especially local relationships (such as HC and density-based clustering methods). Therefore, we introduced β to weigh global connections and local relationships, see (16), and we also discussed its impact on ShVEEGc here. The results are illustrated in Fig. 2, which demonstrates that similar to α, on most EEG datasets, ShVEEGc with a compromise β can achieve the best clustering performance with repect to *RI*, *F-score*, $kappa$, and *NMI*. In other words, from the perspective of cooperative games, the combination of collective rationality and individual desire can achieve a better coalition.

2) Impact of Similarity Measures: As we introduced before, similarity measures play a significantly important role in ShVEEGc, but it is difficult to choose the most suitable one for EEG datasets from various similarity measures. Consequently, to select the relatively applicable similarity measure for most EEG datasets in our study, we here discussed the influence of different similarity measures on the ShVEEGc, including Euclidean similarity (ED), normalized cross-correlation (NCCc), dynamic time warping (DTW), conventional cosine similarity (CCOS) and improved cosine similarity (ICOS). The results are shown in Fig. 3, which demonstrates that the ShVEEGc with

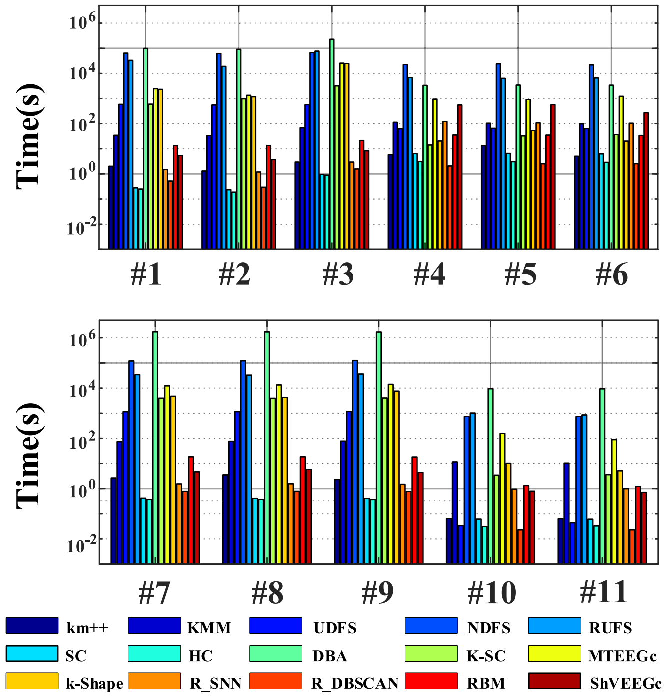


Fig. 4. Running time comparison of different EEG time series clustering algorithms on eleven EEG datasets.

ICOS can yield better performance compared to ED, NCCc, DTW and CCOS for all eleven EEG datasets, even though several criteria are not as good as Euclidean distance. Based on a large number of experiments, we find that either similarity measures or clustering algorithms can perform well on all EEG datasets. Interestingly, the results in Fig. 3 have a great relationship with the results in Table III. Although there are some exceptions, they are generally consistent with each other. In detail, ED is a kind of robust similarity measure and it can achieve good results in most fields, which is also in keeping with the conclusion in [53], but it is not the best choice for the EEG dataset according to the experimental results. NCCc just focuses on the segment with the largest sequence binding, ignoring the global sequence's binding properties. Besides, DTW minimizes sequence distance by distorting local time series. In response to this, DTW is over-optimized and is affected by the instability and complexity of the EEG sequence. For another similarity measure, CCOS measures the angle information of the space very well, but it cannot capture the information of the global sequence displacement. As we introduced before, ICOS considers both the spatial angle information and global sequence displacement information of EEG data, so it can assist ShVEEGc to obtain better clustering performance on EEG data compared with ED, NCCc, DTW and CCOS.

3) Execution Time: As mentioned earlier, the time complexity of ShVEEGc is $O(n^3 + n^2\, m^2)$, and on large-scale datasets, the execution time of this algorithm is mainly concentrated on the calculation of the Shapley value. Once the primary cluster is changed, the algorithm updates the Shapley value matrix immediately. Of course, for the small and medium scale datasets, ShVEEGc is very efficient. The time consumption on eleven EEG datasets is shown in Fig. 4, which clearly indicates that ShVEEGc runs faster than most of those fourteen baseline algorithms. Among these methods, km++, SC, HC, R_SNN and R_DBSCAN are all efficient algorithms because of their low complexity of similarity measures and efficiently simple clustering strategy. In addition, UDFS, NDFS, RUFS, k-shape, DBA, K-SC, and MTEEGc take a large amount of running time to optimize the objective function to seek optimal cluster centers and extract features, so they have higher time consumption. Although ShVEEGc is relatively slow on large-scale datasets (i.e., dataset bciIII_V_s1, bciIII_V_s2 and bciIII_V_s3) according to the analysis of time complexity, in fact, the running time of ShVEEGc is less than that of NDFS, RUFS, DBA and MTEEGc, and the clustering accuracy and quality are significantly improved. Besides, the ShvEEGc algorithm can be potentially well applied to real-time EEG analysis. After forming stable cooperative game coalitions, we only need to calculate the Shapley value of real-time EEG to cluster it, which only requires $O(n)$ time, where n represents the number of EEG.

VI. Conclusion and Future Work

Inspired by cooperative game theory and Shapley value, we proposed an EEG clustering method in this study, called ShVEEGc. First, we transform EEG clustering into a cooperative game and prove its convexity. Then we calculate the Shapley values of EEG data based on an improved cosine similarity measure modified by cross-correlation and finally cluster the EEG with the improved Shapley value. ShVEEGc avoids the calculation of cluster centers and forms stable and tight clusters through the internal relationship between EEG data. Finally, we compare ShVEEGc with fourteen state-of-the-art EEG time series clustering methods on eleven real-world EEG datasets to illustrate the effectiveness and superiority of ShVEEGc. In addition, due to the lack of similarity measure research in the field of EEG clustering, we design three indicators to quantify the representation ability of similarity measures on EEG data. The results also demonstrate that cosine distance, correlation and normalized cross-correlation have good characterization capabilities.

To be honest, the computational efficiency of the ShVEEGc algorithm is not outstanding on large-scale EEG datasets for now, so we plan to improve its efficiency from two aspects in our future work: 1) considering more efficient data linking methods, such as tree structure, and 2) adopting adaptive strategies [61], [62]. Besides, in the future, we will try to evaluate and improve the efficacy of our method on higher spatial-resolution EEG datasets that recorded by more channels, and we will also plan to apply ShVEEGc to real-world EEG clustering and other related applications, such as ECG and EMG clustering.

References

[1] Z. Lv, L. Qiao, Q. Wang, and F. Piccialli, "Advanced machine-learning methods for brain-computer interfacing," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 5, pp. 1688–1698, Sep./Oct. 2021.
[2] C. W. Thomas, W. P. Huebner, and R. J. Leigh, "A low-pass notch filter for bioelectric signals," *IEEE Trans. Biomed. Eng.*, vol. 35, no. 6, pp. 496–498, Jun. 1988.

[3] J. J. Vidal, "Real-time detection of brain events in EEG," *Proc. IEEE*, vol. 65, no. 5, pp. 633–641, May 1977.
[4] T. Nakamura, V. Goverdovsky, M. J. Morrell, and D. P. Mandic, "Automatic sleep monitoring using Ear-EEG," *IEEE J. Transl. Eng. Health Med.*, vol. 5, 2017, Art. no. 2800108.
[5] L. Aron and B. A. Yankner, "Neural synchronization in Alzheimer's disease," *Nature*, vol. 540, no. 7632, pp. 207–208, 2016.
[6] C. Babiloni *et al.*, "Abnormal cortical sources of resting state electroencephalographic rhythms in single treatment-naïve HIV individuals: A statistical z-score index," *Clin. Neurophysiol.*, vol. 127, no. 3, pp. 1803–1812, 2016.
[7] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "ECG analysis: A new approach in human identification," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 3, pp. 808–812, Jun. 2001.
[8] M. Wang, S. Abdelfattah, N. Moustafa, and J. Hu, "Deep Gaussian mixture-hidden Markov model for classification of EEG signals," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 2, no. 4, pp. 278–287, Aug. 2018.
[9] A. Saha, A. Konar, and A. K. Nagar, "EEG analysis for cognitive failure detection in driving using type-2 fuzzy classifiers," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 1, no. 6, pp. 437–453, Dec. 2017.
[10] R. Barea, L. Boquete, M. Mazo, and E. López, "Wheelchair guidance strategies using EOG," *J. Intell. Robot. Syst.*, vol. 34, no. 3, pp. 279–299, 2002.
[11] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: Detection, processing, classification and applications," *Biol. Procedures Online*, vol. 8, no. 1, pp. 11–35, 2006.
[12] R. Martins, S. Selberherr, and F. A. Vaz, "A CMOS IC for portable EEG acquisition systems," *IEEE Trans. Instrum. Meas.*, vol. 47, no. 5, pp. 1191–1196, Oct. 1998.
[13] P. Fiedler *et al.*, "Novel multipin electrode cap system for dry electroencephalography," *Brain Topogr.*, vol. 28, no. 5, pp. 647–656, 2015.
[14] C. Dai, D. Pi, L. Cui, and Y. Zhu, "MTEEGC: A novel approach for multi-trial EEG clustering," *Appl. Soft. Comput.*, vol. 71, pp. 255–267, 2018.
[15] C. Dai *et al.*, "Brain EEG time-series clustering using maximum-weight clique," *IEEE Trans. Cybern.*, vol. 52, no. 1, pp. 357–371, Jan. 2022.
[16] C. Dai, J. Wu, D. Pi, L. Cui, B. Johnson, and S. I. Becker, "Electroencephalogram signal clustering with convex cooperative games," *IEEE Trans. Knowl. Data Eng.*, early access, pp. 1–14, Feb. 22, 2021, doi: 10.1109/TKDE.2021.3060742.
[17] M. F. Glasser *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, pp. 171–178, 2016.
[18] T. K. Reddy, V. Arora, S. Kumar, L. Behera, Y.-K. Wang, and C.-T. Lin, "Electroencephalogram based reaction time prediction with differential phase synchrony representations using co-operative multi-task deep neural networks," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 3, no. 5, pp. 369–379, Oct. 2019.
[19] Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task," *Sci. Data*, vol. 6, no. 1, pp. 1–8, 2019.
[20] S. J. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy," *J. Neurol. Neurosurg. Psychiatry*, vol. 76, no. suppl 2, pp. ii2–ii7, 2005.
[21] Y. Zhang *et al.*, "Epilepsy signal recognition using online transfer TSK fuzzy classifier underlying classification error and joint distribution consensus regularization," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 5, pp. 1667–1678, Sep./Oct. 2021.
[22] P. Langhorne, J. Bernhardt, and G. Kwakkel, "Stroke rehabilitation," *Lancet*, vol. 377, no. 9778, pp. 1693–1702, 2011.
[23] J. Liu, M. Li, W. Lan, F.-X. Wu, Y. Pan, and J. Wang, "Classification of Alzheimer's disease using whole brain hierarchical network," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 15, no. 2, pp. 624–632, Mar./Apr. 2018.
[24] B. Faraji, M. Gheisarnejad, Z. Esfahani, and M.-H. Khooban, "Smart sensor control for rehabilitation in Parkinson's patients," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 6, no. 2, pp. 267–275, Apr. 2022.
[25] G. Tan *et al.*, "Meta-analysis of EEG biofeedback in treating epilepsy," *Clin. EEG Neurosci.*, vol. 40, no. 3, pp. 173–179, 2009.
[26] I. Iturrate, J. Antelis, and J. Minguez, "Synchronous EEG brain-actuated wheelchair with automated navigation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2009, pp. 2318–2325.
[27] T. K. Reddy and L. Behera, "Online eye state recognition from EEG data using deep architectures," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, 2016, pp. 000712–000717.
[28] J. Lehtonen, P. Jylanki, L. Kauhanen, and M. Sams, "Online classification of single EEG trials during finger movements," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 2, pp. 713–720, Feb. 2008.
[29] J. Metsomaa, J. Sarvas, and R. J. Ilmoniemi, "Multi-trial evoked EEG and independent component analysis," *J. Neurosci. Methods*, vol. 228, pp. 15–26, 2014.
[30] T. W. Liao, "Clustering of time series data—A survey," *Pattern Recognit.*, vol. 38, no. 11, pp. 1857–1874, 2005.
[31] R. S. Michalski, "Knowledge acquisition through conceptual clustering: A theoretical framework and an algorithm for partitioning data into conjunctive concepts," *Int. J. Policy Anal. Inf. Syst.*, vol. 4, no. 3, pp. 219–244, 1980.
[32] Y. Zhang, P. Tiňo, A. Leonardis, and K. Tang, "A survey on neural network interpretability," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 5, no. 5, pp. 726–742, Oct. 2021.
[33] G. Chalkiadakis, E. Elkind, and M. Wooldridge, "Computational aspects of cooperative game theory," *Synth. Lectures Artif. Intell. Mach. Learn.*, vol. 5, no. 6, pp. 1–168, 2011.
[34] B. Rajasekhar, N. Pindoriya, W. Tushar, and C. Yuen, "Collaborative energy management for a residential community: A non-cooperative and evolutionary approach," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 3, no. 3, pp. 177–192, Jun. 2019.
[35] I. Razzak, M. Blumenstein, and G. Xu, "Multiclass support matrix machines by maximizing the inter-class margin for single trial EEG classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1117–1127, Jun. 2019.
[36] J. Paparrizos and L. Gravano, "k-Shape: Efficient and accurate clustering of time series," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2015, pp. 1855–1870.
[37] D. Arthur and S. Vassilvitskii, "k-means : The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.
[38] J. Xu, J. Han, K. Xiong, and F. Nie, "Robust and sparse fuzzy K-means clustering," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 2224–2230.
[39] F. Nie and C.-L. Wang, "K-multiple-means: A multiple-means clustering method with specified K clusters," in *Proc. IEEE 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 959–967.
[40] Y. Yang, H. Tao Shen, Z. Ma, Z. Huang, and X. Zhou, "L2, 1-norm regularized discriminative feature selection for unsupervised," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 1589–1594.
[41] Z. Li, Y. Yang, J. Liu, X. Zhou, and H. Lu, "Unsupervised feature selection using nonnegative spectral analysis," in *Proc. 26th AAAI Conf. Artif. Intell.*, 2012, pp. 1026–1032.
[42] M. Qian and C. Zhai, "Robust unsupervised feature selection," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, 2013, pp. 1621–1627.
[43] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.
[44] U. V. Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2004.
[45] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognit.*, vol. 44, no. 3, pp. 678–693, 2011.
[46] J. Yang and J. Leskovec, "Patterns of temporal variation in online media," in *Proc. 4th ACM Int. Conf. Web Search Data Mining*, 2011, pp. 177–186.
[47] Q. Zhang, J. Wu, P. Zhang, G. Long, and C. Zhang, "Salient subsequence learning for time series clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 9, pp. 2193–2207, Sep. 2019.
[48] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discov. Data Mining*, 1996, pp. 226–231.
[49] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: Ordering points to identify the clustering structure," *ACM SIGMOD Rec.*, vol. 28, no. 2, pp. 49–60, 1999.
[50] L. Ertöz, M. Steinbach, and V. Kumar, "Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data," in *Proc. SIAM Int. Conf. Data Mining*, 2003, pp. 47–58.
[51] Y. Zhu, K. Ting, and M. J. Carman, "Density-ratio based clustering for discovering clusters with varying densities," *Pattern Recognit.*, vol. 60, pp. 983–997, 2016.
[52] G. E. Batista, E. J. Keogh, O. M. Tataw, and V. de Souza, "CID: An efficient complexity-invariant distance for time series," *Data Mining Knowl. Discov.*, vol. 28, no. 3, pp. 634–669, 2014.
[53] J. Paparrizos and L. Gravano, "Fast and accurate time-series clustering," *ACM Trans. Database Syst.*, vol. 42, no. 2, pp. 1–49, 2017.

[54] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 26, no. 1, pp. 43–49, Feb. 1978.

[55] M. Senoussaoui, P. Kenny, T. Stafylakis, and P. Dumouchel, "A study of the cosine distance-based mean shift for telephone speech diarization," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 1, pp. 217–227, Jan. 2014.

[56] W. Huang, W. Zhang, T. Zhen, F. Zhang, and F. Li, "A cross-correlation method in wavelet domain for demodulation of FBG-FP static-strain sensors," *IEEE Photon. Technol. Lett.*, vol. 26, no. 16, pp. 1597–1600, Aug. 2014.

[57] E. Winter, "The shapley value," *Handbook Game Theory Econ. Appl.*, vol. 3, 2002, pp. 2025–2054.

[58] V. K. Garg, Y. Narahari, and M. N. Murty, "Novel biobjective clustering (BiGC) based on cooperative game theory," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 1070–1082, May 2013.

[59] L. S. Shapley, "Cores of convex games," *Int. J. Game Theory*, vol. 1, no. 1, pp. 11–26, 1971.

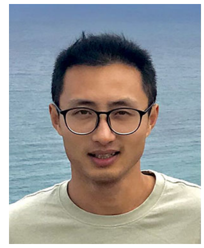
[60] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series," *Math. Comput.*, vol. 19, no. 90, pp. 297–301, 1965.

[61] T. K. Reddy, V. Arora, and L. Behera, "HJB-equation-based optimal learning scheme for neural networks with applications in brain-computer interface," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 4, no. 2, pp. 159–170, Apr. 2020.

[62] A. Kamble, P. Ghare, and V. Kumar, "Machine-learning-enabled adaptive signal decomposition for a brain-computer interface using EEG," *Biomed. Signal Process. Control*, vol. 74, 2022, Art. no. 103526.

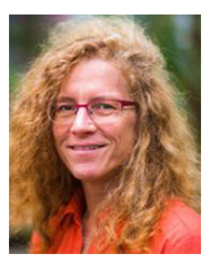
Guanghui Li received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005. He is currently a Professor with the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China. He has authored or coauthored more than 80 papers in journals or conferences. His research interests include edge computing, wireless sensor networks, fault tolerant computing, EEG processing, and nondestructive testing and evaluation. His research was supported by the National Natural Science Foundation of China, Jiangsu Provincial Science and Technology Foundation, and other governmental and industrial agencies.

Jiahua Shen received the B.S. degree in Internet of Things Engineering from Changzhou University, Changzhou, China, in 2020. He is currently working toward the master's degree with the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China. His research interests include time series clustering, EEG processing, and anti-artifacting technology.



Chenglong Dai received the M.S. and Ph.D. degrees in computer science and technology from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2014 and 2020, respectively. He is currently a Lecturer with the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China. He has authored or coauthored several related papers in prestigious journals and top conferences, including IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Cybernetics, *ACM Transactions on Knowledge Discovery from Data*, *ACM Transactions on Intelligent Systems and Technology*, and SIAM SDM. He was the recipient of SDM'18 Best Paper Award in Data Science Track. His research interests include brain-computer interfaces, EEG data mining, and machine learning. He was a Recognition Reviewer for *Knowledge-Based Systems* and a Reviewer for IJCNN'18–21.

Jia Wu (Senior Member, IEEE) received the Ph.D. degree in computer science from the University of Technology Sydney, Ultimo, NSW, Australia. He is currently an ARC DECRA Fellow with the Department of Computing, Macquarie University, Sydney, NSW, Australia. He has authored or coauthored more than 100 refereed journal and conference papers, including IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Multimedia, *ACM Transactions on Knowledge Discovery from Data*, NIPS, WWW, KDD, ICDM, IJCAI, and AAAI. His current research interests include data mining and machine learning. He was the recipient of SDM'18 Best Paper Award in Data Science Track, IJCNN'17 Best Student Paper Award, and ICDM'14 Best Paper Candidate Award. He is the Associate Editor for the *ACM Transactions on Knowledge Discovery from Data* and *Neural Networks*.



Stefanie I. Becker received the Ph.D. degree in 2007. She is currently an Associate Professor and an ARC Future Fellow with the University of Queensland, Brisbane, QLD, Australia. She is an internationally recognised expert on visual attention, and received several awards for her work involving eye tracking, EEG, and fMRI. She has authored or coauthored more than 60 papers in high-ranking journals. She was an Associate Editor for the journal *Journal of Experimental Psychology: Human Perception and Performance*.