Semi-supervised EEG Clustering with Multiple Constraints

Chenglong Dai, Jia Wu, *Senior Member, IEEE*, Jessica J.M. Monaghan, Guanghui Li, Hao Peng, Stefanie I. Becker, David McAlpine

***Abstract*—Electroencephalogram (EEG)-based applications in Brain-Computer Interfaces (BCIs, or Human-Machine Interfaces, HMIs), diagnosis of neurological disease, rehabilitation, *etc*, rely on supervised techniques such as EEG classification that requires given class labels or markers. Incomplete or incorrectly labeled or unlabeled EEG data are increasing with the ever-expanding amount of EEG data generated by such applications and the ambiguities these generate degrade the performance of supervised techniques. To address the challenging task of clustering EEG data with limited *priori* knowledge, we introduce a semi-supervised graph embedding EEG clustering approach termed *ConsEEGc* with multiple constraints, *i.e.*, label-transformed connectivity constraints that constrains the connection or disconnection among EEG data, compactness-and-scatter constraint that constrains the intra-cluster compactness and inter-cluster scatter of EEG clusters, and fairness constraint that constrains the fair ratio of elements between EEG clusters, to make best use of limited *priori* knowledge of EEG data and to achieve better EEG clustering results. *ConsEEGc* is conducted with an optimization objective function that integrates pseudo label learning, least-square error minimization and multiple constraints, and it can quickly converge to local optima. The experiments demonstrate that *ConsEEGc* can efficiently yield good clustering results on various types of real-world EEG datasets, compared to state-of-the-art standard unsupervised and semi-supervised EEG/time series clustering algorithms.**

***Index Terms*—Semi-supervised EEG clustering, connectivity constraints, compactness-and-scatter constraint, fairness constraint.**

I. Introduction

ELECTROENCEPHALOGRAM (EEG) is an established technique that has been widely applied in Brain-Computer Interfaces (BCIs) [1]–[3] as a non-invasive means of analysing brain working status and body functions [4]. Its utility in Human-Machine Interfaces (HMIs) [5], diagnosis of neuropsychiatric disease [6], [7], and rehabilitation [2], [3] has expanded in recent decades, but most of these applications are concentrated on supervised tasks that require a *priori* knowledge such as EEG labels or markers that defines the class they belong to. Unfortunately, not all EEG labels associated with specific patterns of brain activity can be completely or correctly obtained from subjects across different recording sessions separated by days or weeks. This is especially so for those complex and challenging patients groups, such as those suffering from Alzheimer's disease (AD) [6], epileptic seizures [8], stroke [2], [3] or amyotrophic lateral sclerosis (ALS) [9], *etc*. Consequently, the increasing amount of unlabeled or mislabeled EEG data likely degrades the efficacy of supervised methods, *e.g.*, classification that is vital to BCI-based applications and disease diagnosis. With the increased utility of EEG and other brain-monitoring or neuro-imaging methods, labelling signals is becoming a labor-intensive and time-consuming task with the ever-increasing unlabeled signals in every-day applications. A means of overcoming the problem of increasing amounts of unlabeled data is clustering. Clustering is a way to assign objects into corresponding groups based on the correlations or patterns of objects without any supervised information such as data labels or markers. In other words, the notion of clustering is group those objects with high similarity (similar patterns) into same class while separating different groups as far as possible. In real-world applications, labeled and unlabeled EEG data are often mixed together, hence labeled EEG data containing known information can, intuitively, improve the performance of unlabeled EEG clustering. That is, mixtures of labeled and unlabeled EEG data provide a semi-supervised way to analyze unlabeled EEG data. Here, we demonstrate that with a small amount of labeled EEG data, a semi-supervised-inspired strategy for unlabeled EEG clustering can improve the performance of neural signature analyzing using EEG, potentially enhancing its efficacy and utility in real-world applications.

A. Motivation

As a promising strategy to conduct unsupervised tasks, the emerging technique of EEG clustering [10], [11], is of interest to a wide range of researchers and clinicians. However, most of traditional clustering methods based on analysis of EEG time-series aim to group EEG signals based on inherent similarities/dissimilarities in the time course of the EEG data without any supervision of a *priori* knowledge. Unfortunately, however, the features of EEG time series in corresponding clusters obtained by completely unsupervised clustering appear unpredictable. Besides, as a specific type of bio-electrical signals, EEG's characteristics of high cerebration

Chenglong Dai and Guanghui Li are with the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China. E-mail: chenglongdai@jiangnan.edu.cn; ghli@jiangnan.edu.cn.

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

Jessica J.M. Monaghan is with the National Acoustic Laboratories, NSW 2109, Australia. E-mail: jessica.monaghan@nal.gov.au.

Hao Peng is with the Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100191, China. E-mail: penghao@buaa.edu.cn.

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.

David McAlpine is with the Department of Linguistics, Macquarie University, NSW 2109, Australia. E-mail: david.mcalpine@mq.edu.au.

relevance, high dimension, high complexity, low signal-to-noise, strong non-linearity and oscillation probably make it inapplicable for traditional time series clustering methods to handle partially labeled EEG. Consequently, semi-supervised clustering of EEG time series provides a promising solution to the problem of unpredictable data, not only guiding clustering methods to improve the performance of EEG clustering techniques. Unlike unsupervised analysis for completely unlabeled EEG data, semi-supervised analysis for partially labeled EEG data has not been reported. Further, a *priori* information from multiple sources, such as the limited label information, linkage and cluster fairness, provides constrained but potentially important information for semi-supervised EEG clustering. However, most traditional constrained semi-supervised clustering approaches do not make full use of a *priori* knowledge. This leads to four limitations: (1) most traditional constrained clustering approaches only consider single constraint, such as Must-link and Cannot-link constraints [12], or fairness [13] or label information [14]. Multiple constraints integrated with such single constraints are rarely considered simultaneously; (2) connectivity constraints, such as connection constraints (similar to Must-link constraints) and disconnection constraints (similar to Cannot-link constraints) are not well or fully used; (3) the intra-cluster compactness and inter-cluster scatter are not considered as a constraint for semi-supervised clustering; and (4) the fairness constraint is not considered as an additional constraint to monitor the balance of clustered partitions and improve cluster quality. Motivated to make full use of limited a *priori* knowledge of the small amount of labeled EEG data to improve clustering performance, we explored the potential benefits of semi-supervised clustering for analysing partially labeled EEG data. Specifically, we propose a multi-constraint semi-supervised clustering algorithm termed *ConsEEGc* for partially labeled EEG data, which simultaneously considers the connectivity constraints, compactness-and-scatter constraints, and fairness constraint.

B. Contributions and Outline

Semi-supervised EEG clustering is a non-trivial task in such applications as BCI, diagnosis of disease and rehabilitation, but few related studies on semi-supervised EEG clustering have been reported so far. To address the problem of semi-supervised EEG clustering, this paper proposes a multi-constraint semi-supervised graph embedding EEG clustering. In summary, the main contributions of the paper are briefly introduced as follows.

- Addressing the problem of semi-supervised EEG clustering, we proposed a novel optimization objective function with label-transformed connectivity constraints, compactness-and-scatter constraint, and fairness constraint. This represents the first attempt to address semi-supervised EEG clustering with multiple constraints.
- We exploited connectivity constraints to link EEG signals with a *priori* knowledge of a small amount of labeled EEG signals, with the aim of minimizing the potential connection or disconnection among partially labeled EEG data based on their similarities.
- We utilized compactness-and-scatter constraint to generate high-quality EEG clusters by minimizing inter-cluster scatter and maximizing intra-cluster compactness, and transformed this to fairness-constrained NCut (i.e., normalized Cut), to relax compactness-and-scatter constraint and achieve approximately equal-sized EEG clusters.
- We proposed a novel approach *ConsEEGc* for semi-supervised EEG clustering by applying a gradient descent strategy to obtain optimal minimization of an objective function that integrates pseudo label classifier learning and least-square error minimization of pseudo labels with connectivity constraints, compactness-and-scatter constraint, and fairness constraint. We also demonstrate the efficacy and superiority of *ConsEEGc* by comparing its performance with various types of clustering methods on several categories of EEG datasets.

The remainder of this paper is presented as follows: Related works on EEG/time series clustering are summarized in Section II. The preliminaries are introduced in Section III. Subsequently, the proposed model of multi-constraint semi-supervised EEG clustering with connectivity and fairness constraints is described in Section IV, and the algorithm *ConsEEGc* is presented in Section V, including convergence analysis, initialization analysis. Detailed experiments are presented in Section VI. In the end, we give conclusions and future works in Section VII.

II. Related Works

In this section, we summarize some works related to our work, including traditional unsupervised clustering and semi-supervised clustering.

A. Traditional Clustering

As the technique to deal with unlabeled data, various promising traditional unsupervised clustering approaches have been intensively emerged in recent decades. The reported unsupervised clustering approaches can be roughly grouped into five categories: Classical k-means-type clustering, such as k-means++ [15] and k-multiple-means [16], aims to enhance or improve the standard k-means algorithm by boosting the initialization of centers for standard k-means with specific strategies of probability enhancement or multiple subcenter modification; Density-based clustering, such as DBSCAN [17], OPTICS [18], and SNN [19], identifies dense regions with density estimators and then links neighbouring dense regions to cluster; Distance-based clustering, such as k-DBA [20] and K-SC [21], clusters data based on distance measures or similarity metrics, which tries to learn representative centroid sequences with an optimization objective on a specific distance measure such as Euclidean distance, dynamic time warping, *etc*; Feature-selection clustering, such as nonnegative spectral feature selection [22], robust unsupervised feature selection [23], and robust spectral feature selection [24], *etc*, transforms EEG time series into feature space, and then applies k-means to cluster with selected features; and shape-based or shapelet-based clustering such as k-Shape [25] and USSL [26], transforms EEG time series into shape/shapelet space by learning

distinct shapes/shapelets of EEG time series, and then clusters based on the correlations between EEG time series and the learnt shapes/shapelets. In summary, unsupervised clustering without any supervision or constraint clusters EEG time series just by learning the representative patterns of completely unlabeled data without considering *priori* knowledge and then groups them with higher similarities into same cluster while those with lower correlations into different clusters. However, with limited *priori* knowledge (e.g., a small amount of labels or markers in original data), the performance of clustering intuitively can be enhanced or even improved.

B. Semi-Supervised Clustering

In practice, unsupervised clustering may result unexpected clusters that differs from user's expectation due to its lack of data labels. To solve this problem, semi-supervised clustering has been reported in [27], [28], which integrates *priori* knowledge such as label information [14], [29], pairwise constraints of linkage [30], and other additional information about their similarity or distance [31]–[34], to obtain better EEG clusters with what is known or assumed about the true categories. Since Wagstaff [12] put forward the concept of constrained clustering by embedding pairwise constraints of Must-Link and Cannot-Link into clustering algorithms, many related constrained studies have been reported, such as [35]–[37]. Moreover, a large number of classical clustering methods have been modified or extended with pairwise constraints, such as semi-supervised mean-shift clustering [38], weighted consensus semi-supervised clustering [39], constrained spectral clustering [40], [41], semi-supervised DenPeak clustering [42], [43], and fairness-constrained clustering [13], [44], [45], and other approaches [46], [47]. In detail, the must-link constraint constrains that pairwise samples must be in the same cluster, and the cannot-link constraint demands samples should be grouped into different clusters. Most of these semi-supervised clustering approaches mainly address one type of constraints or single constraint, but they seldom explore the integration of different constraints, such as the fusion of pairwise constraints and fairness constraint, to improve the efficacy of semi-supervised clustering. In other words, the performance those clustering approaches with single constraints mainly rely on the quality of such constraints, which is easily influenced by noises or incorrect single constraints.

III. Preliminaries

In a semi-supervised clustering notion, partially labeled EEG signals should be clustered into corresponding subgraphs in an EEG graph according to the prior known labels and their similarities (or weights). As a promising technique, spectral clustering utilizes graph Laplacian matrix to cluster. For an undirected weighted complete EEG graph $\boldsymbol{G_E} = (\boldsymbol{V}, \boldsymbol{S})$ where $\boldsymbol{V} = \{e_1, \cdots, e_n\}$ denotes the graph vertexes transformed by n EEG signals with length of m and $\boldsymbol{S} \in (\mathbb{R}^+ \cup 0)^{n \times n}$ defines the normalized weighted adjacency matrix constructed by pairwise non-negative similarities between any two EEG signals, where $s_{ij} \in \boldsymbol{S}$, based on Gaussian similarity, in the paper is defined as

$$s_{ij} = \exp^{-\frac{\|\boldsymbol{e}_i - \boldsymbol{e}_j\|_2^2}{2\sigma^2}}, \tag{1}$$

where σ is the kernel parameter. Particularly, $s_{ij} = s_{ji}$, where i, j denote EEG signal e_i and e_j ($\forall e_i, e_j \in \boldsymbol{V}$), respectively. Besides, the degree $d_{ii} \in \boldsymbol{D}$ of an EEG vertex e_i in $\boldsymbol{G_E}$ is defined as

$$d_i = d_{ii} = \sum_{j=1}^{n} s_{ij}. \tag{2}$$

Then, $\boldsymbol{D}$ denotes the degree matrix that is diagonal with degrees of every EEG vertex: $d_1, \cdots, d_n$. In the end, the EEG graph Laplacian $\boldsymbol{\mathcal{L}_{G_E}}$ is correspondingly defined as

$$\boldsymbol{\mathcal{L}_{G_E}} = \boldsymbol{D} - \boldsymbol{S}, \tag{3}$$

where $rank(\boldsymbol{\mathcal{L}_{G_E}}) = n - k$, and k is the multiplicity of eigenvalues of 0 for $\boldsymbol{\mathcal{L}_{G_E}}$, which indicates the number of connected subgraphs in $\boldsymbol{G_E}$, as introduced in [48]–[50].

IV. Semi-supervised EEG Clustering with Multiple Constraints

To address the challenging task of semi-supervised EEG clustering with limited *priori* knowledge and to make best use of such knowledge from partially labeled EEG data, we proposed a semi-supervised clustering method *ConsEEGc* with multiple constraints, including the label-transformed connectivity constraints that considers the connection or disconnection among partially labeled EEG data, compactness-and-scatter constraint that guarantees the intra-cluster compactness and inter-cluster scatter of EEG clusters, and fairness constraint that constrains the balance of EEG amount in clusters. The framework of our method *ConsEEGc* is briefly illustrated in Fig.1, which illustrates the main process to cluster EEG data with multiple constraints. In detail, *ConsEEGc* is conducted with a minimization optimization objective function, which integrates the pseudo label classifier learning, least-square error minimization of pseudo labels to clustering labels, with connectivity constraints, compactness-and-scatter constraint, and fairness constraint.

A. Optimal Pseudo Adjacency Matrix

Motivated by the Theorem introduced by Mohar [49] and Chung [50] that the multiplicity k of eigenvalue 0 of Laplacian matrix $\boldsymbol{\mathcal{L}_S}$ equals the number of connected components (i.e., clusters). Given an initial adjacency matrix of EEG similarity $\boldsymbol{M} \in \mathbb{R}^{n \times n}$, this paper, similarly to [36], tries to learn an optimal pseudo adjacency matrix $\boldsymbol{S}$ to approximate $\boldsymbol{M}$ with a minimization optimization strategy, i.e.,

$$\begin{aligned} &\min_{\boldsymbol{S} \in \mathbb{R}^+ \cup 0} \|\boldsymbol{S} - \boldsymbol{M}\|_F^2 \\ &s.t. \quad \boldsymbol{S}\mathbf{1}_n = \mathbf{1}_n, rank(\boldsymbol{\mathcal{L}_S}) = n - c, \end{aligned} \tag{4}$$

where c denotes the number of EEG clusters, i.e., the number of eigenvalues 0 of $\boldsymbol{\mathcal{L}_S}$, and $\mathbf{1}_n = [1, \cdots, 1]^T$.

B. Pseudo Labels for Partially Labeled EEG Signals

In this paper, we cluster EEG signals with their similarities, where similar EEG signals are probably clustered into same

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2022.3206330

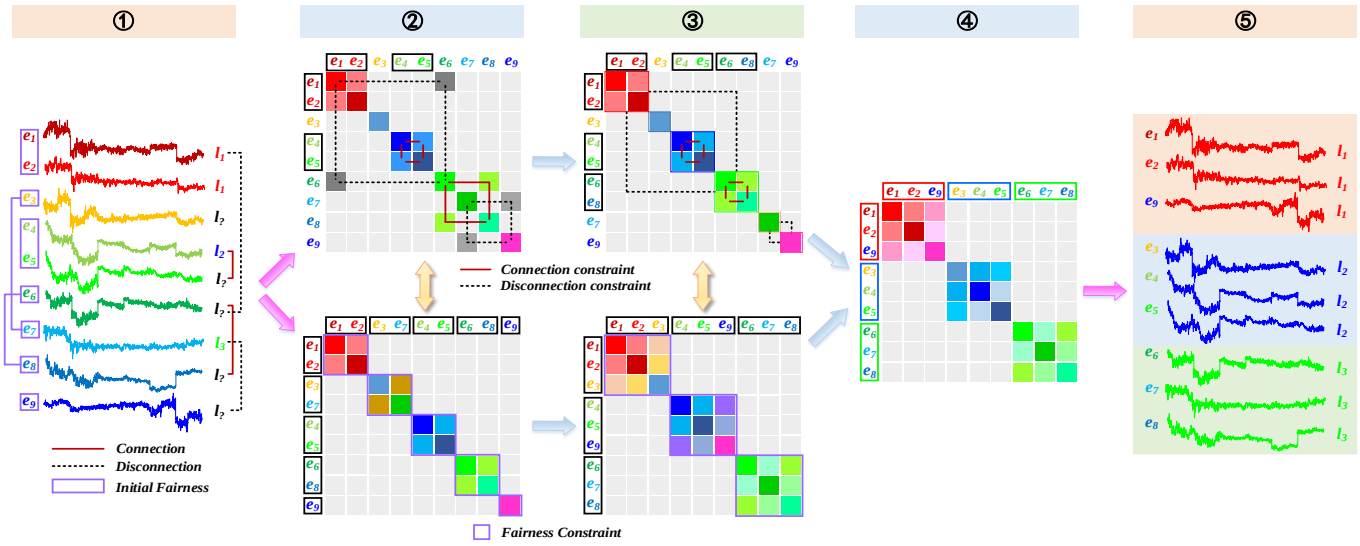


Fig. 1. The framework of the method. ① indicates the original EEG data with a *priori* knowledge such as label-transformed connection/disconnection information and number of EEG data in every category (or cluster) that is called fairness constraint in following introduction; ② shows the similarity adjacency matrix (i.e., spectral graph) of EEG data with original *priori* knowledge transformed constraints; ③ illustrates the clustering process with connectivity, compactness-scatter constraint, and fairness constraint transformed with limited *priori* knowledge from EEG data; ④ learns the locally optimal graph-based solution that holds the connectivity and fairness constraints; ⑤ outputs the learned EEG cluster labels.

category. Namely, EEG signals with high similarity are likely assigned with same pseudo labels when clustering. Suppose partially labeled EEG signals contain c classes, then the pseudo label matrix is defined as $\boldsymbol{L} \in \mathbb{R}^{c\times n}$. In fact, the pseudo label matrix can be also defined as a probability matrix that $\boldsymbol{L}_{iv}$ denotes the probability (pseudo label) of EEG e_i belongs to v^{th} category. For example, for two categories v_1 and v_2, if the pseudo label $\boldsymbol{L}_{iv_1} > \boldsymbol{L}_{iv_2}$ for EEG e_i, then e_i is clustered into category v_1, rather than v_2.

C. Regularized Least-Square Error Minimization

In the paper, it tries to partition similar EEG signals into same cluster with partial labels. Hence, the proposed method clusters incompletely labeled EEG signals by minimizing the regularized least-square error, where the optimal adjacency matrix, optimal pseudo labels, and pseudo label classifier are learnt simultaneously. Mathematically,

$$\begin{aligned} \min_{\boldsymbol{S},\boldsymbol{W},\boldsymbol{L}} & \|\boldsymbol{W}^T\boldsymbol{S} - \boldsymbol{L}\|_F^2 + \gamma\|\boldsymbol{W}\|_F^2 \\ s.t. \quad & \boldsymbol{W}^T\boldsymbol{S}\mathbf{1}_n = \mathbf{1}_n, \end{aligned} \tag{5}$$

where $\boldsymbol{W} \in \mathbb{R}^{n\times c}$ denotes the pseudo label classifier, $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ denotes the optimized EEG similarity adjacency matrix, and $\gamma > 0$ is regularization parameter.

D. Limited Label-Transformed Connectivity Constraints

For limited labeled EEG data, those, obviously, with different labels must be disconnected with each other when clustering, and those EEG with same labels must be categorized into same cluster together. In other words, EEG clustering must satisfy the constraints of partially priori knowledge, i.e., *connectivity constraints*.

In detail, those partially labeled EEG signals with different labels that holding *disconnection constraint* (similar to *Cannot-link*) must be assigned to different clusters.

Theorem 1. (***Disconnection Constraint***) *Given a nonnegative n-EEG-constructed graph $\boldsymbol{G}_{\boldsymbol{E}}$, if EEG signals e_i and e_j ($e_i, e_j \in \boldsymbol{E}$) hold disconnection constraint with labeling indicators that $l_i \neq l_j \in \boldsymbol{L}$, $\boldsymbol{L} \in \mathbb{R}^n$, there exists no paths between e_i and e_j such that*

$$\boldsymbol{L}\boldsymbol{\mathcal{L}}_{\boldsymbol{G}_{\boldsymbol{E}}}\boldsymbol{L}^T = 0.$$

That is, e_i and e_j must be clustered into different subgraphs $\boldsymbol{G}_{e_i}, \boldsymbol{G}_{e_j} \subset \boldsymbol{G}_{\boldsymbol{E}}, \boldsymbol{G}_{e_i} \cap \boldsymbol{G}_{e_j} = \varnothing$.

Proof: The proof is shown in Appendix A in the Supplementary file. ■

Similarly, those partially labeled EEG signals with same labels that holding *connection constraint* (similar to *Must-link*) must be assigned into same clusters when clustering.

Theorem 2. (***Connection Constraint***) *Given a nonnegative n-EEG-constructed graph $\boldsymbol{G}_{\boldsymbol{E}}$, if EEG signals e_i and e_j ($e_i, e_j \in \boldsymbol{E}$) hold connection constraint with same labeling indicator $l_i = l_j \in \mathbb{R}^n$. When $\boldsymbol{L}\boldsymbol{\mathcal{L}}_{\boldsymbol{G}_{\boldsymbol{E}}}\boldsymbol{L}^T = 0$, $\forall e_t \in \hat{\boldsymbol{E}}, \hat{\boldsymbol{E}} \subseteq \boldsymbol{E} \setminus \{e_i \cup e_j\}$ on the path between e_i and e_j must be in the same subgraph of $\boldsymbol{G}_{\boldsymbol{E}}$ that e_i and e_j locate in.*

Proof: The proof is presented in Appendix B in the Supplementary file. ■

With Theorem 1 and Theorem 2, EEG signals that connect to partially labeled EEG signals also must satisfy the connectivity constraints that labeled EEG signals hold.

Deduction 1. (***Affinity Connection Constraint***) *Given a*

nonnegative n-EEG-constructed graph $\boldsymbol{G_E}$, where EEG signals e_i and e_j ($e_i, e_j \in \boldsymbol{E}$) hold connection constraint, then $\forall e_t \in \hat{\boldsymbol{E}}, \hat{\boldsymbol{E}} \subseteq \boldsymbol{E} \setminus \{e_i, e_j\}$ that connects to either e_i or e_j must be in the same subgraph $\boldsymbol{G}_{e_i \cup e_j}$ of $\boldsymbol{G_E}$ that e_i and e_j in.

Proof: The proof is similar to Theorem 2. ■

Deduction 2. (***Affinity Disconnection Constraint***) *Given a nonnegative n-EEG-constructed graph $\boldsymbol{G_E}$, where EEG signals e_i and e_j ($e_i, e_j \in \boldsymbol{E}$) hold disconnection constraint, then $\forall e_t, e_q \in \hat{\boldsymbol{E}}, \hat{\boldsymbol{E}} \subseteq \boldsymbol{E} \setminus \{e_i, e_j\}, e_t \neq e_q$ where e_t connects to e_i while e_q connects to e_j, then e_t must be clustered into subgraph $\boldsymbol{G}_{e_i}$ that e_i in and e_q is in the same subgraph $\boldsymbol{G}_{e_j}$ that e_j in, such that $\boldsymbol{G}_{e_i} \cap \boldsymbol{G}_{e_j} = \varnothing$.*

Proof: This deduction can be proven with Theorem 1 and Deduction 1. ■

When clustering, the expected result is that partially labeled EEGs and unlabeled EEGs are partitioned into correct clusters that results in $\sum_k^c \boldsymbol{L}_k \boldsymbol{\mathcal{L}_S} \boldsymbol{L}_k^T = 0$. In practice, we learn the optimal EEG-constructed graph to cluster partially labeled EEG with connectivity constraints:

$$\min_{\boldsymbol{L}} \sum_{k}^{c} \boldsymbol{L}_k \boldsymbol{\mathcal{L}_S} \boldsymbol{L}_k^T. \tag{6}$$

With Ky Fan's Theorem [51], the connectivity constraints in Eq.(6) can be equivalently written as

$$\min_{\boldsymbol{L}} tr(\boldsymbol{L}_k \boldsymbol{\mathcal{L}_S} \boldsymbol{L}_k^T). \tag{7}$$

E. Compactness-and-Scatter Constraint

According to the notion of clustering, a good clustering approach will get intra-cluster compact and inter-cluster scattered results for unlabeled data. Therefore, to achieve the goal, our method clusters partially labeled EEG signals constrained with compactness and scatter by simultaneously minimizing the similarities of EEG signals among different clusters and the similarity difference in same clusters, see Eq.(8).

$$\min_{\boldsymbol{S}} \sum_{\substack{g \in \boldsymbol{C}_i, h \in \boldsymbol{C}_j \\ i \neq j}} s_{gh} + \frac{1}{2} \sum_{\substack{p,o,l,t \in \boldsymbol{C}_i \\ p \neq o, l \neq t}} (s_{po} - s_{lt})^2, \tag{8}$$

where s denotes EEG similarity and $\boldsymbol{C}$ is EEG cluster. The first term denotes scatter constraint and the second one is compactness constraint for partially EEG clustering. Actually, the difference of EEG similarities in the same cluster is relatively small, *i.e.*, $s_{po} - s_{lt}$ is close to 0. Hence, the second term in Eq.(8) can be ideally approximated to 0, *i.e.*,

$$\min \frac{1}{2} \sum_{\substack{p,o,l,t \in \boldsymbol{C}_i \\ p \neq o, l \neq t}} (s_{po} - s_{lt})^2 \to 0. \tag{9}$$

To the end, Eq.(8) can be simply relaxed as

$$\min_{\boldsymbol{S}} \sum_{\substack{g \in \boldsymbol{C}_i, h \in \boldsymbol{C}_j \\ i \neq j}} s_{gh}. \tag{10}$$

Then, Eq.(10) can be transformed to minimize the normalized Cut (*i.e.*, NCut [48]) that is defined as

$$\min NCut(\boldsymbol{E}) = \sum_{i=1}^{c} \frac{cut(\boldsymbol{C}_i, \boldsymbol{E} \backslash \boldsymbol{C}_i)}{vol(\boldsymbol{C}_i)}, \tag{11}$$

where $vol(\boldsymbol{C}_i) = \sum_{t \in \boldsymbol{C}_i} \sum_{j \in [1, \cdots, n]} s_{tj}$.

By defining a matrix $\boldsymbol{R} \in \mathbb{R}^{n \times c}$ as

$$R_{ik} = \begin{cases} \frac{1}{\sqrt{vol(\boldsymbol{C}_k)}}, i \in \boldsymbol{C}_k \\ 0, i \notin \boldsymbol{C}_k \end{cases}, \tag{12}$$

Equation (11) can be relaxed to solve the minimization of Eq.(13) with a relaxation requirement $\boldsymbol{R}^T \boldsymbol{D} \boldsymbol{R} = \boldsymbol{I}$:

$$\begin{aligned} &\min_{\boldsymbol{R}} tr(\boldsymbol{R}^T \boldsymbol{\mathcal{L}_S} \boldsymbol{R}) \\ &s.t. \quad \boldsymbol{R}^T \boldsymbol{D} \boldsymbol{R} = \boldsymbol{I}, \end{aligned} \tag{13}$$

where $\boldsymbol{D}$ is the degree matrix of $\boldsymbol{S}$ such that $d_i = \sum_{j \in [1, \cdots, n]} s_{ij}$. By substituting $\boldsymbol{R} = \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{U}$ such that $\boldsymbol{U} \in \mathbb{R}^{n \times c}$, Eq.(13) is relaxed to

$$\begin{aligned} &\min_{\boldsymbol{U}} tr(\boldsymbol{U}^T \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{\mathcal{L}_S} \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{U}) \\ &s.t. \quad \boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{I}. \end{aligned} \tag{14}$$

where $\boldsymbol{U}$ can be obtained by computing k smallest eigenvalues and corresponding k eigenvectors of $\boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{\mathcal{L}_S} \boldsymbol{D}^{-\frac{1}{2}}$.

F. Fairness Constraint

To achieve a fairly balanced clustering for partially labeled EEG signals, we also bring in the *fairness* to constrain the clustering in the paper. According to the notion of fairness defined by Chierichetti et al. [52], the fairness of EEG clustering is similarly defined as follows.

Definition 1. (***Fairness***) *For a clustering $\boldsymbol{\mathcal{C}} = \cup_{i=1}^{c} \boldsymbol{C}_i$ of EEG set $\boldsymbol{E} = \cup_{t=1}^{v} \boldsymbol{E}_t$ such that $\boldsymbol{C}_i \neq \varnothing, \cap_{i=1}^{c} \boldsymbol{C}_i = \varnothing$, where v, c denote the number of EEG categories (i.e., EEG classes) and number of clusters, respectively, its fairness is defined as*

$$fairness(\boldsymbol{\mathcal{C}}) = \min_{i \in [1, \cdots, c]} \min_{m \neq n \in [1, \cdots, v]} \frac{|\boldsymbol{E}_m \cap \boldsymbol{C}_i|}{|\boldsymbol{E}_n \cap \boldsymbol{C}_i|} \in [0, 1]. \tag{15}$$

With the definition of fairness, the EEG clustering in the paper holds Lemma 1.

Lemma 1. *For any clustering $\boldsymbol{\mathcal{C}}$ of EEG set $\boldsymbol{E}$, $fairness(\boldsymbol{\mathcal{C}}) \leq fairness(\boldsymbol{E})$.*

Proof: The proof is shown in Appendix C in the Supplementary file. ■

As noted in Lemma 1, we try to cluster EEG signals with fairness that approximately equals to original EEG set. To achieve this, we add *fairness constraint* on EEG clustering in the paper, as introduced in Lemma 2.

Lemma 2. (***Fairness Constraint***) *For a v-category (i.e., v-class) EEG set $\boldsymbol{E} = \cup_{t=1}^{v} \boldsymbol{E}_t$, its clustering $\boldsymbol{\mathcal{C}} = \cup_{i=1}^{c} \boldsymbol{C}_i$ holds a linear fairness constraint with a binary membership indicator $b_{ot} = \begin{cases} 1, o \in \boldsymbol{E}_t \\ 0, o \notin \boldsymbol{E}_t \end{cases}$ that $\frac{|\boldsymbol{E}_t \cap \boldsymbol{C}_i|}{|\boldsymbol{C}_i|} = \frac{|\boldsymbol{E}_t|}{|\boldsymbol{E}|}$ for $t \in [1, \cdots, v]$ iff*

$\sum_{o=1}^{|\boldsymbol{E}|}\left(b_{ot}-\frac{|\boldsymbol{E}_t|}{|\boldsymbol{E}|}\right)R_{ot}=0$ *for* $t\in[1,\cdots,v-1]$.

Proof: The proof is shown in Appendix D in the Supplementary file. ■

With Lemma 2, the compactness-and-scatter constraint defined in Eq.(13) can be transformed to the fairness constraint that is reformulated to

$$\begin{aligned}&\min_{\boldsymbol{R}} tr(\boldsymbol{R}^T\boldsymbol{\mathcal{L}_S}\boldsymbol{R})\\ &s.t.\quad \boldsymbol{R}^T\boldsymbol{D}\boldsymbol{R}=\boldsymbol{I},\boldsymbol{B}^T\boldsymbol{R}=\boldsymbol{0}^{(v-1)\times c},\end{aligned}\tag{16}$$

where $\boldsymbol{B}\in\mathbb{R}^{n\times(v-1)}$, subject to $rank(\boldsymbol{B})=v-1$, is the binary membership indicator matrix constructed by column vectors of Eq.(17) with respect to corresponding position (o,t),

$$\boldsymbol{B}=\{colomn\quad vectors|b_{ot}-\frac{|\boldsymbol{E}_t|}{n}\},\tag{17}$$

where $t\in[1,\cdots,v-1], o\in[1,\cdots,n]$, $n=|\boldsymbol{E}|, v, c$ denote the number of EEG signals, number of EEG categories, and number of partitioned EEG clusters, respectively. As introduced in [13] that if $v-1<n-c$, there does not exist solutions for Eq.(16), so we define a matrix $\boldsymbol{Q}\in\mathbb{R}^{n\times(n-v+1)}$ as

$$\boldsymbol{Q}=orth(\ker(\boldsymbol{B}^T)),\tag{18}$$

where $orth()$ defines the orthogonal basis of a matrix, and $\ker(\boldsymbol{B}^T)$ is defined as

$$\ker(\boldsymbol{B}^T)=\{\boldsymbol{x}\in\mathbb{R}^n:\boldsymbol{B}^T\boldsymbol{x}=\boldsymbol{0}\}.\tag{19}$$

By substituting $\boldsymbol{R}=\boldsymbol{Q}\boldsymbol{X}$ with $\boldsymbol{X}\in\mathbb{R}^{(n-v+1)\times c}$, Eq,(16) can be transformed as

$$\begin{aligned}&\min_{\boldsymbol{X}} tr(\boldsymbol{X}^T\boldsymbol{Q}^T\boldsymbol{\mathcal{L}_S}\boldsymbol{Q}\boldsymbol{X})\\ &s.t.\quad \boldsymbol{X}^T\boldsymbol{Q}^T\boldsymbol{D}\boldsymbol{Q}\boldsymbol{X}=\boldsymbol{I}.\end{aligned}\tag{20}$$

Since $\boldsymbol{Q}^T\boldsymbol{D}\boldsymbol{Q}$ is positive semidefinite, we then define a positive semidefinite $\boldsymbol{Z}\in\mathbb{R}^{(n-v+1)\times(n-v+1)}$ such that

$$\boldsymbol{Z}^2=\boldsymbol{Q}^T\boldsymbol{D}\boldsymbol{Q}.\tag{21}$$

Subsequently, by substituting $\boldsymbol{X}=\boldsymbol{Z}^{-1}\boldsymbol{U}$ with $\boldsymbol{U}\in\mathbb{R}^{(n-v+1)\times c}$, then the problem of Eq.(20) can be reformulated to Eq.(22) that aims to solve NCut with fairness constraint.

$$\begin{aligned}&\min_{\boldsymbol{U}} tr(\boldsymbol{U}^T\boldsymbol{Z}^{-1}\boldsymbol{Q}^T\boldsymbol{\mathcal{L}_S}\boldsymbol{Q}\boldsymbol{Z}^{-1}\boldsymbol{U})\\ &s.t.\quad \boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I}.\end{aligned}\tag{22}$$

The solution $\boldsymbol{U}$ of Eq.(22) can be solved by computing the corresponding eigenvectors as columns of $\boldsymbol{Z}^{-1}\boldsymbol{Q}^T\boldsymbol{\mathcal{L}_S}\boldsymbol{Q}\boldsymbol{Z}^{-1}$ corresponding to its c smallest eigenvalues, i.e.,

$$\boldsymbol{U}=\frac{\gamma_5}{2}\cdot c_{eigenvectors}(\boldsymbol{Z}^{-1}\boldsymbol{Q}^T\boldsymbol{\mathcal{L}_S}\boldsymbol{Q}\boldsymbol{Z}^{-1}).\tag{23}$$

To the end, $\boldsymbol{R}$ with fairness constraint in Eq.(16) is solved by

$$\boldsymbol{R}=\boldsymbol{Q}\boldsymbol{Z}^{-1}\boldsymbol{U}.\tag{24}$$

G. Problem Formulation

Our method clusters incompletely labeled EEG signals by a minimization optimization objective function that integrates pseudo adjacency matrix, pseudo label classifier, least-square error minimization, connectivity constraints, compactness-and-scatter constraints, and fairness constraint. Mathematically, the problem we are addressing is defined as

$$\begin{aligned}\min_{\boldsymbol{R},\boldsymbol{W},\boldsymbol{S},\boldsymbol{L}}\boldsymbol{\mathcal{F}}=&\frac{\gamma_1}{2}\|\boldsymbol{S}-\boldsymbol{M}\|_F^2+\frac{\gamma_2}{2}\|\boldsymbol{W}^T\boldsymbol{S}-\boldsymbol{L}\|_F^2\\&+\frac{\gamma_3}{2}\|\boldsymbol{W}\|_F^2+\frac{\gamma_4}{2}tr(\boldsymbol{L}\boldsymbol{\mathcal{L}_S}\boldsymbol{L}^T)\\&+\frac{\gamma_5}{2}tr(\boldsymbol{R}^T\boldsymbol{\mathcal{L}_S}\boldsymbol{R})\\ s.t.\quad rank(\boldsymbol{\mathcal{L}_S})=n-c,&\boldsymbol{S}\boldsymbol{1}_n=\boldsymbol{1}_n,\boldsymbol{W}^T\boldsymbol{S}\boldsymbol{1}_n=\boldsymbol{1}_n,\\ \boldsymbol{R}=\boldsymbol{Q}\boldsymbol{Z}^{-1}&\boldsymbol{U},\boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I},\end{aligned}\tag{25}$$

where $\gamma_1,\gamma_2,\gamma_3,\gamma_4,\gamma_5$ are weight parameters. The first term aims to learn an optimal pseudo adjacency matrix to achieve c EEG clusters from $\boldsymbol{\mathcal{L}_S}$, the seconde and third terms try to learn the optimal class label of EEG; The fourth term constrains the connectivity of partially labeled EEG signals and it also minimizes the connectivity degrees from one EEG to another; The last term that is relaxed to a *NCut* with fairness constraint constrains the intra-cluster compactness and inter-cluster scatter among different clusters and same clusters.

V. The Proposed Algorithm: ConsEEGc

In this section, we introduce the proposed algorithm *ConsEEGc*, followed by its convergence analysis, initialization and time complexity.

A. ConsEEGc

As introduced in the former section, the objective function (i.e., Eq.(25)) is not globally convex, so we apply gradient descent to compute $\boldsymbol{R},\boldsymbol{W},\boldsymbol{S},\boldsymbol{L}$ in a way that we update every single variable by fixing other variables, and it finally leads to the algorithm *ConsEEGc*, see **Algorithm 1**.

1) Updating $\boldsymbol{R}$*:* We update $\boldsymbol{R}$ by fixing $\boldsymbol{L}$, $\boldsymbol{W}$, and $\boldsymbol{S}$. Then the objective function Eq.(25) is degenerated to

$$\begin{aligned}&\min_{\boldsymbol{R}}\boldsymbol{\mathcal{F}}(\boldsymbol{R})=\frac{\gamma_5}{2}tr(\boldsymbol{R}^T\boldsymbol{\mathcal{L}_S}\boldsymbol{R})\\ &s.t.\quad \boldsymbol{B}^T\boldsymbol{R}=^{(v-1)\times c},\boldsymbol{U}^T\boldsymbol{U}=\boldsymbol{I}.\end{aligned}\tag{26}$$

where $\boldsymbol{R}=\boldsymbol{Q}\boldsymbol{Z}^{-1}\boldsymbol{U}$.

The derivative of Eq.(26) *w.r.t* $\boldsymbol{R}$ is

$$\frac{\partial\boldsymbol{\mathcal{F}_R}}{\partial\boldsymbol{R}}=\gamma_5\boldsymbol{\mathcal{L}_S}\boldsymbol{R}.\tag{27}$$

Then, the $\boldsymbol{R}_{t+1}$ at time $t+1$ can be updated by

$$\boldsymbol{R}_{t+1}=\boldsymbol{R}_t-\eta\frac{\partial\boldsymbol{\mathcal{F}_R}}{\partial\boldsymbol{R}}.\tag{28}$$

where η is the well-determined learning rate.

2) Updating $\boldsymbol{W}$*:* When fixing $\boldsymbol{L}$, $\boldsymbol{S}$ and $\boldsymbol{R}$, $\boldsymbol{W}$ is updated. Then the objective function Eq.(25) can be degenerated to

$$\min_{\boldsymbol{W}}\boldsymbol{\mathcal{F}}(\boldsymbol{W})=\frac{\gamma_2}{2}\|\boldsymbol{W}^T\boldsymbol{S}-\boldsymbol{L}\|_F^2+\frac{\gamma_3}{2}\|\boldsymbol{W}\|_F^2.\tag{29}$$

The derivative of Eq.(29) *w.r.t* $\boldsymbol{W}$ is

$$\begin{aligned}\frac{\partial\boldsymbol{\mathcal{F}_W}}{\partial\boldsymbol{W}}&=\gamma_2\boldsymbol{S}(\boldsymbol{W}^T\boldsymbol{S}-\boldsymbol{L})^T+\gamma_3\boldsymbol{W}\\&=(\gamma_2\boldsymbol{S}\boldsymbol{S}^T+\gamma_3\boldsymbol{I})\boldsymbol{W}-\gamma_2\boldsymbol{S}\boldsymbol{L}^T.\end{aligned}\tag{30}$$

Then, the $\boldsymbol{W}_{t+1}$ at time $t+1$ can be updated by

$$\boldsymbol{W}_{t+1} = \boldsymbol{W}_t - \eta \frac{\partial \boldsymbol{\mathcal{F}}_{\boldsymbol{W}}}{\partial \boldsymbol{W}}. \tag{31}$$

3) Updating $\boldsymbol{S}$: Similarly, we update $\boldsymbol{S}$ by fixing $\boldsymbol{L}$, $\boldsymbol{W}$ and $\boldsymbol{U}$. Then the objective function Eq.(25) is degenerated to

$$\begin{aligned}\min_{\boldsymbol{S}} \boldsymbol{\mathcal{F}}(\boldsymbol{S}) = &\frac{\gamma_1}{2}\|\boldsymbol{S}-\boldsymbol{M}\|_F^2 + \frac{\gamma_2}{2}\|\boldsymbol{W}^T\boldsymbol{S}-\boldsymbol{L}\|_F^2 \\ &+ \frac{\gamma_4}{2} tr(\boldsymbol{L}\boldsymbol{\mathcal{L}}_{\boldsymbol{S}}\boldsymbol{L}^T) + \frac{\gamma_5}{2} tr(\boldsymbol{R}^T\boldsymbol{\mathcal{L}}_{\boldsymbol{S}}\boldsymbol{R}).\end{aligned} \tag{32}$$

Its derivative of Eq.(32) *w.r.t* $\boldsymbol{S}$ is

$$\begin{aligned}\frac{\partial \boldsymbol{\mathcal{F}}_{\boldsymbol{S}}}{\partial \boldsymbol{S}} = &\gamma_1(\boldsymbol{S}-\boldsymbol{M}) + \gamma_2\boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{S}-\boldsymbol{L}) \\ &+ \frac{\gamma_4}{2}\boldsymbol{L}^T\boldsymbol{L}\frac{\partial \boldsymbol{\mathcal{L}}_{\boldsymbol{S}}}{\partial \boldsymbol{S}} + \frac{\gamma_5}{2}\boldsymbol{R}\boldsymbol{R}^T\frac{\partial \boldsymbol{\mathcal{L}}_{\boldsymbol{S}}}{\partial \boldsymbol{S}},\end{aligned} \tag{33}$$

where $\boldsymbol{\mathcal{L}}_{\boldsymbol{S}} = \boldsymbol{D} - \boldsymbol{S}$ and $d_{ii} = \sum_j^n s_{ij}$. Then, to simplify Eq.(33), we get

$$\frac{\partial \boldsymbol{\mathcal{L}}_{\boldsymbol{S}}}{\partial \boldsymbol{S}} = \frac{\partial(\boldsymbol{D}-\boldsymbol{S})}{\partial \boldsymbol{S}} = \frac{\partial \boldsymbol{D}}{\partial \boldsymbol{S}} - \boldsymbol{I} = \frac{\partial \boldsymbol{I}}{\partial \boldsymbol{S}} - \boldsymbol{I} = \boldsymbol{0} - \boldsymbol{I}, \tag{34}$$

since $\boldsymbol{S}$ is subject to $\boldsymbol{S}\boldsymbol{1}_n = \boldsymbol{1}_n$ and $d_i = \sum_j s_{ij}$, then $\boldsymbol{D} = diag(\boldsymbol{S}\boldsymbol{1}_n) = \boldsymbol{I}$ for simplification. To the end, Eq.(33) can be reformulated as

$$\begin{aligned}\frac{\partial \boldsymbol{\mathcal{F}}_{\boldsymbol{S}}}{\partial \boldsymbol{S}} = &(\gamma_1\boldsymbol{I} + \gamma_2\boldsymbol{W}\boldsymbol{W}^T)\boldsymbol{S} - \gamma_1\boldsymbol{M} - \gamma_2\boldsymbol{W}\boldsymbol{L} \\ &- \frac{\gamma_4}{2}\boldsymbol{L}^T\boldsymbol{L} - \frac{\gamma_5}{2}\boldsymbol{R}\boldsymbol{R}^T.\end{aligned} \tag{35}$$

Subsequently, the $\boldsymbol{S}_{t+1}$ at time $t+1$ can be updated by

$$\boldsymbol{S}_{t+1} = \boldsymbol{S}_t - \eta \frac{\partial \boldsymbol{\mathcal{F}}_{\boldsymbol{S}}}{\partial \boldsymbol{S}}. \tag{36}$$

4) Updating $\boldsymbol{L}$: With the gradient descent strategy, we update $\boldsymbol{L}$ by fixing $\boldsymbol{W}$, $\boldsymbol{S}$ and $\boldsymbol{U}$. Then, the objective function Eq.(25) can be degenerated to

$$\min_{\boldsymbol{L}} \boldsymbol{\mathcal{F}}(\boldsymbol{L}) = \frac{\gamma_2}{2}\|\boldsymbol{W}^T\boldsymbol{S}-\boldsymbol{L}\|_F^2 + \frac{\gamma_4}{2} tr(\boldsymbol{L}\boldsymbol{\mathcal{L}}_{\boldsymbol{S}}\boldsymbol{L}^T). \tag{37}$$

The derivative of Eq.(37) *w.r.t* $\boldsymbol{L}$ is defined as

$$\begin{aligned}\frac{\partial \boldsymbol{\mathcal{F}}_{\boldsymbol{L}}}{\partial \boldsymbol{L}} &= -\gamma_2(\boldsymbol{W}^T\boldsymbol{S}-\boldsymbol{L}) + \boldsymbol{\gamma_4}\boldsymbol{L}\boldsymbol{\mathcal{L}}_{\boldsymbol{S}} \\ &= \boldsymbol{L}(\gamma_4\boldsymbol{\mathcal{L}}_{\boldsymbol{S}} + \gamma_2\boldsymbol{I}) - \gamma_2\boldsymbol{W}^T\boldsymbol{S},\end{aligned} \tag{38}$$

where $\boldsymbol{I}$ denotes the identity matrix.

To the end, the $\boldsymbol{L}_{t+1}$ at $t+1$ can be updated by

$$\boldsymbol{L}_{t+1} = \boldsymbol{L}_t - \eta \frac{\partial \boldsymbol{\mathcal{F}}_{\boldsymbol{L}}}{\partial \boldsymbol{L}}. \tag{39}$$

Algorithm 1 ConsEEGc

Input:

$\boldsymbol{E}^{n\times m} = \boldsymbol{E}_{\boldsymbol{L}} \cup \boldsymbol{E}_{\tilde{\boldsymbol{L}}} = \cup_i^v \boldsymbol{E}_i$: n EEG trials of v categories with length of m, where $\boldsymbol{E}_{\boldsymbol{L}}$, $\boldsymbol{E}_{\tilde{\boldsymbol{L}}}$ denote labeled EEG signals and unlabeled ones, respectively

$\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5$: Weight parameters

η: Learning rate

c: Number of clusters

Output:

$\boldsymbol{L}^{*c\times n}$: Optimal pseudo labels for partially labeled $\boldsymbol{E}$

Compute $\boldsymbol{M}$ of $\boldsymbol{E}$ with Eq.(1)
Compute $\boldsymbol{D}_{\boldsymbol{M}}$, $\boldsymbol{\mathcal{L}}_{\boldsymbol{M}}$ with Eq.(2), Eq.(3), respectively
Initialize $\boldsymbol{R}_0, \boldsymbol{W}_0, \boldsymbol{S}_0, \boldsymbol{L}_0$:
  Compute $\boldsymbol{R}_0$ by Eqs.(17)–(19), Eq.(21), Eqs.(23)–(24)
  Construct $\boldsymbol{S}_0$ with connectivity constraints by Eq.(40) and then normalize it such that $\boldsymbol{S}_0\boldsymbol{1}_n = \boldsymbol{1}_n$
  Construct $\boldsymbol{L}_0 = [\boldsymbol{L}_{label}; \tilde{\boldsymbol{L}}]$, where $\tilde{\boldsymbol{L}}$ is computed with k-means on unlabeled EEG $\boldsymbol{E}_{\tilde{L}}$
  Compute $\boldsymbol{W}_0$ with Eq.(41), based on $\boldsymbol{L}_0$ and $\boldsymbol{M}$
while not converging **do**
  Compute $\boldsymbol{D}_{\boldsymbol{S}_t}, \boldsymbol{\mathcal{L}}_{\boldsymbol{S}_t}$:
    $\boldsymbol{D}_{\boldsymbol{S}_t} \leftarrow diag(d_i)$ with Eq.(2)
    $\boldsymbol{\mathcal{L}}_{\boldsymbol{S}_t} \leftarrow \boldsymbol{D}_{\boldsymbol{S}_t} - \boldsymbol{S}_t$ with Eq.(3)
  Update $\boldsymbol{R}_{t+1}$, $\boldsymbol{W}_{t+1}$, $\boldsymbol{S}_{t+1}$, $\boldsymbol{L}_{t+1}$:
    $\boldsymbol{R}_{t+1} \leftarrow \boldsymbol{R}_t - \eta\frac{\partial \boldsymbol{\mathcal{F}}_{\boldsymbol{R}}}{\partial \boldsymbol{R}}$ with Eqs.(27) and Eq.(28)
    $\boldsymbol{W}_{t+1} \leftarrow \boldsymbol{W}_t - \eta\frac{\partial \boldsymbol{\mathcal{F}}_{\boldsymbol{W}}}{\partial \boldsymbol{W}}$ with Eq.(30) and Eq.(31)
    $\boldsymbol{S}_{t+1} \leftarrow \boldsymbol{S}_t - \eta\frac{\partial \boldsymbol{\mathcal{F}}_{\boldsymbol{S}}}{\partial \boldsymbol{S}}$ with Eqs.(33)–(36)
    $\boldsymbol{L}_{t+1} \leftarrow \boldsymbol{L}_t - \eta\frac{\partial \boldsymbol{\mathcal{F}}_{\boldsymbol{L}}}{\partial \boldsymbol{L}}$ with Eq.(38) and Eq.(39)
  $t \leftarrow t+1$
end while
return $\boldsymbol{U}^* = \boldsymbol{U}_{t+1}, \boldsymbol{L}^* = \boldsymbol{L}_{t+1}, \boldsymbol{S}^* = \boldsymbol{S}_{t+1}, \boldsymbol{W}^* = \boldsymbol{W}_{t+1}$

B. Convergence Analysis

The proposed algorithm *ConsEEGc* exploits the gradient descent strategy to solve the multi-variable optimization objective function with respect to optimal similarity adjacency matrix $\boldsymbol{S}$, pseudo label classifier $\boldsymbol{W}$, and optimal pseudo labels $\boldsymbol{L}$, that can converge to a local optima with a well-determined learning rate η.

Theorem 3. *The proposed gradient descent algorithm ConsEEGc converges to a local optima with a well-determined learning rate η.*

Proof: The proof is presented in Appendix E in the Supplementary file. ■

C. Initialization

As introduced above that the algorithm, *ConsEEGc*, converges to its local optima, the well-organized initialization probably yields a good clustering. Here we introduce the initialization of $\boldsymbol{R}_0, \boldsymbol{W}_0, \boldsymbol{S}_0, \boldsymbol{L}_0$.

Firstly, $\boldsymbol{R}_0$ can be initialized as $\boldsymbol{R}_0 = \boldsymbol{Q}\boldsymbol{Z}^{-1}\boldsymbol{U}$, see Eq.(24), where $\boldsymbol{U}$ can be achieved by c eigenvectors of $\boldsymbol{Z}^{-1}\boldsymbol{Q}^T\boldsymbol{\mathcal{L}}_{\boldsymbol{M}}\boldsymbol{Q}\boldsymbol{Z}^{-1}$ as columns, corresponding to the first c smallest eigenvalues, see Eqs.(17–19), Eq.(21), and Eq.(23). In the paper, we try to learn the nonnegative and normalized pseudo adjacent matrix $\boldsymbol{S}$ to approximate the original ground-truth adjacent similarity matrix $\boldsymbol{M}$, such that $\sum_j s_{ij} = 1$. In the paper, we simply initialize $\boldsymbol{S}_0$ by Eq.(40) and then normalize it such that $\boldsymbol{S}_0\boldsymbol{1}_n = \boldsymbol{1}_n$, where $\boldsymbol{1}_n = [1, \cdots, 1]_n^T$.

$$s_{ij} = \begin{cases} 1, & e_i, e_j \in \boldsymbol{E} \quad \text{with connection constraint,} \\ 0, & e_i, e_j \in \boldsymbol{E} \quad \text{with disconnection constraint,} \\ \dfrac{1}{|\boldsymbol{E}|}, & \text{otherwise.} \end{cases} \tag{40}$$

To initialize $\boldsymbol{L}_0$, the class labels of labeled EEGs $\boldsymbol{E_L}$ are set as their original ones and unlabeled ones are set by applying k-means on unlabeled EEG data $\boldsymbol{E}_{\tilde{\boldsymbol{L}}}$, then $\boldsymbol{L}_0 = \boldsymbol{L}_{label} \cup \tilde{\boldsymbol{L}}$. Subsequently, we simply initialize $\boldsymbol{W}_0$ based on the initialization of $\boldsymbol{L}_0$ and $\boldsymbol{M}$:

$$\boldsymbol{W}_0 = (\boldsymbol{L}_0 \boldsymbol{M}^{-1})^T. \tag{41}$$

With the initialization of such variables, the gradient descent algorithm, *ConsEEGc*, iteratively learns the local optima with a well-determined learning rate η.

D. Time Complexity

As introduced in Algorithm 1, we learn the local optima of $\boldsymbol{W}, \boldsymbol{S}, \boldsymbol{U}, \boldsymbol{L}$ with a well-determined learning rate η during many iterations that contributes to the time consumption. Here, we discuss the time complexity of Algorithm 1 in detail.

To compute $\boldsymbol{M}$ (Line 1), $\boldsymbol{D_M}$ and $\boldsymbol{\mathcal{L}_M}$ (see Line 2), it takes $O(n)$, respectively. The initialization for $\boldsymbol{R}_0, \boldsymbol{W}_0, \boldsymbol{S}_0, \boldsymbol{L}_0$, see Lines 4–7, requires $O\big(n^2(n-v+1) + n(n-v+1)^2 + (n-v+1)^3 + nc(n-v+1)\big) = O(n^3)$ (Line 4), $O(n)$ (Line 5), $O(n^2)$ (Line 6), and $O(cn^2)$ (Line 7), respectively, where n, c, v denote the total number of EEG signals, number of clusters, and number of EEG categories, respectively. With the initialization, Algorithm 1 updates $\boldsymbol{R}, \boldsymbol{W}, \boldsymbol{S}, \boldsymbol{L}$ iteratively, along with updating $\boldsymbol{D_S}$ and $\boldsymbol{\mathcal{L}_S}$. In each iteration, the computation of $\boldsymbol{D}_{\boldsymbol{S}_t}$ and $\boldsymbol{\mathcal{L}}_{\boldsymbol{S}_t}$, see Line 10–11, takes $O(n)$, respectively. Subsequently, it takes $O\big(cn(n-v+1) + c(n-v+1)^2 + cn^2 + n^2(n-v+1) + n(n-v+1)^2\big) = O(n^3)$ to update $\boldsymbol{R}$, see Line 13, and $O(n^3 + cn^2) = O(n^3)$ for updating $\boldsymbol{W}$ as Line 14 shows, since $c \ll n, v \ll n$. When updating $\boldsymbol{S}$ and $\boldsymbol{L}$, it takes $O(n^3 + cn^2) = O(n^3)$ and $O(cn^2)$, respectively. In a result, the time complexity of Algorithm 1 with $iter$ iterations converging to local optima is $O\big(5n + n^3 + cn^2 + iter \cdot (2n + 3n^3 + cn^2)\big) = O(iter \cdot n^3)$.

VI. Experiment And Discussion

This section presents the experimental results of *ConsEEGc* comparing with several state-of-the-art unsupervised and semi-supervised clustering approaches on different types of real-world EEG datasets.

A. EEG Datasets

As Table I shows, we assessed 12 EEG datasets [1] for our approach *ConsEEGc*, including (1) slow cortical potentials (SCPs, a type of EEG data), one set recorded from a healthy subject, i.e., II_Ia, and the other from a patient with amyotrophic lateral sclerosis (ALS; or motor-neuron disease) (i.e., II_Ib); A cross-healthy-patient EEG dataset is also constructed with datasets Ia and Ib, i.e., II_Ia_Ib; (2) Three-class mental imagery EEG with pre-computed features from two healthy subjects, i.e., III_V_s1 and III_V_s2; (3) Four-class motor imagery EEG from three healthy subjects, i.e., IV_2a_s1, IV_2a_s2 and IV_2a_s3; (4) Two-class simple motor imagery EEG datasets from two healthy subjects, i.e., IV_2b_s1 and IV_2b_s2; (5) Four-class hand movement EEG datasets from two healthy subjects, i.e., IV_3_s1 and IV_3_s2. Importantly, as the EEG data were all originally labeled, we deleted these labels prior to performing clustering algorithms. Besides, all the EEG data are z-normalized and randomly shuffled in rows.

[1]The preprocessed EEG datasets are publicly available at https://github.com/Jackie-Day/EEG-data-and-descriptions.

B. Evaluation Criteria

We evaluated the semi-supervised clustering performance of the proposed method with five criteria, *Accuracy* (*i.e.*, *Rand index*), *normalized mutual information*, *adjusted rand index*, *F-score*, and *Fleiss' kappa*.

• ***Accuracy (Acc)*** (*i.e.*, *Rand index* [53]) evaluates clustering performance according to the percentage of correct clustered numbers over total elements, which is mathematically defined as $Acc = \frac{TP+TN}{TP+TN+FP+FN}$, where TP, TN, FP, FN denote the number of true positives, true negatives, false positives, and false negatives, respectively.

• ***Normalized Mutual Information (NMI)*** [54] measures the quality of clusters, which is mathematically defined as $NMI = \frac{\sum_i^v \sum_j^v |\boldsymbol{C}_i \cap \boldsymbol{C}_j| \log \frac{N \cdot |\boldsymbol{C}_i \cap \boldsymbol{C}_j|}{|\boldsymbol{C}_i||\boldsymbol{C}_j|}}{\sqrt{(\sum_i^v |\boldsymbol{C}_i| \log \frac{|\boldsymbol{C}_i|}{N})(\sum_j^v |\boldsymbol{C}_j| \log \frac{|\boldsymbol{C}_j|}{N})}}$, where $|\boldsymbol{C}_i|$ and $|\boldsymbol{C}_j|$ denote the number of EEG signals in Cluster $\boldsymbol{C}_i$ and $\boldsymbol{C}_j$, respectively; v and N denote the number of categories and the total number of EEG signals, respectively.

• ***Adjusted Rand Index (ARI)*** [55] measures the agreement between two partitions: one given by the cluster process and the other defined by external criteria. Namely, $ARI = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{n}{2}}$, where n_{ij}: number of objects in both class u_i and cluster v_i; a_i: number of objects in class u_i and b_j: number of objects in cluster v_j; n: total number of objects.

• ***F-score*** [56] unequally weighs the false positives FP and false negatives FN in *RI* with a scale parameter $\beta \geq 0$ on *recall*. Mathematically, $F - score = \frac{(1+\beta^2)pr}{\beta^2 p + r}$, where *precision* $p = \frac{TP}{TP+FP}$, *recall* $r = \frac{TP}{TP+FN}$, TP is true positives and commonly $\beta = 1$.

• ***Fleiss' kappa*** [57] is a statistical measure for assessing the coherence of decision ratings among classes. 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 $\overline{P_e} = \sum_{j=1}^k (\frac{1}{Nn} \sum_{i=1}^N n_{ij})^2$.

Higher *NMI*, *ARI*, *F-score*, and *kappa* correspond to better performance of EEG clustering approaches.

C. Baselines

To illustrate the efficacy of *ConsEEGc*, we compare it to several state-of-the-art EEG time series clustering algorithms, such as classic unsupervised clustering, spectral clustering, semi-supervised clustering with constraints, and fair clustering. These baselines are in detail introduced as follows.

k-means++ [15]: As a variant of classic k-means, k-means++ modifies the initialization of the first cluster center by replacing initializing randomly with probability, and then it performs the standard k-means to cluster.

KMMC [16]: K-Multiple-Means Clustering, unlike the classic k-means-type algorithms, groups EEG time series with

TABLE I
EEG DATASETS.

Dataset	Descriptions	# EEG Trials	Length of EEG Signals	# EEG Channels	# Clusters	Category Ratio
II_Ia	SCPs from a healthy subject	268	5377	6	2	133:135
II_Ib	SCPs from an ALS patient	200	8065	7	2	100:100
II_Ia_Ib	Mixed SCPs of dataset Ia (from a healthy subject) and Ib (from an ALS patient)	468	5377	6	4	100:100:133:135
III_V_s1	Mental imagery EEG with precomputed features of left hand, right hand, and word association	3488	97	8	3	1440:928:1120
III_V_s2		3472				1472:880:1120
IV_2a_s1	Motor imagery EEG of left hand, right hand, both feet, and tongue	288	6887	22	4	72:72:72:72
IV_2a_s2						
IV_2a_s3						
IV_2b_s1	Simple motor imagery EEG of left hand and right hand	120	940	3	2	60:60
IV_2b_s2						
IV_3_s1	Hand movement EEG in directions of left, right, forward, and backward	160	4001	10	4	40:40:40:40
IV_3_s2						

multiple sub-cluster centers into specified k clusters by formalizing it as an optimization problem that updates the partitions of m sub-cluster centers and k clusters with an alternating optimization strategy.

USPEC [41]:Ultra-Scalable Spectral Clustering uses a hybrid representative selection strategy and a fast approximating method of k-nearest representatives to construct a sparse affinity sub-matrix, and then utilizes transfer cut to partition EEG time series.

USENC [41]: Ultra-Scalable Ensemble Clustering integrates multiple USPEC clusters into an ensembe clustering framework to enhance the robustness of USPEC.

SSDC [43]: Semi-Supervised DenPeak Clustering exploits pairwise constraints or side information to guide EEG time series cluster process, which improves the performance of clustering.

SCPCOG [37]: Semi-Supervised Clustering via Pairwise Constrained structured Optimal Graph cluster EEG signals with the pairwise constraints of must-link and cannot-link in semi-supervised graph-based clustering.

FNSC [13]: Fair Normalized Spectral Clustering partitions EEG signals into corresponding fairer EEG clusters by embedding fairness constraints in graph spectral clustering.

The number of clusters for clustering algorithms are set according to the number of classes in original EEG datasets, and parameter values in baseline methods are set as same as the original papers (For details, please refer to the original paper). For *ConsEEGc* [2], the kernel parameter σ for Gaussian similarity that is used to calculate the similarities between EEG data, and the optimal combinations of weight parameters $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5$ for the objective function are flexibly auto-searched from $\{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$ according to the size of EEG datasets. For example, IV_2b_s1, IV_2b_s2, IV_3_s1, and IV_3_s2 are relatively small-sized EEG datasets, so it prefers to auto-search the optimal combination of parameters from the range $\{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$; II_Ia, II_Ib, IV_2a_s1, IV_2a_s2, and IV_2a_s3 are relatively large-sized EEG datasets, so it prefers to auto-search the optimal combination of parameters from the range $\{10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}\}$; and for II_Ia_Ib, III_V_s1, and III_V_s2, the optimal combination of parameters is preferred to auto-search from the range $\{10^{-1}, 10^{0}, 10^{1}\}$. The learning rate η is kept fixed at 0.01, and the loop is set no more than 50 iterations for convergence. All baseline algorithms (i.e., those seven state-of-the-art clustering algorithms) are conducted with Matlab R2014b, on Windows 10 machines with 8∗3.0 GHz CPUs and 16 GB memory. To render the clustering results more reliable, all algorithms are run 20 times.

[2]The code is available online at https://github.com/Jackie-Day/EEG-data-and-descriptions.

D. EEG Clustering Results Compared with Baselines

We present the experimental results and discussion for the proposed *ConsEEGc* that is designed for EEG semi-supervised clustering with multiple constraints, including the completely unlabeled EEG clustering performance compared to state-of-the-art clustering algorithms, multi-constrained semi-supervised clustering performance comparison, execution time analysis, convergence analysis, and sensitivity analysis. Note that the results of *ConsEEGc* are achieved in best parameters that are auto searched. In fact, we transformed label information of original EEG signals to connectivity constraints in the experiments. For example, 10% connectivity constraint means the labels of 10% EEG signals are prior given.

1) Completely Unlabeled EEG Clustering Results: To show the efficacy of *ConsEEGc* for completely unlabeled EEG clustering, we compared its performance with that of seven state-of-the-art clustering algorithms without using any initialization of label-transformed connectivity constraints. That is, the value of connectivity constraint was initialized to 0%, and we initialized the fairness constraint according to the definition of fairness in Section IV-F that can be described as the minimum ratio of EEG signals between different classes (categories) in EEG datasets. In detail, the fairness constraint values for EEG datasets II_Ia, II_Ib, II_Ia_Ib, III_V_s1, II_V_s2, IV_2a_s1, IV_2a_s2, IV_2a_s3, IV_2b_s1, IV_2b_s2, IV_3_s1, and IV_3_s2 are set as 133/135, 100/100, 100/135, 928/1440, 880/1472, 72/72, 72/72, 72/72, 60/60, 60/60, 40/40, and 40/40, respectively, according to Table I. The

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2022.3206330

TABLE II
EEG CLUSTERING RESULTS ON TWELVE EEG DATASETS WITH RESPECT TO *NMI*, *ARI*, *F-score* AND *kappa*.

EEG dataset	Criteria	k-means++	KMMC	USPEC	USENC	SSDC	SCPCOG	FNSC	ConsEEGc
II_Ia	Acc	0.5115 ± 0.0221	0.4985 ± 0.0003	0.4982 ± 0.0	0.501 ± 0.0121	0.5175 ± 0.0208	0.4982 ± 0.0001	0.4991 ± 0.0	**0.5324 ± 0.0**
	NMI	0.0352 ± 0.0175	0.0308 ± 0.0218	0.0001 ± 0.0	0.0104 ± 0.0369	0.0307 ± 0.0322	0.0296 ± 0.0003	0.0017 ± 0.0	**0.0509 ± 0.0**
	ARI	0.0432 ± 0.0241	0.0005 ± 0.0005	-0.0035 ± 0.0	0.0033 ± 0.0242	0.0352 ± 0.0415	0.0 ± 0.0002	-0.0015 ± 0.0	**0.0648 ± 0.0**
	F-score	0.4869 ± 0.11	0.5049 ± 0.0132	0.4996 ± 0.0038	0.5047 ± 0.0382	0.522 ± 0.0982	0.5056 ± 0.0033	0.5045 ± 0.0225	**0.5261 ± 0.1313**
	kappa	0.2084 ± 0.0611	0.0217 ± 0.0142	0.0093 ± 0.0	0.0326 ± 0.0695	0.1675 ± 0.1042	0.0148 ± 0.0002	0.0464 ± 0.0	**0.2618 ± 0.0**
II_Ib	Acc	0.4986 ± 0.0012	0.4975 ± 0.0001	0.4975 ± 0.0	0.4976 ± 0.0002	0.4983 ± 0.0009	0.4977 ± 0.0	0.4976 ± 0.0001	**0.4999 ± 0.0**
	NMI	0.0017 ± 0.0018	0.0026 ± 0.0027	0.0001 ± 0.0	0.0002 ± 0.0002	0.0014 ± 0.0015	0.0254 ± 0.0	0.0001 ± 0.0001	**0.0337 ± 0.0415**
	ARI	-0.0026 ± 0.0024	-0.0002 ± 0.0001	-0.0049 ± 0.0	-0.0046 ± 0.0004	-0.0027 ± 0.0018	-0.0002 ± 0.0	-0.0049 ± 0.0001	**0.0 ± 0.0**
	F-score	0.4997 ± 0.0242	0.5 ± 0.0054	0.499 ± 0.005	0.5018 ± 0.0085	0.4975 ± 0.0204	0.51 ± 0.0	0.4983 ± 0.0059	**0.521 ± 0.0287**
	kappa	0.0385 ± 0.028	0.009 ± 0.0055	0.01 ± 0.0	0.0145 ± 0.0089	0.033 ± 0.0232	0.02 ± 0.0	0.0115 ± 0.0037	**0.07 ± 0.0**
II_Ia_Ib	Acc	0.7119 ± 0.0057	0.2842 ± 0.0016	0.7187 ± 0.0	0.7205 ± 0.0062	0.7054 ± 0.0418	0.2806 0.0001	0.7268 ± 0.0109	**0.7344 ± 0.0054**
	NMI	0.4021 ± 0.0106	0.0565 ± 0.0069	0.4062 ± 0.0	0.4109 ± 0.0283	0.367 ± 0.0982	0.0825 ± 0.0001	0.4062 ± 0.0195	**0.4161 ± 0.0105**
	ARI	0.2987 ± 0.017	0.0007 ± 0.0021	0.3081 ± 0.0	0.3058 ± 0.0318	0.26 ± 0.081	0.0028 ± 0.0001	0.3038 ± 0.0246	**0.3161 ± 0.0172**
	F-score	0.1956 ± 0.1259	0.1848 ± 0.0083	0.1972 ± 0.1243	0.2162 ± 0.1419	0.2306 ± 0.083	0.1659 ± 0.0005	0.245 ± 0.1587	**0.2495 ± 0.1062**
	kappa	0.3462 ± 0.0409	0.0176 ± 0.006	0.2771 ± 0.0	0.3238 ± 0.0694	0.3464 ± 0.0783	0.0374 ± 0.0001	0.3309 ± 0.0051	**0.3633 ± 0.0425**
III_V_s1	Acc	0.5544 ± 0.0006	0.3456 ± 0.0002	0.5645 ± 0.013	0.5667 0.0162	0.5606 ± 0.0096	0.3452 ± 0.0002	0.5606 ± 0.0117	**0.5742 ± 0.0005**
	NMI	0.1453 ± 0.0006	0.0082 ± 0.0021	0.1499 ± 0.0103	0.1547 ± 0.02	0.0282 ± 0.0222	0.0146 ± 0.0014	0.1503 ± 0.0168	**0.1651 ± 0.0005**
	ARI	0.0798 ± 0.0008	-0.0006 ± 0.0003	0.0834 ± 0.0196	0.0831 ± 0.0203	0.0267 ± 0.0199	-0.0001 ± 0.0003	0.0768 ± 0.0153	**0.0995 ± 0.0007**
	F-score	0.2783 ± 0.0523	0.3115 ± 0.0006	0.3291 ± 0.1343	0.3196 ± 0.1078	0.3055 ± 0.0457	0.2925 ± 0.054	0.3353 ± 0.1156	**0.354 ± 0.0175**
	kappa	0.3104 ± 0.0008	-0.0015 ± 0.0011	0.3131 ± 0.0217	0.3184 ± 0.027	0.1077 ± 0.041	0.0008 ± 0.0011	0.3044 ± 0.0229	**0.3301 ± 0.0008**
III_V_s2	Acc	0.5177 ± 0.0364	0.3495 ± 0.0005	0.505 ± 0.0585	0.5113 ± 0.0409	0.5137 ± 0.0404	0.3489 ± 0.0001	0.5155 ± 0.0288	**0.5359 ± 0.0217**
	NMI	0.1046 ± 0.0198	0.0137 ± 0.0043	0.1001 ± 0.0329	0.1028 ± 0.0162	0.0175 ± 0.0124	0.0155 ± 0.0009	0.0769 ± 0.0014	**0.1103 ± 0.0114**
	ARI	0.0516 ± 0.034	-0.0002 ± 0.0007	0.0588 ± 0.0254	0.0575 ± 0.0293	0.0237 ± 0.0181	0.0001 ± 0.0002	0.0471 ± 0.0025	**0.0604 ± 0.0235**
	F-score	0.2794 ± 0.0638	0.3232 ± 0.0214	0.3348 ± 0.1015	0.2808 ± 0.104	0.3283 ± 0.0549	0.3156 ± 0.0383	0.3284 ± 0.1078	**0.3463 ± 0.0081**
	kappa	0.1773 ± 0.0547	0.0001 ± 0.0025	0.1603 ± 0.0601	0.174 ± 0.0589	0.0913 ± 0.0521	0.0006 ± 0.0007	0.1418 ± 0.0011	**0.1945 ± 0.0444**
IV_2a_s1	Acc	0.6098 ± 0.0279	0.2898 ± 0.0029	0.6171 ± 0.0016	0.6123 ± 0.0199	0.5962 ± 0.0162	0.2787 ± 0.0023	0.6214 ± 0.0056	**0.6329 ± 0.0099**
	NMI	0.0319 ± 0.0336	0.0446 ± 0.015	0.0456 ± 0.0239	0.0323 ± 0.0122	0.0222 ± 0.0116	0.0482 ± 0.0238	0.0333 ± 0.0147	**0.0546 ± 0.0129**
	ARI	0.0164 ± 0.0229	0.0002 ± 0.0003	0.0281 ± 0.0137	0.0186 ± 0.0123	0.0074 ± 0.0082	0.0005 ± 0.0001	0.0193 ± 0.0135	**0.0363 ± 0.0128**
	F-score	0.2416 ± 0.0519	0.2483 ± 0.0078	0.2446 ± 0.0443	0.2497 ± 0.0427	0.2487 ± 0.0589	0.2313 ± 0.0511	0.2509 ± 0.0406	**0.2589 ± 0.0518**
	kappa	0.1156 ± 0.0394	0.0264 ± 0.008	0.1151 ± 0.033	0.1 ± 0.0251	0.0829 ± 0.0254	0.0417 ± 0.0014	0.116 ± 0.0205	**0.141 ± 0.0295**
IV_2a_s2	Acc	0.5872 ± 0.0366	0.284 ± 0.0041	0.5996 ± 0.0	0.6087 ± 0.0148	0.5925 ± 0.0173	0.2816 ± 0.0126	0.6126 ± 0.0031	**0.6261 ± 0.0037**
	NMI	0.0131 ± 0.0066	0.0114 ± 0.0108	0.0069 ± 0.0	0.0135 ± 0.0071	0.0149 ± 0.0094	0.0145 ± 0.0098	0.0153 ± 0.0073	**0.0164 ± 0.0066**
	ARI	0.0036 ± 0.0054	0.0 ± 0.0004	-0.0028 ± 0.0	0.0039 ± 0.0055	0.0037 ± 0.008	0.0004 ± 0.0002	0.0041 ± 0.0077	**0.0046 ± 0.0063**
	F-score	0.2455 ± 0.0294	0.2488 ± 0.008	0.2439 ± 0.0243	0.2465 ± 0.0312	0.2324 ± 0.0287	0.2411 ± 0.047	0.2464 ± 0.0448	**0.2533 ± 0.04**
	kappa	0.0683 ± 0.0175	0.0206 ± 0.0055	0.0417 ± 0.0	0.0704 ± 0.0192	0.0671 ± 0.0214	0.0414 ± 0.001	0.0719 ± 0.0277	**0.0808 ± 0.0198**
IV_2a_s3	Acc	0.4808 ± 0.1144	0.2819 ± 0.0035	0.5939 ± 0.0	0.5642 ± 0.0399	0.5948 ± 0.0153	0.2792 ± 0.0022	0.6151 ± 0.0016	**0.6307 ± 0.0043**
	NMI	0.025 ± 0.0086	0.0265 ± 0.0104	0.0215 ± 0	0.0256 ± 0.0111	0.0163 ± 0.0072	0.0276 ± 0.0098	0.0272 ± 0.0055	**0.0291 ± 0.0091**
	ARI	0.005 ± 0.0069	0.0 ± 0.0003	0.0138 ± 0.0	0.01 ± 0.0068	0.0026 ± 0.0049	0.0004 ± 0.0001	0.0148 ± 0.0045	**0.0164 ± 0.0085**
	F-score	0.2429 ± 0.031	0.2422 ± 0.0049	0.2444 ± 0.04	0.2444 ± 0.0299	0.2415 ± 0.0254	0.2409 ± 0.0416	0.2375 ± 0.0499	**0.266 ± 0.0464**
	kappa	0.062 ± 0.0331	0.0169 ± 0.0053	0.0828 ± 0.0	0.0924 ± 0.0214	0.0669 ± 0.0193	0.0412 ± 0.0021	0.1014 ± 0.0089	**0.112 ± 0.0263**
IV_2b_s1	Acc	0.4968 ± 0.0013	0.496 ± 0.0002	0.4973 ± 0.0	0.4968 ± 0.0014	0.4976 ± 0.0018	0.4971 ± 0.0003	0.4969 ± 0.0003	**0.4977 ± 0.0028**
	NMI	0.0034 ± 0.007	0.0065 ± 0.0105	0.0054 ± 0.0	0.0052 ± 0.012	0.0038 ± 0.004	0.0062 ± 0.0369	0.0017 ± 0.0005	**0.0074 ± 0.0376**
	ARI	-0.0043 ± 0.0028	**-0.0005 ± 0.0004**	-0.0011 ± 0.0	-0.002 ± 0.0033	-0.0033 ± 0.0039	-0.0018 ± 0.0032	-0.0059 ± 0.0008	-0.0043 ± 0.0058
	F-score	0.4683 ± 0.0528	0.495 ± 0.0087	0.5025 ± 0.0424	0.4142 ± 0.0554	0.4791 ± 0.0432	0.4858 ± 0.0426	0.5029 ± 0.0235	**0.5054 ± 0.043**
	kappa	0.0358 ± 0.0261	0.0167 ± 0.0108	0.0333 ± 0.0	0.0325 ± 0.0299	0.0417 ± 0.0405	0.045 ± 0.0051	0.0467 ± 0.0103	**0.0475 ± 0.036**
IV_2b_s2	Acc	0.498 ± 0.0028	0.4959 ± 0.0002	0.4959 ± 0.0	0.4971 ± 0.0022	0.4983 ± 0.0031	0.4971 ± 0.0017	0.4987 ± 0.0012	**0.4992 ± 0.0022**
	NMI	0.0041 ± 0.0049	0.0078 ± 0.0118	0.0002 ± 0.0001	0.0059 ± 0.0056	0.0046 ± 0.0061	0.0162 ± 0.072	0.0047 ± 0.0018	**0.0175 ± 0.057**
	ARI	-0.0028 ± 0.0057	**-0.0005 ± 0.0003**	-0.0082 ± 0.0001	-0.0031 ± 0.0031	-0.0021 ± 0.0062	-0.0017 ± 0.0034	-0.0019 ± 0.0022	-0.001 ± 0.004
	F-score	0.4925 ± 0.0343	0.4971 ± 0.0082	0.5 ± 0.0076	0.4883 ± 0.0293	0.4842 ± 0.0321	0.5025 ± 0.0113	0.4979 ± 0.0386	**0.5113 ± 0.0388**
	kappa	0.0533 ± 0.0388	0.0142 ± 0.0098	0.015 ± 0.0051	0.0342 ± 0.0373	0.0567 ± 0.0424	0.05 ± 0.0011	0.0642 ± 0.0138	**0.0775 ± 0.0277**
IV_3_s1	Acc	0.6059 ± 0.0182	0.3274 ± 0.0062	0.6224 ± 0.0	0.5954 ± 0.0342	0.5835 ± 0.0252	0.3019 ± 0.0067	0.6218 ± 0.0011	**0.6254 ± 0.004**
	NMI	0.0173 ± 0.008	0.0181 ± 0.0073	0.0084 ± 0.0	0.0119 ± 0.0039	0.0147 ± 0.01	0.0126 ± 0.0073	0.0159 ± 0.0008	**0.0184 ± 0.0059**
	ARI	-0.0057 ± 0.0056	-0.0044 ± 0.0051	-0.0115 ± 0.0	-0.0078 ± 0.0025	**-0.0037 ± 0.0056**	-0.0046 ± 0.0029	-0.0065 ± 0.0005	-0.0039 ± 0.0053
	F-score	0.2488 ± 0.0241	0.2459 ± 0.0276	0.2466 ± 0.0235	0.2475 ± 0.0241	0.2349 ± 0.0248	0.2494 ± 0.0219	0.2515 ± 0.0247	**0.2584 ± 0.0261**
	kappa	0.0642 ± 0.0228	0.0433 ± 0.011	0.05 ± 0.0	0.0558 ± 0.0121	0.0729 ± 0.0186	0.0732 ± 0.0046	0.0722 ± 0.0046	**0.0746 ± 0.0079**
IV_3_s2	Acc	0.6172 ± 0.0139	0.3226 ± 0.0086	0.6137 ± 0.0	0.6021 ± 0.0195	0.5943 ± 0.0248	0.3019 ± 0.0019	0.6162 ± 0.0018	**0.6253 ± 0.0031**
	NMI	0.0151 ± 0.005	0.0153 ± 0.0083	0.0144 ± 0.0	0.0131 ± 0.0037	**0.0168 ± 0.0131**	0.0163 ± 0.0046	0.016 ± 0.004	0.0165 ± 0.0052
	ARI	-0.005 ± 0.0047	-0.0052 ± 0.0023	-0.0039 ± 0.0	-0.0066 ± 0.0028	-0.0042 ± 0.0099	-0.0036 ± 0.0025	-0.0041 ± 0.0037	**-0.0034 ± 0.0051**
	F-score	0.2358 ± 0.0316	0.2444 ± 0.0206	0.245 ± 0.0345	0.2374 ± 0.037	0.2309 ± 0.0245	0.2463 ± 0.0314	0.2447 ± 0.0425	**0.2544 ± 0.0287**
	kappa	0.0767 ± 0.021	0.035 ± 0.0088	0.0711 ± 0.0	0.0675 ± 0.0177	0.0767 ± 0.0296	0.075 ± 0.0019	0.0761 ± 0.0148	**0.0796 ± 0.0144**
Avg *Acc* Rank		4.25 ± 1.5877	7 ± 0.4082	4.5 ± 1.7078	4.3333 ± 1.1055	4.1667 ± 1.5184	7 ± 1.5811	3.1667 ± 1.2133	**1.0 ± 0.0**
# Best *Acc*		0	0	0	0	0	0	0	**12**
Avg *NMI* Rank		5.0 ± 1.7795	4.75 ± 2.2776	6.0 ± 2.0817	4.9167 ± 1.8008	5.3333 ± 1.8856	4.1667 ± 2.0344	4.6667 ± 1.795	**1.0833 ± 0.2764**
# Best *NMI*		0	0	0	0	1	0	0	**11**
Avg *ARI* Rank		4.75 ± 1.0897	5.5 ± 2.8136	4.5833 ± 2.7525	4.6667 ± 1.748	4.8333 ± 1.4625	5.1667 ± 1.9076	4.8333 ± 1.993	**1.6667 ± 1.6499**
# Best *ARI*		0	2	0	0	1	0	0	**9**
Avg *F-score* Rank		6.25 ± 1.4215	4.8333 ± 1.2802	4.3333 ± 1.748	4.75 ± 1.8314	5.9167 ± 2.0999	5 ± 2.3452	3.8333 ± 2.0344	**1.0 ± 0.0**
# Best *F-score*		0	0	0	0	0	0	0	**12**
Avg *kappa* Rank		3.5 ± 1.3229	7.8333 ± 0.5528	5.6667 ± 1.4907	4.75 ± 1.5877	4.0 ± 1.4719	5.6667 ± 1.748	3.5 ± 1.3844	**1.0 ± 0.0**
# Best *kappa*		0	0	0	0	0	0	0	**12**

results in Table II demonstrate that *ConsEEGc* outperforms all state-of-the-art clustering algorithms for EEG clustering, with *ConsEEGc* achieving the lowest average ranks of *Acc*, *NMI*, *ARI*, *F-score* and *kappa*, as well as the largest number of best *Acc*, *NMI*, *ARI*, *F-score* and *kappa* across all twelve EEG datasets compared. In other words, although no label-transformed connectivity constraints are initialized in the beginning, our method still can obtain good clustering results with the iteratively learning and optimizing strategy for the multi-constrained objective function.

2) Multi-Constrained Semi-Supervised Clustering Results: We next compared the *ConsEEGc* with the state-of-the-art clustering algorithms for semi-supervised EEG clustering, with different connectivity constraints and fairness constraint included. The connectivity percentage among EEG data is set as $\{1\%, 5\%, 10\%, 20\%, 30\%, 50\%, 75\%\}$ and the fairness constraint is set as the minimum value of the category ratio. The results of this comparison, see Fig.2 (Only the results on II_Ia, II_Ib, II_Ia_Ib, III_V_s1, III_V_s2, and III_2a_s1 are presented, and those on IV_2a_s2, IV_2a_s3, IV_2b_s1, IV_2b_s2, IV_3_s1,

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2022.3206330

Fig. 2. EEG clustering on the first six of twelve EEG dataset: (a) II_Ia, (b) II_Ib, (c) II_Ia_Ib, (d) III_V_s1, (e) III_V_s2, and (f) IV_2a_s1 *w.r.t. Acc*, *NMI*, *ARI*, *F-score*, and *kappa*. All the results were calculated without using known EEG labels, *e.g.*, when connectivity constraint is set 10% with known EEG labels, then the rest 90% EEG signals are used to cluster and calculate the averaged *Acc*, *NMI*, *ARI*, *F-score*, and *kappa*.

and IV_3_s2 are illustrated in Fig.S1 in Appendix F), were achieved with those EEG data where those are used to build constraints are excluded. The results clearly show that the *ConsEEGc* yields the highest *Acc*, *NMI*, *ARI*, *F-score* and *kappa* on different connectivity and fairness constraints for twelve different EEG datasets, which indicates the superiority of *ConsEEGc* on semi-supervised EEG clustering.

3) Execution Time Comparison: As analyzed in Section V-D, the time complexity of *ConsEEGc* requires $O(iter \cdot n^3)$, which indicates that the running time of *ConsEEGc* mainly depends on the number of EEG signals. But with the fast convergence of *ConsEEGc* that will be introduced subsequently, it can learn and output EEG labels in a short time. The running time presented in Fig.3 (Only the results on II_Ia, II_Ib, II_Ia_Ib, III_V_s1, III_V_s2, and III_2a_s1 are provided here, and those on IV_2a_s2, IV_2a_s3, IV_2b_s1, IV_2b_s2, IV_3_s1, and IV_3_s2 are shown in Fig.S2 in Appendix G) also indicates that *ConsEEGc* has competitive efficiency on different EEG datasets with various sizes compared to state-of-the-art clustering algorithms, thanks to its quick convergence.

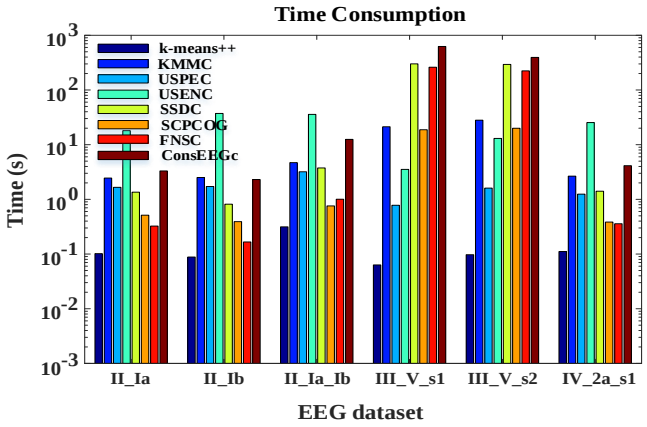


Fig. 3. Average running time for semi-supervised EEG clustering on EEG datasets II_Ia, II_Ib, II_Ia_Ib, III_V_s1, III_V_s2, and III_2a_s1.

E. Sensitivity Analysis

Supervised approaches, i.e., classification, are inappropriate for unlabeled EEG data, and most unsupervised methods, i.e., traditional clustering, do not make full use of limited given information of EEG data or just consider single constraint of EEG data when clustering. We have designed a novel model *ConsEEGc* to cluster unlabeled EEG data in a semi-supervised way that considers multiple constraints such as connectivity constraints, compactness-and-scatter constraint and fairness constraint. Our experimental results demonstrate that our novel technique *ConsEEGc* with multiple constraints achieved the best clusters on various types of EEG datasets compared to all state-of-the-art clustering algorithms that includes unsupervised clustering, semi-supervised clustering with single constraints, and fair clustering. Possible reasons for its apparent superiority over other methods are that (1) *ConsEEGc* utilizes multiple constraints that can constrain the correlations from different perspectives of a *priori* knowledge, ignored by previously reported clustering techniques and (2) *ConsEEGc* potentially makes the best use of a *priori* knowledge of partially labeled EEG data.

The clustering performance of *ConsEEGc* is also affected by several factors that is mainly from three perspectives: the percentage of label-transformed connectivity, the impact of parameters in objective function, and the convergence of *ConsEEGc*. Therefore, we also discussed their impact on the clustering of *ConsEEGc* respectively in the following.

1) Impact of Label-Transformed Constraints: The clustering results in the condition that the EEG data used to build constraints are excluded (see Fig.2 and Fig.S1 in Appendix F in the Supplementary file) demonstrate that the clustering performance of *ConsEEGc* changes smoothly along with the increase in the percentage of connectivity constraints. We further evaluated the impact of connectivity constraints and the fairness constraint on *ConsEEGc* in the condition that those EEG data used to build constraints are included. Specifically, the connectivity percentage is set as $\{1\%, 5\%, 10\%, 15\%, 20\%, 25\%, 30\%, 40\%, 50\%, 60\%, 75\%\}$ and the fairness constraint remains as the minimum of the category ratio among any two categories. The cluster performance of *ConsEEGc* applied to a range of EEG datasets is clearly enhanced (see Fig.4, here only the results on EEG datasets II_Ia, II_Ib, II_Ia_Ib, III_V_s1, III_V_s2, and III_2a_s1 are presented, and those on IV_2a_s2, IV_2a_s3, IV_2b_s1, IV_2b_s2, IV_3_s1, and IV_3_s2 are introduced in Fig.S3 in Appendix H) with an increasing percentage of connectivity constraints, according to *Acc*, *NMI*, *ARI*, *F-score*, and *kappa*. Consequently, *ConsEEGc* can learn a better label indicator with higher percentage of label-transformed connectivity constraints from given a *priori* knowledge of EEG data and then help produce correct labels for EEG data.

2) Impact of Regularization Parameters: The parameters in our model of *ConsEEGc* include the kernel parameter σ for computation of Gaussian similarity of EEG data, and weight parameters $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5$ for the function terms in objective function defined in Eq.(25). Here, we discuss the impact of a single parameter by setting the values of other parameters to 10^0, confining our assessment to three different EEG datasets; II_Ia (Fig.5) in which connectivity constraints and fairness constraint are set as 10% and 133/135, respectively), II_2a_s1 (Fig.S4 in Appendix I) in which connectivity constraints and fairness constraint are set as 10% and 72/72, respectively), and IV_3_s1 (Fig.S5 in Appendix I) in which connectivity constraints and fairness constraint are set as 10% and 40/40, respectively. The results demonstrate that there is no linear relation between the performance of *ConsEEGc* and these parameters, but they clearly show that *ConsEEGc* yields relatively best clustering performance when $\sigma, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5$ are set around 10^0 *w.r.t. Acc*, *NMI*, *ARI*, *F-score* and *kappa*. Besides, *ConsEEGc* is more sensitive to γ_4 than $\sigma, \gamma_1, \gamma_2, \gamma_3, \gamma_5$ on EEG datasets II_Ia, IV_2a_s1, and IV_3_s1. This is especially so when $\gamma_4 > 10^1$ the performance of *ConsEEGc* is considerably degraded.

3) Impact of Convergence: The convergence of *ConsEEGc* for twelve EEG datasets, see Fig.6 (here only the results on the EEG datasets II_Ia, II_Ib are presented, and those on other EEG datasets are shown in Fig.S6 in Appendix J

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2022.3206330

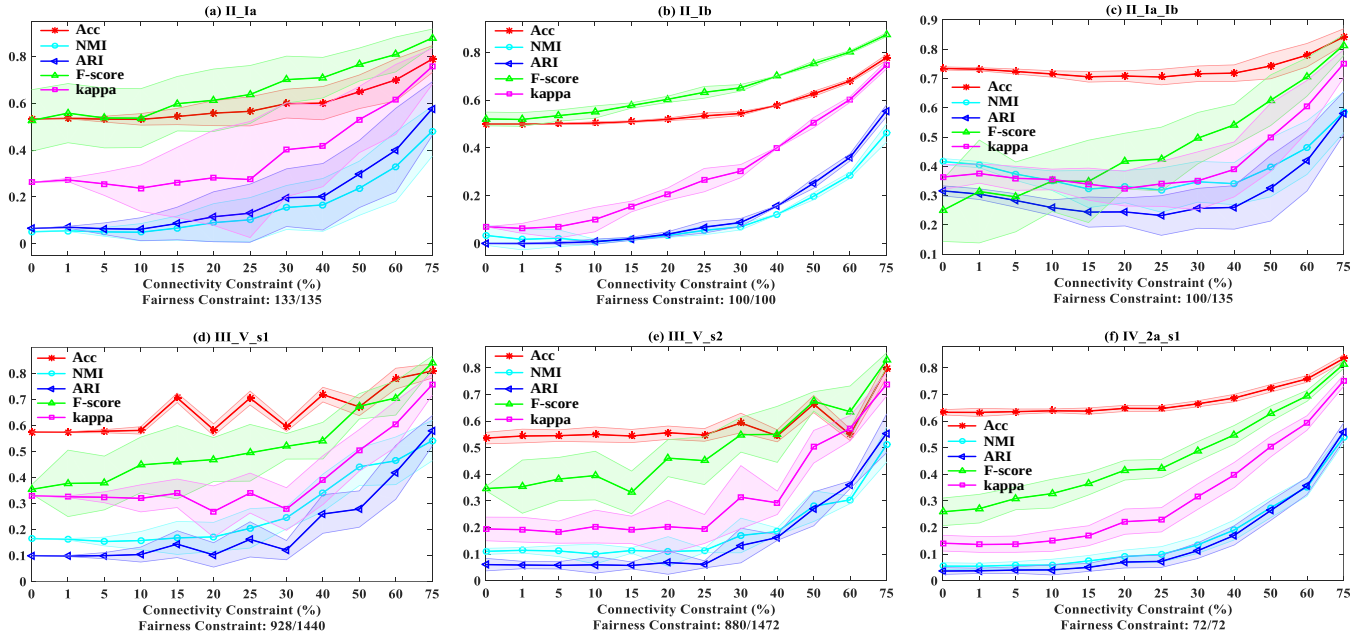


Fig. 4. The clustering performance of *ConsEEGc w.r.t. Acc*, *NMI*, *ARI*, *F-score*, and *kappa* in different constraints on EEG datasets II_Ia, II_Ib, II_Ia_Ib, III_V_s1, III_V_s2, and III_2a_s1.

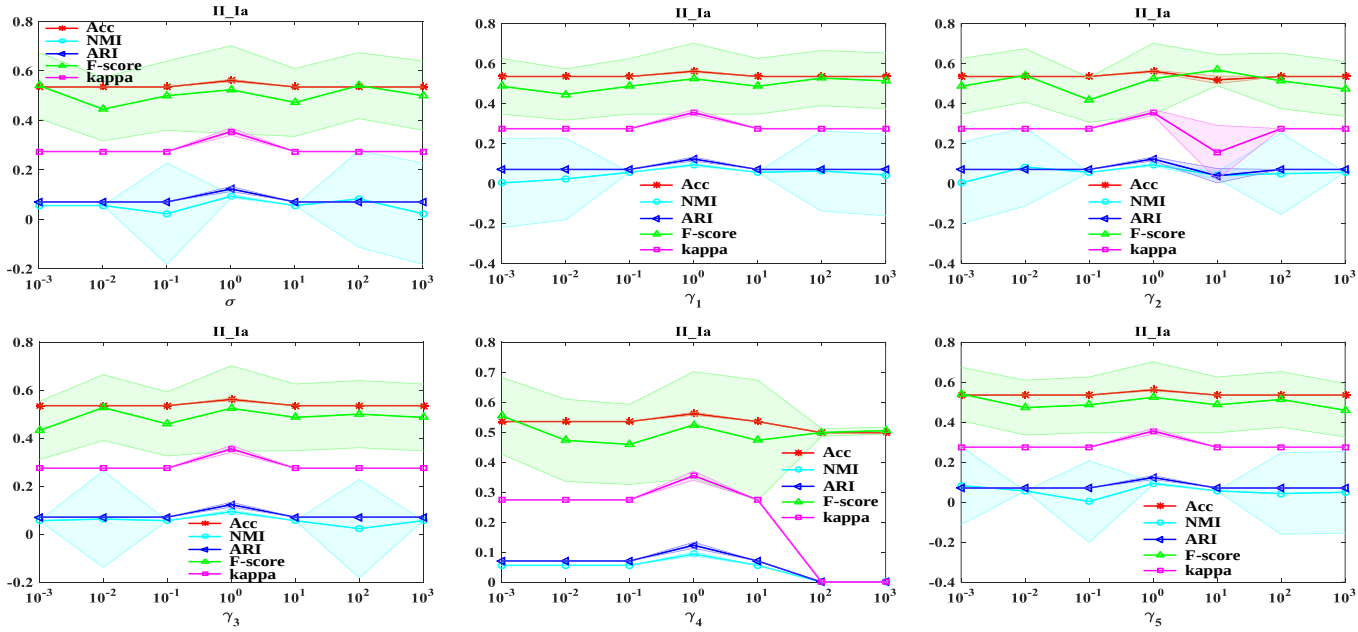


Fig. 5. Parameter impact on *ConsEEGc* for EEG dataset II_Ia *w.r.t. Acc*, *NMI*, *ARI*, *F-score*, and *kappa*.

of the Supplementary file), demonstrates convincingly that *ConsEEGc* yields the locally optimal solution with relatively low time cost, since *ConsEEGc* converges to local optima within the first five iterations, and the impact of learning rate on convergence is slight but *ConsEEGc* achieves the relatively best results on $\eta = 0.001$ for most situations, particularly.

F. Ablation Analysis

To further illustrate the effectiveness of multiple constraints on our method, we conducted the ablation analysis in this section. The results illustrated in Fig.S7 and Fig.S8 clearly demonstrate that the full model with multiple constraints outperforms that without constraints. Due to the page limit, please refer to the details in Appendix K.

VII. Conclusions

With the ever-increasing amount of unlabeled EEG signals in BCI, disease diagnosis, rehabilitation, etc, analysis on partially labeled or incorrectly mislabeled EEG has become an urgent and challenging task. However, to the best of our knowledge, related works on semi-supervised EEG clustering with multiple constraints are rarely reported. Therefore, to

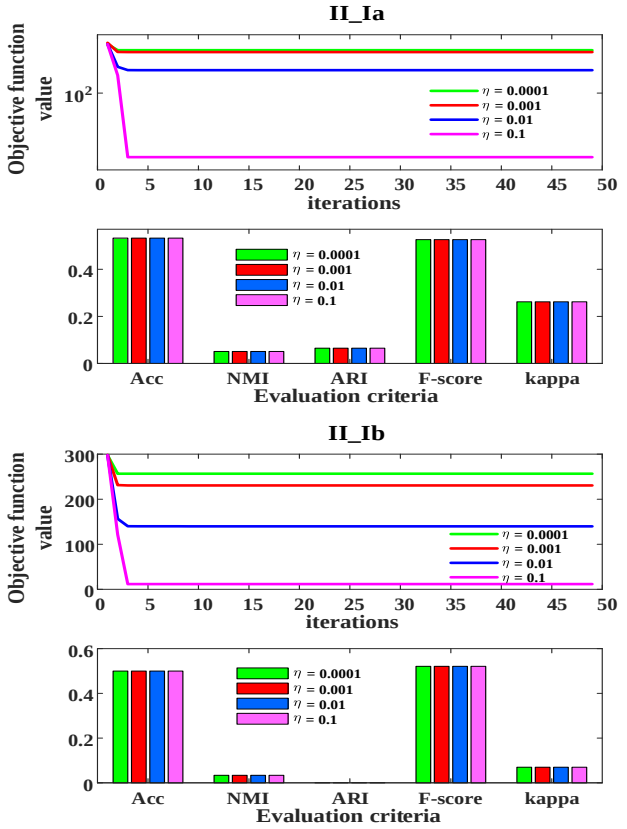


Fig. 6. The impact of learning rate on convergence and performance of *ConsEEGc* on EEG datasets II_Ia and II_Ib.

address the tough problem, this paper investigated semi-supervised EEG clustering with multiple constraints from original EEG datasets and proposed a novel approach that simultaneously considers partial label-transformed connectivity and compactness-and-scatter-transformed fairness to cluster partially labeled EEG signals, which, to the end, leads to a quickly converging algorithm *ConsEEGc*. With the detailed experimentation on twelve real-world EEG datasets, the results demonstrated the superiority of *ConsEEGc* over seven state-of-the-art unsupervised, semi-supervised and constrained clustering algorithms on partially labeled EEG signals.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 62106087), Natural Science Foundation of Jiangsu Province (Grant No. BK20210455) and the Fundamental Research Funds for the Central Universities (Grant No. JUSRP122033).

REFERENCES

[1] F. R. Willett, D. T. Avansino, L. R. Hochberg, et al., "High-performance brain-to-text communication via handwriting," *Nature*, vol. 593, pp. 249–254, 2021.
[2] F. Cincotti, F. Pichiorri, P. Aricò, et al., "EEG-based brain-computer interface to support post-stroke motor rehabilitation of the upper limb," in *EMBC*, 2012, pp. 4112–4115.
[3] R. Foong, N. Tang, E. Chew, et al., "Assessment of the efficacy of EEG-based MI-BCI with visual feedback and EEG correlates of mental fatigue for upper-limb stroke rehabilitation," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 3, pp. 786–795, 2020.
[4] M. F. Glasser, T. S. Coalson, E. C. Robinson, et al., "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, pp. 171–178, 2016.
[5] G. S. Thirunavukkarasu, H. Abdi, and N. Mohajer, "A smart HMI for driving safety using emotion prediction of EEG signals," in *SMC*, 2016, pp. 4148–4153.
[6] H. Yu, X. Lei, Z. Song, et al., "Supervised network-based fuzzy learning of EEG signals for Alzheimer's disease identification," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 1, pp. 60–71, 2020.
[7] Y. Zhao, Y. Zhao, P. Durongbhan, et al., "Imaging of nonlinear and dynamic functional brain connectivity based on EEG recording with the application on the diagnosis of Alzheimer's disease," *IEEE Trans. Medical Imaging*, vol. 39, no. 5, pp. 21571–1581, 2020.
[8] K. Xia, T. Ni, H. Yin, et al., "Cross-domain classification model with knowledge utilization maximization for recognition of epileptic EEG signals," *IEEE ACM Trans. Compt. Biol. Bioinform.*, vol. 18, no. 1, pp. 53–61, 2021.
[9] V. Jayaram, N. Widmann, C. Forster, et al., "Brain-computer interfacing in amyotrophic lateral sclerosis: Implications of a resting-state EEG analysis," in *EMBC*, 2015, pp. 6979–6982.
[10] C. Dai, J. Wu, D. Pi, et al., "Brain EEG time-series clustering using maximum-weight clique," *IEEE Trans. Cybern.*, early access, pp. 1–15, 2020.
[11] C. Dai, J. Wu, D. Pi, et al., "Electroencephalogram signal clustering with convex cooperative games," *IEEE Trans. Knowl. Data Eng.*, early access, pp. 1–14, 2021.
[12] K. Wagstaff and C. Cardie, "Clustering with instance-level constraints," in *ICML*, 2000, pp. 1103–1110.
[13] M. Kleindessner, S. Samadi, P. Awasthi, et al., "Guarantees for spectral clustering with fairness constraints," in *ICML*, 2019, pp. 3458–3467.
[14] H. Liu, Z. Tao, and Y. Fu, "Partition level constrained clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2469–2483, 2018.
[15] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *SODA*, 2007, pp. 1027–1035.
[16] F. Nie, C. Wang, and X. Li, "K-multiple-means: A multiple-means clustering method with specified k clusters," in *KDD*, 2019, pp. 959–967.
[17] M. Ester, H. P. Kriegel, J. Sander, et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," in *KDD*, 1996, pp. 226–231.
[18] M. Ankerst, M. M. Breunig, H. P. Kriegel, et al., "OPTICS: Ordering points to identify the clustering structure," in *ACM SIGMOD*, 1999, pp. 49–60.
[19] L. Ertöz, M. Steinbach, V. Kumar, et al., "Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data," in *SDM*, 2003, pp. 47–58.
[20] F. Petitjean, A. Ketterlin, and P. Gancarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recogn.*, vol. 44, no. 3, pp. 678–693, 2011.
[21] J. Yang and J. Leskovec, "Patterns of temporal variation in online media," in *WSDM*, 2011, pp. 177–186.
[22] Z. Li, Y. Yang, J. Liu, et al., "Unsupervised feature selection using nonnegative spectral analysis," in *AAAI*, 2012, pp. 1016–1032.
[23] M. Qian and C. Zhai, "Robust unsupervised feature selection," in *IJCAI*, 2013, pp. 1621–1627.
[24] L. Shi, L. Du, and Y. D. Shen, "Robust spectral learning for unsupervised feature selection," in *ICDM*, 2014, pp. 977–982.
[25] J. Paparrizos and L. Gravano, "k-shape: Efficient and accurate clustering of time series," in *ACM SIGMOD*, 2015, pp. 1855–1870.
[26] Q. Zhang, J. Wu, P. Zhang, et al., "Salient subsequence learning for time series clustering," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 41, no. 9, pp. 2193–2207, 2019.
[27] S. Basu, M. Bilenko, R. Mooney, et al., "A probabilistic framework for semi-supervised clustering," in *KDD*, 2004, pp. 59–68.
[28] X. Zhu and A. Goldberg, *Introduction to semi-supervised learning*, San Rafael, CA, USA: Morgan Claypool, 2009.
[29] H Wu and Z. Liu, "Non-negative matrix factorization with constraints," in *AAAI*, 2010, pp. 506–511.
[30] B. Brubach, D. Chakrabarti, J. P. Dickerson, et al., "Fairness, semi-supervised learning, and more: A general framework for clustering with stochastic pairwise constraints," in *AAAI*, 2021, pp. 6822–6830.
[31] M. Bilenko, S. Basu, and R. J. Mooney, "Integrating constraints and metric learning in semi-supervised clustering," in *ICML*, 2004, pp. 81–88.
[32] J. Zhang and R. Yan, "On the value of pairwise constraints in classification and consistency," in *ICML*, 2007, pp. 1111–1118.

[33] L. Wu, Steven C. H. Hoi, R. Jin, et al., "Learning bregman distance functions for semi-supervised clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 3, pp. 478–491, 2012.

[34] W. Kalintha, S. Ono, M. Numao, et al., "Kernelized evolutionary distance metric learning for semi-supervised clustering," in *AAAI*, 2017, pp. 4945–4946.

[35] I. Davidson and S. S. Ravi, "Clustering with constraints: Feasibility issues and the k-means algorithm," in *SDM*, 2005, pp. 201–211.

[36] F. Nie, X. Wang, M. I. Jordan, et al., "The constrained Laplacian rank algorithm for graph-based clustering," in *AAAI*, 2016, pp. 1969–1976.

[37] F. Nie, H. Zhang, R. Wang, et al., "Semi-supervised clustering via pairwise constrained optimal graph," in *IJCAI*, 2020, pp. 3160–3166.

[38] S. Anand, S. Mittal, O. Tuzel, et al., "Semi-supervised kernel mean shift clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1201–1215, 2014.

[39] Y. Lai, S. He, Z. Lin, et al., "An adaptive robust semi-supervised clustering framework using weighted consensus of random k-means ensemble," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 5, pp. 1877–1890, 2021.

[40] X. Wang, B. Qian, and I. Davidson, "On constrained spectral clustering and its applications," *Data Mining Knowl. Discov.*, vol. 28, no. 1, pp. 1–30, 2014.

[41] D. Huang, C. Wang, J. Wu, et al., "Ultra-scalable spectral clustering and ensemble clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1212–1226, 2020.

[42] L. Lelis and J. Sander, "Semi-supervised density-based clustering," in *ICDM*, 2009, pp. 842–847.

[43] Y. Ren, X. Hu, K. Shi, et al., "Semi-supervised DenPeak clustering with pairwise constraints," in *PRICAI*, 2018, pp. 837–850.

[44] I. M. Ziko, J. Yuan, E. Granger, et al., "Variational fair clustering," in *AAAI*, 2021, pp. 11202–11209.

[45] B. Li, L. Li, A. Sun, et al., "Approximate group fairness for clustering," in *ICML*, 2021, pp. 6381–6391.

[46] S. Xiong, J. Azimi, and X. Z. Fern, "Active learning of constraints for semi-supervised clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 43–54, 2014.

[47] L. Bai, J. Liang, and F. Cao, "Semi-supervised clustering with constraints of different types from multiple information sources," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 3247–3258, 2021.

[48] U. von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, pp. 395–416, 2007.

[49] B. Mohar, "The Laplacian spectrum of graphs," in *Graph Theory, Combinatorics, and Applications*, 1991, pp. 871–898.

[50] F. R. K. Chung, "Spectral graph theory," in *CBMS Regional Conference Series in Mathematics, No. 92*, 1997, pp. 1–208.

[51] K. Fan, "On a theorem of Weyl concerning eigenvalues of linear transformations I," *PNAS*, vol. 35, pp. 652–655, 1949.

[52] F. Chierichetti, R. Kumar, S. Lattanzi, et al., "Fair clustering through fairlets," in *NeurIPS*, 2017, pp. 5029–5037.

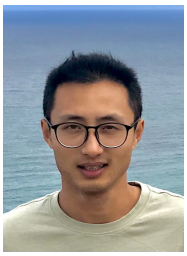
[53] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.

[54] H. Zhang, T. B. Ho, Y. Zhang, et al., "Unsupervised feature extraction for time series clustering using orthogonal wavelet transform," *Informatica*, vol. 30, no. 3, pp. 305–319, 2006.

[55] L. Hubert and P. Arabie, "Comparing partitions," *J. Classif.*, vol. 2, no. 1, pp. 193–218, 1985.

[56] C. J. van Rijsbergen, *Information Retrieval*, 2nd ed., London: Butterworths, 1979.

[57] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychol. Bull.*, 1971, vol. 76, no. 5, pp. 378–382.



Chenglong Dai received the MS, PhD degree in computer science and technology from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2014, 2020, respectively. He is currently a Lecturer at the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China. His research interests include Brain-Computer Interfaces (BCIs), EEG data mining, and machine learning. He has published several related papers in prestigious journals and top conferences, including IEEE TKDE, IEEE TCYB, ACM TKDD, ACM TIST, and SIAM SDM (awarded the Best Paper Award in Data Science Track). He also has served as a Recognition Reviewer for Knowledge-Based Systems, and for IJCNN 18-22.

Jia Wu (SM'21) received the Ph.D. degree in computer science from the University of Technology Sydney, Ultimo, NSW, Australia. He is currently an ARC DECRA Fellow in the Department of Computing, Macquarie University, Sydney, Australia. His current research interests include data mining and machine learning. Since 2009, he has published 100+ refereed journal and conference papers, including IEEE TPAMI, IEEE TKDE, IEEE TNNLS, IEEE TMM, ACM TKDD, NIPS, WWW, KDD, ICDM, IJCAI and AAAI. Dr. Wu 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 of the ACM Transactions on Knowledge Discovery from Data (TKDD) and Neural Networks (NN). He is a Senior Member of the IEEE.

Jessica J.M Monaghan received the BA and MSci in Physics from the University of Cambridge and a PhD in Electrical and Electronic Engineering from the University of Nottingham, UK. She is a Senior Research Scientist at the National Acoustic Laboratories, Sydney and an honorary Senior Research Fellow at Macquarie University. Her current research focuses on how machine learning and technology can be used tackle problems in hearing healthcare, such as improving diagnosis of hearing conditions and communication for people with hearing loss.

Guanghui Li received his PhD degree in Computer Science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005. He is currently a professor in the School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China. He has published over 70 papers in journals or conferences. His research interests include wireless sensor networks, fault tolerant computing, nondestructive testing and evaluation, and EEG processing.

Hao Peng is currently an Assistant Professor at the School of Cyber Science and Technology, and Beijing Advanced Innovation Center for Big Data and Brain Computing in Beihang University. His research interests include representation learning, social network mining and reinforcement learning. To date, Dr Peng has published over 60 research papers in top-tier journals and conferences, including the IEEE TKDE, TC, TPDS, TITS, ACM TOIS, TKDD, TIST, and Web Conference.

Stefanie I. Becker is an Associate Professor and ARC Future Fellow at the University of Queensland, Brisbane, Australia. She is an internationally recognised expert on visual attention, and received several awards for her work involving eye tracking, EEG, and fMRI. Since her PhD in 2007, she has authored or co-authored over 60 papers in high-ranking journals, and served as an associate editor for the journal JEP-HPP.

David McAlpine is Academic Director of Macquarie University Hearing, and Professor of Hearing, Language and the Brain. received his PhD in Neuroscience from the University of Oxford, UK. He is a leading expert on spatial hearing and the brain's ability to adapt to different listening environments.