BMC Bioinformatics

Biclustering fMRI time series: a comparative study

Eduardo N. Castanho, Helena Aidos and Sara C. Madeira*

*Correspondence:
sacmadeira@ciencias.ulisboa.pt

LASIGE, Faculdade de
Ciências, Universidade de
Lisboa, Lisbon, Portugal

Abstract

Background: The effectiveness of biclustering, simultaneous clustering of rows and columns in a data matrix, was shown in gene expression data analysis. Several researchers recognize its potentialities in other research areas. Nevertheless, the last two decades have witnessed the development of a significant number of biclustering algorithms targeting gene expression data analysis and a lack of consistent studies exploring the capacities of biclustering outside this traditional application domain.

Results: This work evaluates the potential use of biclustering in fMRI time series data, targeting the Region × Time dimensions by comparing seven state-in-the-art biclustering and three traditional clustering algorithms on artificial and real data. It further proposes a methodology for biclustering evaluation beyond gene expression data analysis. The results discuss the use of different search strategies in both artificial and real fMRI time series showed the superiority of exhaustive biclustering approaches, obtaining the most homogeneous biclusters. However, their high computational costs are a challenge, and further work is needed for the efficient use of biclustering in fMRI data analysis.

Conclusions: This work pinpoints avenues for the use of biclustering in spatio-temporal data analysis, in particular neurosciences applications. The proposed evaluation methodology showed evidence of the effectiveness of biclustering in finding local patterns in fMRI time series data. Further work is needed regarding scalability to promote the application in real scenarios.

Keywords: Biclustering, fMRI, Neurosciences, Time series analysis

Background

fMRI data is used to evaluate connectivity in the brain, i.e., how brain regions interact together over time. When compared to other neuroimaging techniques, such as Electroencephalography (EEG) and Magnetoencephalography (MEG) (known to have a great temporal resolution), fMRI has a high spatial resolution while maintaining a temporal resolution of a few seconds [1, 2].

There are several strategies for fMRI data analysis, such as independent component analysis, machine learning, graph theory, among other [3]. These strategies require different modeling options for the data, such as 4-dimensional arrays (capturing the

three region dimensions and the temporal one), slices of figures or multivariated time series [4, 5]. In our study, we consider the use of unsupervised machine learning techniques to analyse multivariated fMRI time series. The most popular form of unsupervised machine learning is clustering, which uses the similarity between features (columns) to partition a set of subjects (rows) into groups (clusters).

Traditional clustering approaches are rigid since they search for similarities considering all features of the subjects and do not allow a subject to belong to multiple groups [6]. Biclustering algorithms were proposed to overcome these limitations of clustering by clustering the rows and columns of a data matrix simultaneously [7]. Biclustering has been widely studied over the last 20 years following the development of the Chen and Church algorithm [8]. Multiple algorithms and computational frameworks for biclustering were developed [9], and biclustering is used in domains such as biomedicine, text mining, and marketing analysis [7, 10].

Despite these uses, the main application scenario of biclustering is the traditional gene expression context, which generates a bias in the development and application of biclustering. This bias has two main consequences: **first**, gene expression datasets act as a benchmark both during the development of new biclustering algorithms and in their independent comparison, meaning that these new algorithms are not compared in contexts other than this specific problem. **Second**, results in real datasets are usually compared using measures of biological relevance, such as the Gene Ontology (GO) annotations [11], specific to the gene expression data context, and not useful for any other context. Allied to the fact that the application of biclustering algorithms has not progressed at the same pace as the software development [9], this scenario leads to a collection of biclustering algorithms, whose potential has not been fully explored. The purpose of this study is to evaluate the potentialities of biclustering algorithms in fMRI time series.

Previous exploratory studies considering neurosciences and biclustering did biclustering on extracted features [12–15]. For biclustering, the dimensions were the analysis is done is particularly important, and in contrast to these studies, we do biclustering under the **Region** $\times$ **Time** dimensions. We propose that biclustering can be used to analyze the temporal behaviour of brain (divided into a number of regions), and extract correlation regions. The validity of this approach is suggested by a number of studies who applied biclustering to analyse data with spatial and temporal properties [16–28].

Another difference between our work and previous studies is that we do a comparative analysis of the capacities of biclustering fMRI time series considering state-of-the-art biclustering algorithms instead of testing a proposed new algorithm [14, 16, 29]. Therefore, our study resembles biclustering comparison surveys [11, 30–33] but in contrast to them, it uses internal evaluating metrics, which have wider applicability.

We selected seven state-of-the-art biclustering algorithms (covering different search strategies) and applied them to artificial and real-world fMRI datasets. Additionally, we added to this comparison two variations of three popular clustering algorithms, k-means, spectral, and ward's hierarchical method. In the absence of ground truth to evaluate biclustering/clustering solutions, we opted for internal evaluation metrics and used them to determine the type of patterns to expect in fMRI data.

This study is a comparative study on the application of biclustering in a different biomedical domain than gene expression data analysis. In addition, it proposes a methodology to evaluate biclustering. Besides neurosciences, we suggest that climate science, epidemiology, and sociology, sharing fundamental properties with the former, are scientific fields expected to benefit from this study.

The remaining of this section targets biclustering and related work. First, we position biclustering in the family of unsupervised learning algorithms, describe the relevant biclustering patterns found in real-valued matrices, and categorize biclustering algorithms based on their search strategy. Then we revise previous work on biclustering reviews. Second, we motivate the use of biclustering and review studies about the use of biclustering in neurosciences and spatio-temporal analysis.

The remaining of this paper is organized as follows: In **Methods**, we describe the general comparison methodology followed in our study and how it is different from other comparison studies. In **Results** we introduce and discuss our results. In **Discussion**, we show implications of our study in neurosciences. In **Conclusion**, we give concluding remarks.

Biclustering

Biclustering can be defined as simultaneous clustering of rows and columns [7], allowing the discovery of rows with similar behavior across a subset of columns, and vice-versa (biclusters). Considering a data matrix A, with n rows ($X = \{x_1, \ldots, x_n\}$) and p columns ($Y = \{y_1, \ldots, y_p\}$), the biclustering task aims to discover subsets B of the original data matrix A. Each subset B is called a bicluster, and consists of a subset of rows I, and a subset of columns J, where $I \subseteq X$ and $J \subseteq Y$.

Bicluster patterns

Biclustering is able to find more flexible structures than traditional clustering. While the nomenclature for this possible structures is not standardized in the literature, the definitions provided by Madeira and Oliveira [7] and Aguilar-Ruiz [34] are commonly used. The most simple structure is the **constant bicluster**, which is a submatrix (I,J) where all values b_{ij} are equal: $b_{ij} = \pi$.

In the context of a real-valued matrix, it is of particular interest to discover coherent biclusters. A coherent bicluster uses a row parameter π_i and a column one β_j. The interaction between the two parameters can be either additive (shifting) or multiplicative (scaling). A perfect **shifting bicluster** follows the additive relation among the variable: $b_{ij} = \pi_i + \beta_j$. A **scaling bicluster** follows a multiplicative relation: $b_{ij} = \pi_i \times \beta_j$.

Search strategies

When categorizing biclustering algorithms, a particular concern is related to the search strategies. Table 1 systematizes four main search strategies, defined by Madeira and Oliveira [7] and used in recent comparison studies, such as Padilha and Campello [11] and Henriques et al. [33].

Table 1 Biclustering search strategies as defined by Madeira and Oliveira [7]

Category	General characteristics	Examples
Greedy	Biclusters are generated by adding or removing columns to a initial random bicluster in order to improve some gain function. The final objective is for the algorithm to find a global minimun solution after some iterations. Despite making wrong decisions, and loosing good biclusters due to beeing stuck in local minima, they have the potential of being fast algorithms	ISA XMotifs [35, 36]
Distribution parameter identification	Assume some statistical model behind the data, and then apply some iterative procedure in order to obtain its parameters by minimizing some criterion	FABIA spectral biclustering [37, 38]
Divide and conquer	Divide the original data matrix into smaller instances. With the potential of being very fast, they could fail to find good biclusters, splitted before identified	Bimax [30]
Exhaustive	Based on the premise that finding the best biclusters can only be done by using an exhaustive enumeration of all possible biclusters in the data matrix. Despite being able to find the bests biclusters they do it by imposing restrictions to the biclusters size (since these algorithms are typically very slow)	BicPAM CCC [39, 40]

Reviews

During the last two decades several authors reviewed biclustering methods [6, 7, 9–11, 13, 30–33, 41–44]. These reviews can be fitted into three main categories: (1) General Surveys [6, 7, 9, 10, 13, 33, 44], (2) Comparison Surveys [11, 30–33] and (3) Measure Surveys [41–43, 45].
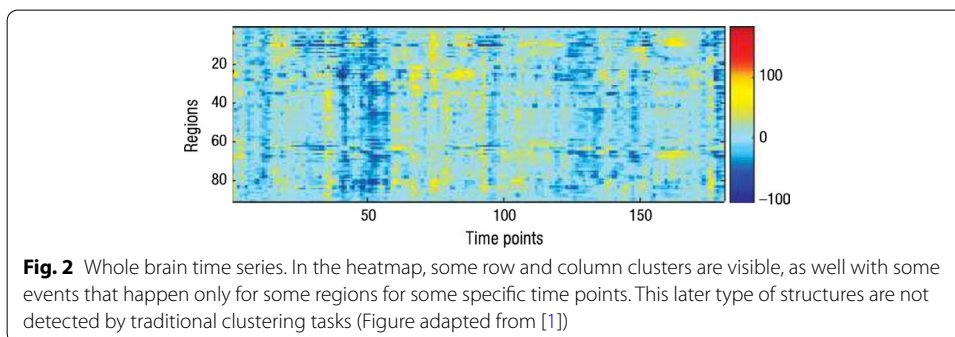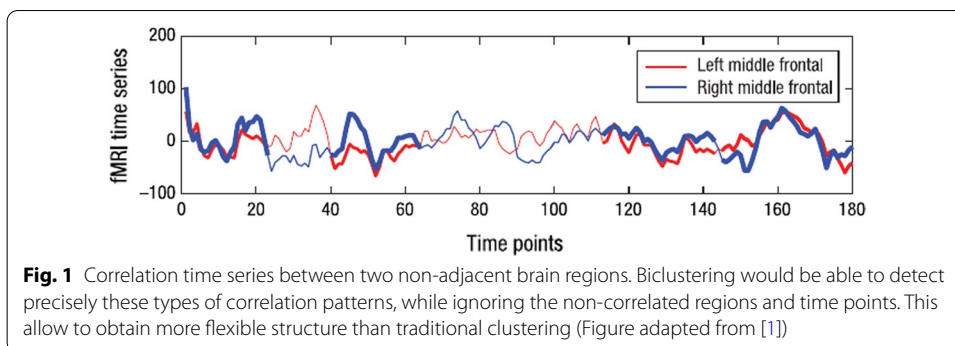
General survey studies categorize biclustering, providing abstract categorizations and reviewing state-of-the-art approaches and typical application cases. These studies are beneficial as a the first look into biclustering.

Comparison surveys, approach the task of finding the "best" biclustering algorithm. Prelić et al. [30] was the first to compare multiple biclustering algorithms by adapting methodologies used in clustering. They choose five biclustering methods, proposed a new algorithm, and compared their performance using synthetic and real gene expression data. The validation was performed using external indices on synthetic data and biological relevance. Bozdağ et al. [31] compared 5 biclustering algorithms following the approach of Prelić et al. [30], but considering the effects of noise, size and bicluster overlap. Eren et al. [32] expanded the research by comparing 12 algorithms, using synthetic datasets following six different biclustering models and eight gene expression datasets. Henriques et al. [33] compared the performance of pattern-based biclustering to the traditional approaches, evaluating a total of 15 state-of-the-art approaches. Their evaluation was done in synthetic and real datasets, considering two external measures and computational efficiency together with biological relevance of results. Padilha and Campello [11] used three synthetic data collections and two real data collections to analyze the performance of 17 algorithms.

Finally, **measure surveys** discuss metrics to evaluate biclustering.

Motivating biclustering

The traditional application area for biclustering is in the context of gene expression data analysis. Therefore the number of applications of biclustering in neuroimaging data

Fig. 1 Correlation time series between two non-adjacent brain regions. Biclustering would be able to detect precisely these types of correlation patterns, while ignoring the non-correlated regions and time points. This allow to obtain more flexible structure than traditional clustering (Figure adapted from [1])

Fig. 2 Whole brain time series. In the heatmap, some row and column clusters are visible, as well with some events that happen only for some regions for some specific time points. This later type of structures are not detected by traditional clustering tasks (Figure adapted from [1])

is limited. This section reviews the most closely related studies. Despite the lack of a research line using biclustering and fMRI data, there are hints in the literature that suggest opportunities for these algorithms to be applied in these data.

Since the brain structure is dynamic, different brain regions will fire together based on different stimuli [1, 5]. A flexible approach is needed to identify brain regions that fire together under different time points, and can be delivered by biclustering. Figure 1 shows the time-series signal from two non-adjacent brain regions, showing that the correlation between two regions is not static and shows only in some specific time points. Figure 2 shows a whole-brain time-series heatmap, where biclusters-like structures are visible. The biclustering task promises to identify these brain regions interacting together over the time points.

The remaining of this section highlights related studies where biclustering was used. In the first part of this section, we review the various studies that associated neurosciences and biclustering, not necessarily the same type of data as the one we cover. Then, we explain how biclustering is used to analyze data with spatio-temporal properties. While these study cases are outside medical data, they motive the capacities of these methods on temporal data.

Biclustering in neurosciences

Busygin et al. [12] used a form of biclustering known as consistent biclustering to analyze EEG data and identify a physiologic marker for optimal vagus nerve stimulator parameters. They conducted biclustering under the **Label** × **Features** dimensions. The labels were created considering experimental settings, and the features

are associated with the EEG channels. Fan et al. [13] cited this EEG study as an example of biclustering application, but it is a general-purpose review about biclustering.

Considering structural MRI data, Rahaman et al. [14] proposed a new biclustering approach under the **Subject** $\times$ **Features**. In this study, the features are computed considering the application of independent component analysis (ICA) of the MRI spatial maps. Gupta et al. [29] proposed a new biclustering approach in the **Subject** $\times$ **Voxel** dimensions to analyse structural MRI data, using an intermediate ICA step.

Yu et al. [15] conducted biclustering analysis to explore relations among brain functional network connectivity (FNC), obtained by using fMRI data and genetic nodes (SNP). Therefore, the biclusters had the **FNC** $\times$ **SNP** shape.

In biclustering, the dimensions of analysis are a important part of the analysis. Each one of the previous studies use different dimensions in their analysis, and a different set of dimensions than we do in this study. The study most closely related to our research is the one conducted by Amar et al. [16], that proposed a new algorithm based on biclustering to perform three-way analysis (triclustering). As an application example, they used multi-subject fMRI data (**Region** $\times$ **Time** $\times$ **Subject**) as an illustrative application of their approach. Their results suggest the capacity of this family of algorithms to identify pertinent brain regions. However, this study is only explorative, meaning that this algorithm's performance is not tested against other three-way clustering algorithms or approaches.

Biclustering spatio-temporal data

In more recent years, more general approaches than clustering have been successfully used in distinct Spatio-temporal application domains. This section focus on the uses of biclustering and triclustering [46] to analyze data with spatio-temporal properties.

Considering biclustering, Wu et al. [17] analysed Chinese meteorological Spatio-temporal data. Shen et al. [18] used biclustering to analyze the global distribution of natural disasters. Kaban et al. [19] used biclustering to identify spatial of social vulnerability in Indonesia and was able to distinguish profiles of social vulnerability. Borgnat et al. [20] identified spatial and temporal profiles in a bike-sharing system in Lyon. Izenman et al. [21] used biclustering to group juvenile-offense data. Dhamodharavadhani and Rathipriya analysed COVID-19 epidemiological data and identified countries with similar epidemic trajectory patterns [27]. Neves et al. [28] used a biclustering algorithm to analyse traffic data.

Considering triclustering, Martínez-Álvarez et al. [22] analyzed parameters to seismogenic zoning using triclustering in the Iberian Peninsula. Guigourès et al. [23] proposed a technique to analyze time-varying graphs, illustrating it in the London bike-sharing system. Wu et al. [24, 25] used triclustering to analyze meteorological data from Duch weather stations. Melgar-García et al. [26] applied a triclustering-based algorithm to discover patterns over time in maize crops in Portugal to help farmers improve their harvests.

It is also relevant to refer that several biclustering algorithms were developed with a focus on temporal data [40, 47–49].
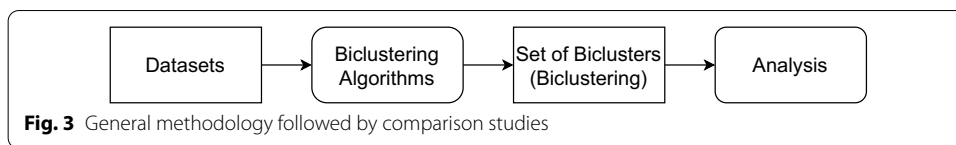
Fig. 3 General methodology followed by comparison studies

Table 2 Data collections used to evaluate the performance of the biclustering algorithms. We have a total of 42 datasets, each one representing one brain scan

Data collection name	Nature	#Brain scans	#Time points	# Regions	References
"First data collection"	Artificial	1	150	≈ 30	[50, 51]
"Artificial data"	Artificial	20	150	≈ 30	[50, 51]
"Real data"	Real	20	94	463	[16, 52]
"Illustrative data"	Real	1	137	45	[53–55]

Methods

Comparisons surveys are a category of studies focused on evaluating the perfomance of biclustering algorithms [11, 30–33]. These studies begin by selecting a set of biclustering algorithms and evaluate their perfomance on some dataset. These studies follow a common general strategy that we sistematize in Fig. 3.

In this section, we explain how our study follows this general methodology. For each sub-task, we explain our criteria for each decision and how it distinguishes from other comparison surveys in biclustering [11, 30–33].

Datasets

Previous studies use two types of data collections: synthetic datasets, where ground truth is known, and real data collections [11, 30–33]. The first is used so that an external metric can be used to evaluate the capacity of the biclustering algorithms to retrieve the artificially-planted biclusters in the dataset [41]. Since data is synthetic, the data parameters can be controlled (such as data size, presence of noise, number of planted biclusters, and possibility of overlapping). The real data collections consists of gene expression datasets from various benchmarks. Typically, the number of datasets in each comparative study is around a few dozens gene expression datasets.

This is a sub-task where our study differ most from previous studies. In our study we work exclusively with fMRI time series datasets, where each dataset represents a brain scan (we use the dimensions **Region** $\times$ **Time**). We use real data and artificial fMRI scans. In our study we use a total of 42 datasets, organized as four data collections, summarized in Table 2.

The **First data collection** has a single artificial brain scan, and was used to test biclustering parameters. The **Artificial data** consists of 20 artificial brain scans, which we use to evaluate the performance of the biclustering algorithms. The artificial datasets were obtained using the SimTB simulator [50] to generate data consistent with an auditory oddball experiment [56]. This approach was validated in previous studies and is part of the SimTB software [50, 51]. The **Real data** consits of 20 brain scans, and is also used to evaluate the performance of the biclustering algorithms. The data was collected by

Vaisvaser et al. [52] to analyze stress response. This data was preprocessed by Amar et al. [16].[1] The **Illustrative data** corresponds of a single brain scan, extracted from the "NYU Slow Flanker" dataset [53–55] and is used exclusivelly to illustrate results from a biclustering algorithm.[2] Typical preprocessing was conducted using the FSL software [57–59]. Finally, the fMRI brain scan was downsampled into regions considering the Harvard-Oxford cortical atlas [60].

Biclustering algorithms

Dozens of biclustering algorithms have been developed during the last two decades [9]. Since it is not viable to test every single available biclustering algorithm, decisions regarding the best algorithm to use must be made. There is not a single strategy to decide what algorithms to test.

Prelić et al. [30] considered three criteria to chose biclustering algorithms: (1) the popularity of the algorithms in the biclustering community, (2) considering similar algorithmic strategies to be better comparable, and (3) the availability of the implementation. Eren et al. [32] choose algorithms considering both the convenience criteria (algorithms that have available implementations) and the criteria of having various biclustering algorithms with differing approaches. Henriques et al. [33] selected what they considered to be the state-of-the-art biclustering algorithms. Padilha and Campello [11] referred the popularity criteria and the availability of implementation when selecting algorithms.

For our study we considered the following criteria to select biclustering algorithms for analysis:

- We began by considering only popular, freely available implementations of biclustering algorithms. We analysed both the previous comparative studies and the studies considering biclustering time series [11, 30–33, 61].
- We considered important for our algorithms to detect shifting and scaling patterns (patterns commonly found on real-valued datasets).
- The algorithms should be robust to noisy datasets (since there is a lot of expected noise in fMRI time series data).
- The algorithms should cover different search strategies.

We selected seven state-of-the-art biclustering algorithms covering four search strategies. Table 3 highlights general characteristics of these algorithms together with the reasons for their selection.
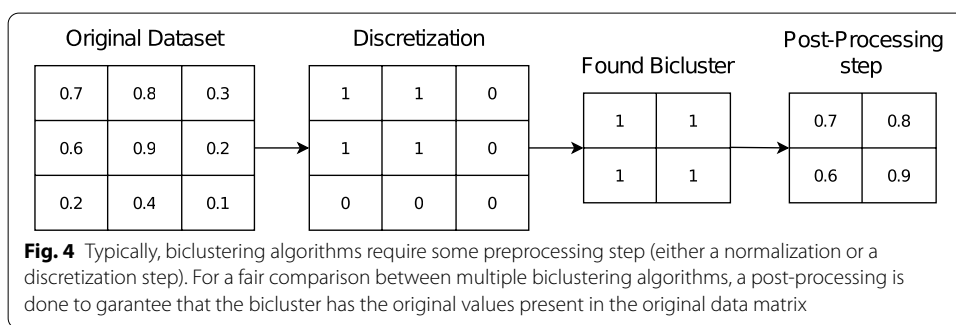
Since clustering can be viewed as a particular case of biclustering, we can compare biclustering approaches to traditional clustering. Clustering approaches are typically used to group brain-regions with similar activity over time [3, 65–67]. Among the most popular clustering approaches are the K-means algorithm, Spectral Clustering, and Hierarchical methods (in particular using the **Ward's algorithm**) [66]. However, both row and column clustering make sense and can be considered. In what follows,

---

[1] Original data available at: acgt.cs.tau.ac.il/twigs/.

[2] Original data available at the openneuro portal: openfmri.org/dataset/ds000102/.

Castanho *et al. BMC Bioinformatics*   (2022) 23:192

Page 9 of 30

Table 3 Biclustering algorithms considered for this study. Additionally, they will be compared to three popular clustering algorithms: k-means, spectral, and ward's hierarchical methods. For clustering, we use scikit-learn implementations [64]

Algorithm	Type of search	Available at	References	Reason to choose it
BicPAM	Exhaustive	BicPAMS	[39, 62]	State-of-the-art pattern mining based biclustering method
CCC	Exhaustive	BiGGEsTS	[40, 61]	Allows to obtain temporal contiguous biclusters efficiently
ISA	Greedy	isa2	[35]	State-of-the-art greedy algorithm able to deal with real data
XMotifs	Greedy	biclust	[36, 63]	State-of-the-art greedy algorithm based on a strategy of discretizating data
Bimax	Divide and conquer	biclust	[30, 63]	Very fast algorithm able to detect simple structures
FABIA	Distribution parameter identification	FABIA	[38]	State-of-the-art algorithm
Spectral Biclustering	Distribution parameter identification	biclust	[37, 63]	State-of-the-art algorithm able to detect a specific type of bicluster structures

Fig. 4 Typically, biclustering algorithms require some preprocessing step (either a normalization or a discretization step). For a fair comparison between multiple biclustering algorithms, a post-processing is done to garantee that the bicluster has the original values present in the original data matrix

and considering data format, traditional row clustering and column clustering will be referred as **region clustering** and **temporal clustering**, respectively.

Biclustering

After running a biclustering algorithm over a dataset, a set of biclusters (biclustering) is generated. Previous studies do not discuss much of this step since it is closely related to the biclustering parametrization step. However, the biclusters must have a format adequate for the posterior analysis.

Different biclustering algorithms will provide biclusterings with different structures. For example, Bimax requires a binary discretization of the original dataset. Therefore, the output of the algorithm will be a binary bicluster. Our strategy was to consider the set of rows and columns obtained by the algorithm and the original values present in the dataset to define the biclusters to be evaluated. Figure 4 illustrates this process.

Some biclustering algorithms can generate biclusters with less than two rows or columns. In the context of our study, these biclusters were considered uninteresting and

were removed from the biclustering, either by parametrization or considering a post-processing step.

For the artificial and real data collections (20 datasets each), biclustering was conducted independently for each dataset. For this study purposes, we consider the union of all generated biclusters as the result for that data collection.

Analysis

Traditional comparative studies use both the external evaluation metrics applied to the synthetic data and the biological significance applied to the real-world gene expression datasets [11, 30–33].

Our data focus on fMRI time series. External measures cannot be used due to the absence of ground truth. Furthermore, biological significance measures are exclusive to the context of gene expression data and not useful for any other context. Therefore our analysis methodology must be done differently. Internal evaluation indexes were selected to evaluate the performance of biclustering algorithms [43].

Our first task is related to the parametrization of the algorithms. For most parameters, we used the default ones provided by the implementations or suggested by the authors. We decided to test different parameters and select the best performance configurations. This approach is consistent with the standard practice followed by previous studies, which use mostly default parameters, discussing only the effects of some parameters [11, 30, 32, 33]. We use these optimized parameters during the remaining of the study. For this part, the first data collection was used.

Then, we conduct a comparison between biclustering algorithms. This comparison is an essential part of our study since it compares the biclustering capacity against traditional clustering and the capacity of each algorithm. Additionally, we analyze the dimensions of the biclusterings, which are distinct between algorithms and can explain the reasons for the previous performance. Finally, we use the correlation between our metrics to hypothesize the patterns hidden in fMRI time series. For this analysis, we used the artificial data collection and the real data collection.

Evaluation metrics

In our study, we use internal evaluation indexes, that translate the internal quality of the bicluster into a single number [43]. However, since there are multiple types of patterns that fall into the definition of bicluster, there is no single metric that can capture all types of patterns. As stated by Hartigan [68], one simple hypothesis to evaluate a bicluster is to calculate its **Variance (VAR)**:

$$VAR(B) = \frac{1}{|I| \times |J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} (b_{ij} - b_{IJ})^2, \tag{1}$$

where b_{ij} refers to the element (i, j) of the bicluster B, b_{IJ} refers to the global mean of the elements belonging to the data matrix and $|I|$, $|J|$ represent the total number of rows and columns of the bicluster. A disadvantage of using the variance is that it is only able to measure constant biclusters. Therefore it will fail to detect most of the more complex patterns we want to detect.

Introduced by Cheng and Church [8], the **Mean Squared Residue (MSR)** aims to find the coherence of the bicluster over rows and columns:

$$MSR(B) = \frac{1}{\mid I \mid \times \mid J \mid} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} (b_{ij} - b_{iJ} - b_{Ij} + b_{IJ})^2, \tag{2}$$

where b_{iJ} and b_{Ij} point to the row i mean and the column j mean. This approach is an improvement compared to the variance since it can capture shifting tendencies. However, it is unable to capture scaling patterns [31].

In order to capture the scaling tendencies that the MSR cannot recognize, Mukhopadhyay et al. [69] developed a new measure called **Scaling Mean Squared Residue (SMSR)**, focused on scaling patterns (however, it fails to capture the shifting patterns).

$$SMSR(B) = \frac{1}{\mid I \mid \times \mid J \mid} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} \frac{(b_{iJ} \times b_{Ij} - b_{ij} \times b_{IJ})^2}{b_{iJ}^2 \times b_{Ij}^2}. \tag{3}$$

An important property of biclusters is the possibility that they could have a different range of values between each other, suggesting the use of standardisation processes. As pointed by Pontes et al. [43], an advantage of this approach is to characterise their tendency. One way of doing it is to standardize data by row. In the context of fMRI, this means normalizing each brain region to have a similar amplitude as follows:

$$\hat{b}_{ij} = \frac{b_{ij} - \mu_{g_i}}{\sigma_{g_i}}, \quad 1 \leq i \leq \mid I \mid, 1 \leq j \leq \mid J \mid, \tag{4}$$

where μ_{g_i} and σ_{g_i} represent the mean and standard deviation over the rows of the bicluster. Based on this approach, a quality measure was proposed under the name of **Virtual Error (VE)** [34, 70], considering the concept of a virtual pattern, which can be defined over rows or columns [43]. The normalized pattern over rows is given by:

$$\hat{\rho}_j = \frac{1}{\mid I \mid} \sum_{i=1}^{|I|} \hat{b}_{ij}. \tag{5}$$

This pattern represents the average brain behaviour over time, and the virtual error is thus defined as a measure of the difference between the real region behaviour compared to this virtual row:

$$VE(B) = \frac{1}{\mid I \mid \times \mid J \mid} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} \mid \hat{b}_{ij} - \hat{\rho}_j \mid. \tag{6}$$

A particular advantage of the virtual error measure is its capacity to detect both shifting and scaling patterns [43]. This advantage means that it is a suitable internal measure to compare different possible patterns obtained by different biclustering methods. When considering the presence of noise, Pontes et al. [43] pointed out that the virtual error can capture both scaling and shifting patterns since its value varies linearly with the induced error. Additionally, this metric is of particular interest since it uses the concept of pattern over a "time" dimension.

Types of signal

While the virtual error is expected to capture both shifting and scaling patterns [43], it does not provide any knowledge in ther of the specific type of the signal found in the fMRI data. Padilha and Carvalho [71], in a study consisting of 9 biclustering algorithms and a benchmark of 19 different real gene expression datasets, showed that the virtual error is not fundamentally correlated with any of the remaining discussed measures (variance, MSR and SMSR).

Since the virtual error promises to be a stable measure for different types of biclusters, we expect it to be correlated with the measure able to find the most common pattern present on our data. Therefore, we used the Pearson coefficient between the three mentioned quality measures (VAR, MSR and SMSR) and the virtual error measure to get insights patterns found in our data.

Results

This section discusses the performance of biclustering algorithms in our evaluation scenarios.

We begin by testing the influence of different parameters on the performance of each algorithm, to select an adequate set of parameters for the algorithms to be compared with each other. Then, we evaluate the perfomance of the biclustering algorithms considering the following criteria: (1) biclustering versus region and temporal clustering; (2) different biclustering algorithms; (3) top-K biclusters; (4) bicluster size; (5) type of bicluster pattern found.

Testing configurations

In this scenario, the objective was to test the behaviour of different parameter configurations on six bicluster algorithms (BicPAM, FABIA, Bimax, CCC, XMotifs and Spectral). The First data collection was used for this step, and Table 4 shows the results of this testing.

In terms of input, the format Region × Time was selected to run most of the algorithms. This approach was chosen since the algorithms are in general implemented to run in this configuration, since they were designed in gene expression context, where genes are commonly used in the rows and the conditions, such as time, fit in the columns [7].

Considering the input format, an exception was made for **BicPAM**, since this algorithm was reported to be more efficient for matrices with a larger number of rows than columns (which for this dataset consists of the Time × Region format), and allows to select a dimension to search for patterns (the temporal dimension was chosen) [39, 62]. Due to its flexibility of finding different types of patterns, BicPAM was also run in three different configurations to search for constant, shifting and scaling patterns. The *minimum number of biclusters before merging* parameter was selected as high as possible in order to guarantee an adequate exploration of the dataset, while forcing it to run in a reasonable time. Additionally, BicPAM has a discretization step, and we used the default five symbols discretization. Based on the virtual error results

Castanho *et al. BMC Bioinformatics* (2022) 23:192

Page 13 of 30

Table 4 Median values for the four selected measures for the first artificial dataset, with uncertainties given by the standard deviation (except for the case of SMSR were the standard deviation is orders of magnitude higher than the median value). From these results it is visible that **A)** The high values of uncertainty discourage focus on optimizing the biclustering method parameters and **B)** Choosing the right evaluation metric is important, however in most of the biclusters cases they seem to agree for the same "best" configuration. Bold represents the choosen parameters for the next sections

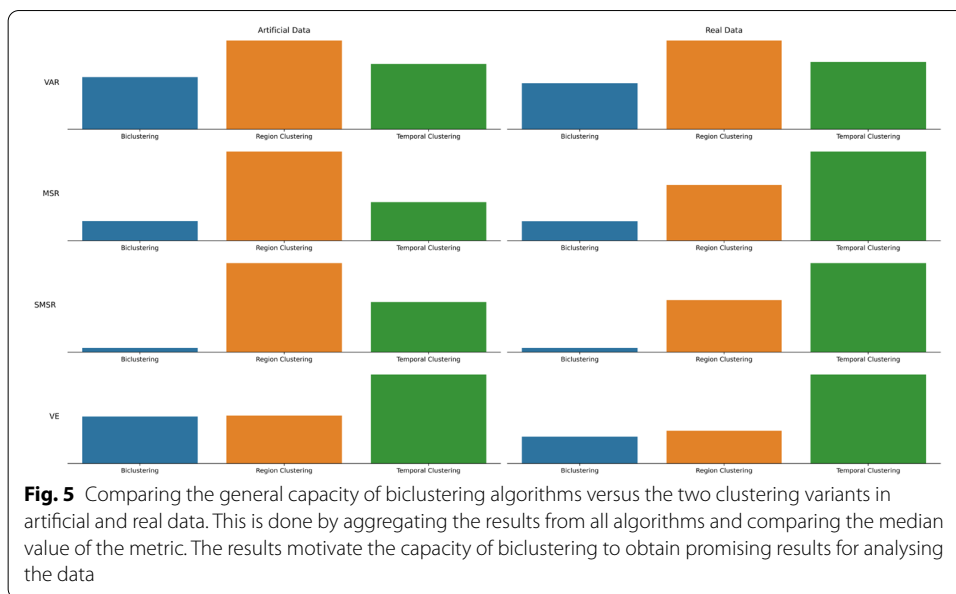| Method	Configuration	VAR	MSR	SMSR	VE
BicPAM	**Additive version**	0.024 ± 0.024	0.001 ± 0.003	2.49	**0.262** ± 0.232
	Constant version	0.014 ± 0.026	0.001 ± 0.002	1.96	0.434 ± 0.185
	Multiplicative	0.025 ± 0.024	0.002 ± 0.005	3.07	0.461 ± 0.263
Bimax	10 biclusters	0.004 ± 0.002	0.001 ± 0.001	0.03	0.881 ± 0.148
	100 biclusters	0.006 ± 0.004	0.003 ± 0.002	0.05	0.729 ± 0.128
	1000 biclusters	0.008 ± 0.008	0.004 ± 0.004	0.06	0.706 ± 0.168
	10,000 biclusters	0.008 ± 0.007	0.004 ± 0.004	0.06	**0.695** ± 0.169
	100,000 Biclusters	0.008 ± 0.007	0.004 ± 0.004	0.06	0.695 ± 0.169
CCC	**Traditional discretization (5 symbols)**	0.189 ± 0.359	0.017 ± 0.021	0.51	**0.288** ± 0.345
	Variation between time points (2 Symbols)	0.596 ± 0.491	0.067 ± 0.108	0.93	0.370 ± 0.315
	Variation between time points (3 Symbols)	0.596 ± 0.491	0.067 ± 0.108	0.93	0.370 ± 0.315
FABIA	**Standard**	0.579 ± 0.502	0.006 ± 0.081	0.51	**0.079** ± 0.167
	Relaxed	0.946 ± 0.053	0.881 ± 0.095	3485.60	0.787 ± 0.036
Spectral Biclustering	**log**	0.039 ± 0.053	0.033 ± 0.004	670.30	**0.726** ± 0.012
	bistochastization	0.040 ± 0.005	0.035 ± 0.004	3232.21	0.731 ± 0.015
	irrc	0.040 ± 0.004	0.035 ± 0.004	1271.71	0.726 ± 0.015
XMotifs	Discretization with 2 symbols	0.017 ± 0.004	0.012 ± 0.001	1168.76	0.662 ± 0.048
	Discretization with 5 symbols	0.007 ± 0.009	0.003 ± 0.001	6.59	**0.585** ± 0.035

between constant, shifting and scaling patterns, we choose the "additive" version of BicPAM.

In **FABIA** the number of biclusters is always limited: it cannot be higher than the number of rows and the number of columns. In this case, we selected the number of columns of the dataset as a number of datasets (following the methodology proposed by Henriques et al. [33]). We considered two threshold options: (1) default thresholds, and (2) relaxed values to try to find more biclusters. This treeshold influences the size and number of generated biclusters.

We observed that the relaxed configuration achieves far worse results than the standard configuration. This suggests that the strategy of relaxing the thresholds does not provide more meaningful results, therefore we used the default threshold values.

In **Bimax** the default values were used but for the number of biclusters, changed as follows: 10 biclusters, 100 biclusters, 1000 biclusters, 10,000 biclusters and 100,000 biclusters.

We observed that generating more biclusters mean adding more noise to the previous generated ones. Additionally, we observed an effect of saturation when generating a large number of biclusters, since asking the algorithm to generate "100,000" biclusters does not produce actually any more biclusters than the "10,000" option.
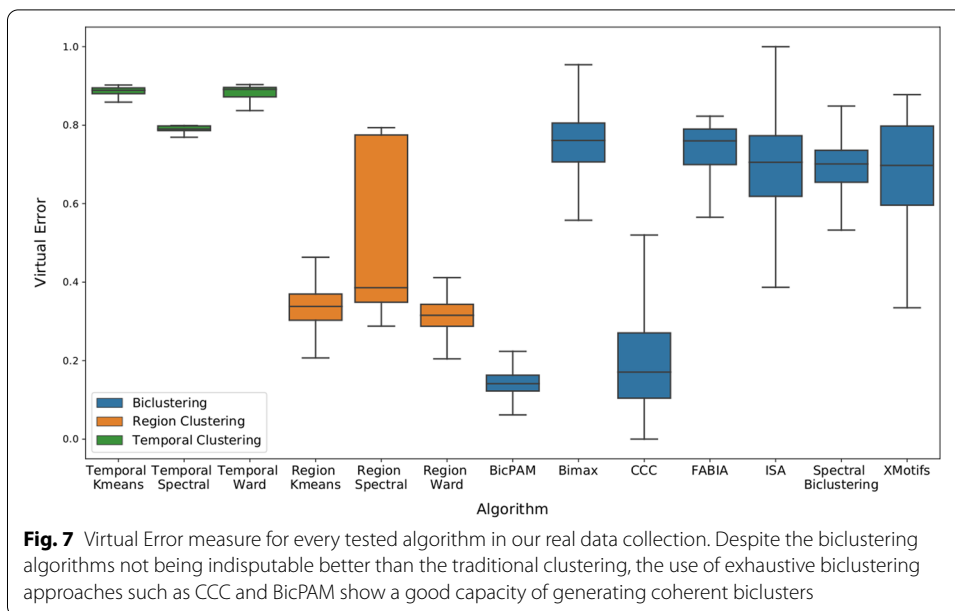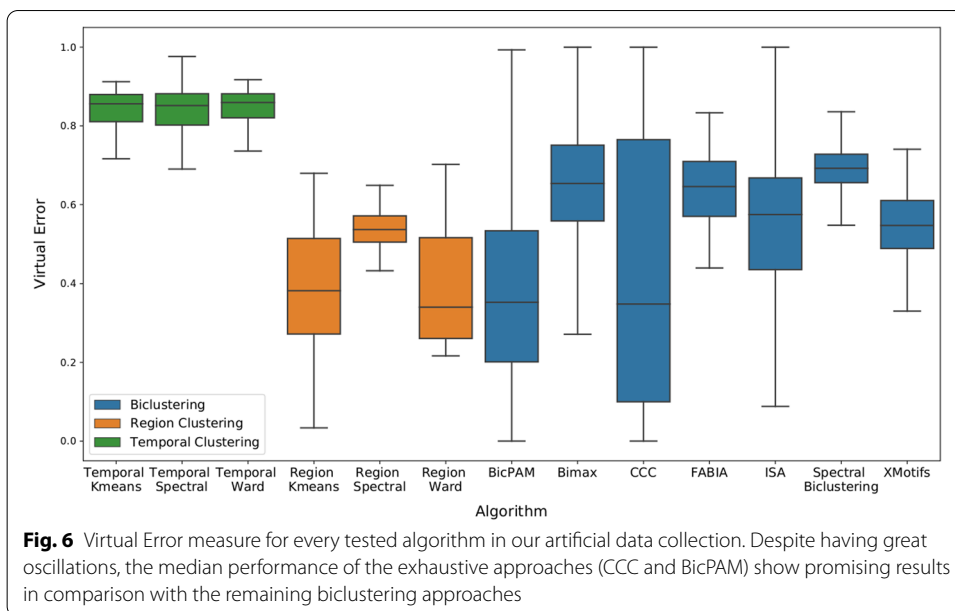
Fig. 5 Comparing the general capacity of biclustering algorithms versus the two clustering variants in artificial and real data. This is done by aggregating the results from all algorithms and comparing the median value of the metric. The results motivate the capacity of biclustering to obtain promising results for analysing the data

The greedy algorithm **XMotifs** was executed 30 times (to avoid interference from the starting seeds, and following the approach proposed by Padilha and Campello [11]). XMotifs uses a discretization step, and 5 symbols were used (in order to be similar to the discretizations used by CCC and BicPAM), as well as 2 symbols (to be similar to Bimax). The strategy of using a discretization with 5 symbols provided biclusters with smaller errors than using a binary discretization. Similar results were obtained by Kemal et al. [32].

Spectral biclustering uses a normalization method. We used the three normalization methods made available by the authors: "logarithmic normalization" (log), "independent rescaling of rows and columns" (Irrc) and "bistochastization" [37]. The three configurations generate similar results in all metrics (except SMSR). The "log" configurations achieved the best performance in all four metrics, thus since it is also the configuration recommended by the authors, it was selected for the next stages of research.

In **CCC** most of the parameters are associated with the discretization step. We used the two major possibilities: use a traditional discretization per row (for this, 5 symbols was used) or use variation between time points (2 or 3 symbols) as described in Madeira et al. [40]. A surprising result is the traditional discretization to generate best biclusters than the variations options, contradicting results obtained by Madeira et al. [40]. This could be due to two possible reasons: different application context or the use of a different evaluation metric.
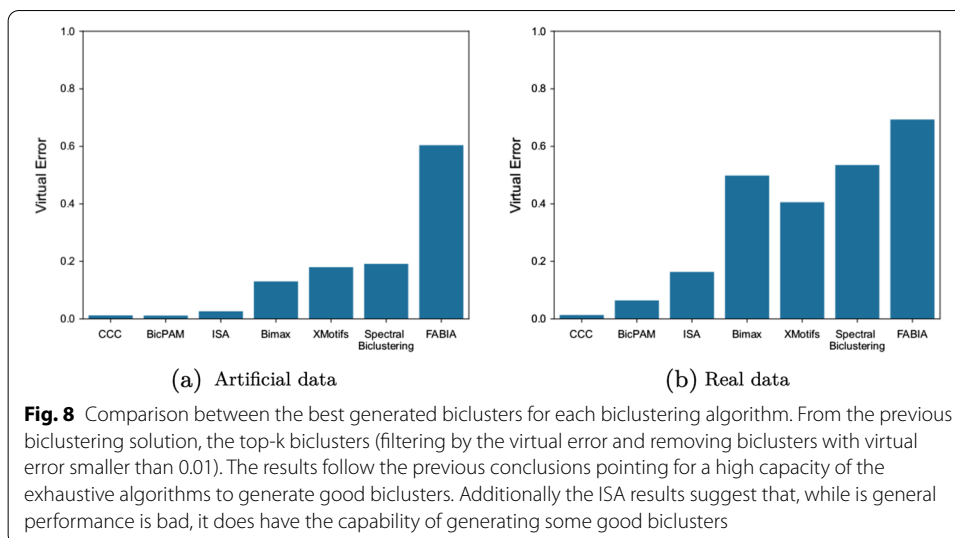
Comparing biclustering and clustering

In this scenario, the general objective is to compare the general capacity of biclustering algorithms with traditional clustering algorithms in both region and temporal clustering. Results are summarized in Fig. 5. The general conclusion anticipates biclustering to obtain more homogeneous structures than the traditional clustering structures, since it achieves better values in all four quality evaluation criteria.

Fig. 6 Virtual Error measure for every tested algorithm in our artificial data collection. Despite having great oscillations, the median performance of the exhaustive approaches (CCC and BicPAM) show promising results in comparison with the remaining biclustering approaches

Fig. 7 Virtual Error measure for every tested algorithm in our real data collection. Despite the biclustering algorithms not being indisputable better than the traditional clustering, the use of exhaustive biclustering approaches such as CCC and BicPAM show a good capacity of generating coherent biclusters

Comparing algorithms

While previous results motivate the use of biclustering, they do not explain the capacity of each individual algorithm to generate homogeneous biclusters. In this scenario we tested the performance of each biclustering (and clustering) algorithm. Figures 6 and 7 illustrate our results in both synthetic and real data.

The first immediate conclusion is related with the lack of capacity of temporal clustering algorithms to generate good results. This makes sense since these clustering scenarios generate groups of time points where all regions behave similarly. Since the brain is heterogeneous, different regions will behave differently under the same time

Fig. 8 Comparison between the best generated biclusters for each biclustering algorithm. From the previous biclustering solution, the top-k biclusters (filtering by the virtual error and removing biclusters with virtual error smaller than 0.01). The results follow the previous conclusions pointing for a high capacity of the exhaustive algorithms to generate good biclusters. Additionally the ISA results suggest that, while is general performance is bad, it does have the capability of generating some good biclusters

points, it was not expected that clustering would generate good results considering the virtual error measure.

More comparable are the tasks of region clustering and biclustering. In this scenario, most biclustering algorithms achieve worse results than the traditional clustering tasks, with the exception of exhaustive biclustering algorithms. However, the capacity of the exhaustive approaches (BicPAM and CCC) to find homogeneous biclusters shows that while the clustering methods are fundamentally limited (due to the cluster structure), the biclustering task is able to achieve very good results.

Focusing on the individual performance, it is not too surprising that the Bimax algorithm would achieve bad results since it discretizes data in a very specific way (it searches for binary biclusters). This step of turning the dataset into a binary one strongly limits the capacity of finding good biclusters. For Spectral Biclustering results are not surprising, since it does search for a very specific type of bicluster. FABIA has similar problems when compared to clustering approaches, since its factor-analysis approach strongly limits the capacity of the algorithm of generating biclusters. Both ISA and XMotifs produce bad results, which is suprising since they were expected to find the same types of biclusters than the exhaustive algorithms.

Comparing top-K biclusters

Biclustering solutions are composed of a different number of generated biclusters, which can be order of magnitude different depeding of the algorithm. Additionally, as part of the methodology, greedy algorithms were executed multiple times to avoid interference from their stochastic nature. Therefore, comparisons that use every single bicluster are not be fair. In this section, our strategy was to use the virtual error as a filter and select the *top-K* biclusters of each algorithm. Additionally, to avoid statistical artifacts, biclusters with virtual error smaller than 0.01 were removed from our study.

In our study, we selected the *K* value empirically considering the number of already generated biclusters. For the artificial data, the *K* number was selected as 50, and for the real dataset, the selected number was 500. Our results are shown in Fig. 8 and reinforce

Table 5 Median values (and associated standard deviation) for the typical bicluster dimension parameters in both data collections: number of regions in each bicluster, number of time points and bicluster area. When comparing this results to the virtual error values, a apparent relation comes between the bicluster size and the associated virtual error, which make sense

Algorithms	Artificial data			Real data		
	Time points	Region points	Area	Time points	Region points	Area
Biclustering						
Bimax	6 ± 7	4 ± 2	24 ± 25	6 ± 2	21 ± 21	126 ± 386
BicPAM	7 ± 3	2 ± 1	16 ± 11	27 ± 8	2 ± 0	54 ± 16
CCC	3 ± 4	3 ± 2	10 ± 9	5 ± 2	4 ± 9	18 ± 22
FABIA	43 ± 5	4 ± 2	184 ± 76	29 ± 4	14 ± 52	403 ± 1204
Spectral Biclustering	23 ± 11	7 ± 3	136 ± 112	20 ± 7	17 ± 6	330 ± 211
ISA	14 ± 9	3 ± 1	44 ± 46	7 ± 5	30 ± 28	152 ± 276
XMotifs	49 ± 13	2 ± 1	114 ± 45	19 ± 11	4 ± 15	78 ± 71
Region clustering						
kmeans	150	3 ± 2	450 ± 283	94	4 ± 1	376 ± 123
pectral	150	4 ± 4	525 ± 610	94	3 ± 117	282 ± 11038
ward	150	6 ± 5	900 ± 700	94	3 ± 1	282 ± 100
Temporal clustering						
kmeans	3 ± 5	26	78 ± 119	3 ± 0	463	1389 ± 197
spectral	3 ± 6	26	78 ± 158	86 ± 36	463	39818 ± 16621
ward	3 ± 9	26	78 ± 247	3 ± 0	463	1389 ± 222

the capacity of exhaustive approaches to generate homogeneous biclusters. Additionally, ISA had a great improvement in performance, indicating that it could be an interesting choice for dealing with this type of data. XMotifs, while being also a greedy algorithm had a worse performance but still better than the FABIA and Spectral algorithms (that are based on distribution parameter identification).

Size remarks

In this section, we studied the size of the biclusters to get insights of possible reasons some algorithms achieve better results than others. Since the algorithms generate biclusters with different properties, now we focus on bicluster size parameters. Our results are shown in Table 5.

A first view on the results show biclustering solutions tend to generate biclusters with highly variable sizes. This makes sense, since the search strategies of each algorithm are different.

In the previous sections we observed the capacity of exhaustive algorithms to generate better biclusters than the remaining algorithms. This results show some possible hits of the reasons why they do it, since CCC and BicPAM generate the smallest biclusters in terms of area. For BicPAM these results show a lack of subspace exploration. BicPAM achieves great results by allowing to generate very small biclusters. Additionally, while the algorithm promised to generate biclusters of all sizes (due to its exhaustive nature), this comes at a price of execution speed (which we observed to be far greater than the remaining algorithms) and a running memory price, making this promise to be potentially unfeasible. Despite being also exhaustive, the temporal contiguity constraint of CCC lead to a faster exploration of the datasets.

Table 6 Correlation between the virtual error and the three specific coherence measures: Variance (constant biclusters), MSR (shifting biclusters) and SMSR (scaling biclusters). Most of the algorithms agree that the expected patterns are of shifting nature

Algorithms	Artificial Data			Real Data			Type of Pattern
	Variance	MSR	SMSR	Variance	MSR	SMSR	
BicPAM	0.009	**0.133**	0.000	0.040	**0.306**	0.000	Shifting
Bimax	0.003	0.037	**0.087**	**0.196**	0.001	0.000	Constant/scaling
CCC	**0.112**	0.007	0.000	**0.200**	0.033	0.000	Constant
FABIA	0.051	**0.627**	0.061	0.003	**0.763**	0.001	Shifting
ISA	0.050	**0.453**	0.000	0.040	**0.045**	0.000	Shifting
Spectral biclustering	0.009	**0.125**	0.000	0.103	**0.361**	0.000	Shifting
XMotifs	0.115	**0.378**	0.016	0.003	**0.670**	0.000	Shifting

FABIA produces the largest biclusters. This is related to the search strategy that uses factor analysis as base. For Spectral, the use of singular value decomposition and the generated checkerboard structures help explain the size of the generated biclusters. Bimax generates biclusters with different sizes due to its simplistic approach. The greedy algorithms ISA and XMotif generate relatively small biclusters.

Finally, clustering approaches generate the largest solutions. This is strongly related with their restriction of including all rows (or columns) in the clustering.

Types of bicluster patterns

To detect the pattern structures found by the biclustering algorithms, we calculated the square of the Pearson coefficient between the VAR, MSR and SMSR and the Virtual Error. A high correlation could be indicative of the expected type of pattern. Our results are shown in Table 6.

Most biclusters agree that the expected patterns are of shifting nature. Bimax is one of the exceptions, supporting constant and scaling patterns. However, it must be pointed that Bimax does not recognize any special type of structure since it works only with binary data. Other exception is CCC, supporting the hypothesis of constant patterns. This could be a consequence of the temporal contiguity constrain that generates biclusters that are fundamentally different than the ones obtained by the other algorithms.

Discussion

Biclustering is a technique that allows the simultaneous clustering of rows and columns. It is worth noticing that the application of biclustering has not progressed in parallel with algorithm design. This has two particular reasons. The first one happens due to a gap between tool development and the understanding of the data properties for each specific study. The second one is a knowledge gap for applying biclustering with other analytical tools such as annotation processes, visualization programs and statistical methods, to derive a more comprehensive interpretation [9].

In the context of neurosciences, this gap is wider for two reasons: first, previous comparative studies consider only the gene expression context in their evaluations. Second, there is a scarcity of a consistent research line for the interpretation and application of biclustering.

Fig. 9 Heatmap of the illustrative data. Interaction between brain regions is local: some brain regions interact together in some time points. Traditional clustering analysis is not able to automatically discover these structures
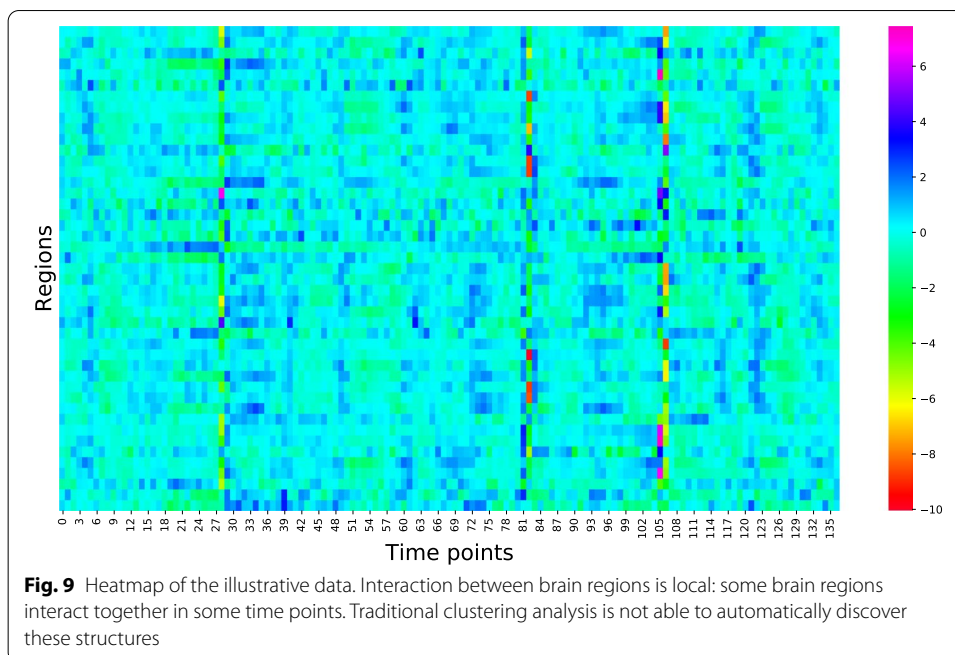
Our study aimed to close this gap between algorithm design, software development and the application of biclustering. The first part of this study was a comparative study for the biclustering capacities of extracting patterns from fMRI time-series data. In this section, we discuss the second issue and highlight the potential of biclustering for fMRI data analysis.

The **first** sub-section is an illustrative analysis of a single fMRI scan, highlighting the spatial and temporal patterns that biclustering can discover in fMRI. The **second** sub-section discusses how biclustering could be integrated with other state-of-the-art techniques of fMRI data analysis.

Illustrative results

To highlight the potentialities of biclustering to detect interesting structures, we consider the "Illustrative data" (already mentioned in methods). This dataset consists of an fMRI scan of 137 time points and 45 brain regions (the Harvard-Oxford atlas was used to group the brain in regions). A heatmap of our data is illustrated in Fig. 9, where several bicluster-like structures are visible.

We choose the CCC biclustering algorithm, implemented in BiGGEsTS [40, 61]. This algorithm was chosen, not only because it achieved good results in the previous analysis, but to generate temporal contiguous biclusters with an easier interpretation. The algorithm returned 749 biclusters, which we sorted according to a temporal statistical significance metric [40, 45]. Figures 10, 11 and 12 show three of the most relevant biclusters.

Biclustering opportunities in fMRI data analysis

Biclustering is not a standalone technique, and is often integrated with an analysis pipeline consisting of several other tools, such as results annotation processes, visualization programs and statistical methods [9]. We explore how biclustering could
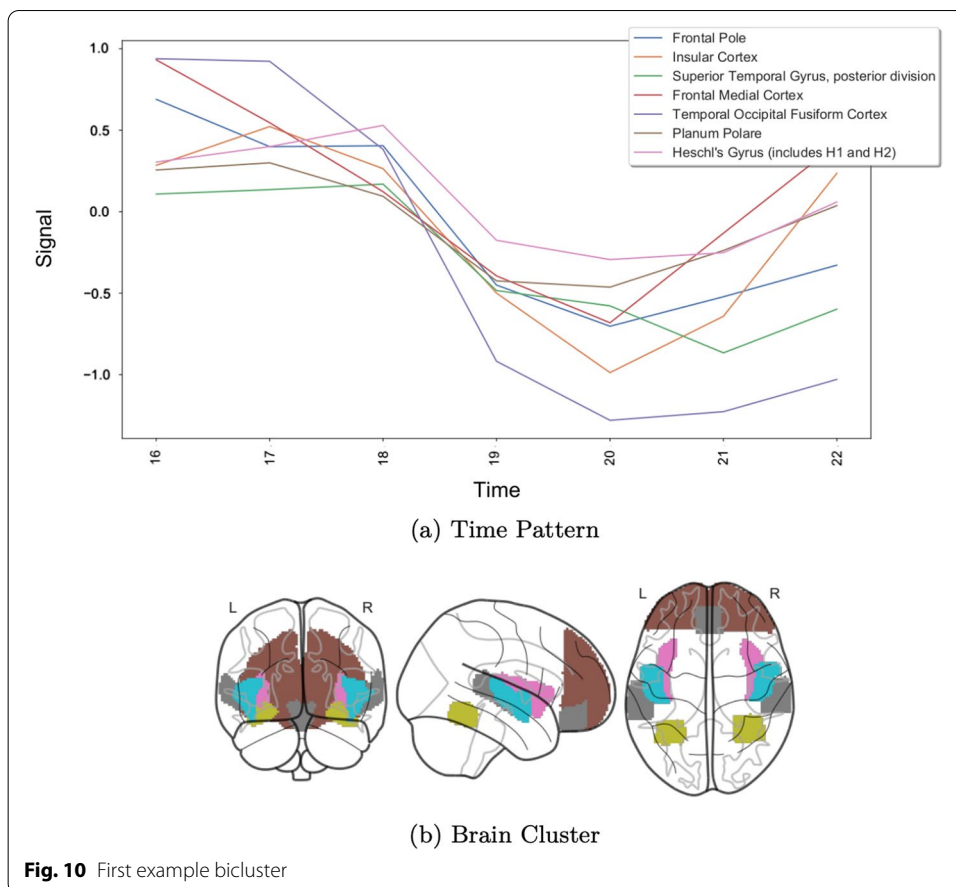
Fig. 10 First example bicluster

be integrated with other popular fMRI analysis techniques: clustering, independent component analysis and graph theory. In each section, we analyse how biclustering compares to them, and how these techniques can be combined to achieve greater interpretations of the phenomena under study. In addition to these, we explore how biclustering, an unsupervised technique of analysis is used to improve the quality of classification tasks.

Clustering

Biclustering algorithms were originally developed to expand on clustering limitations, in particular to allow overlapping between structures and searching for similarity considering only a subset of the features. Therefore, while clustering allows only to discover disjoint structures (subsets of rows or columns), biclustering discovers a larger set of possible interactions. Figure 13 illustrates differences between clustering and biclustering.

If directly applied to fMRI time series (in the format **Region × Time**), clustering discovers subsets of regions that have a similar behaviour considering *all* datasets. This approach assumes that all brain regions behave statically in time. Biclustering is expected to overcome this assumption and search for local patterns.

Fig. 11 Second example bicluster
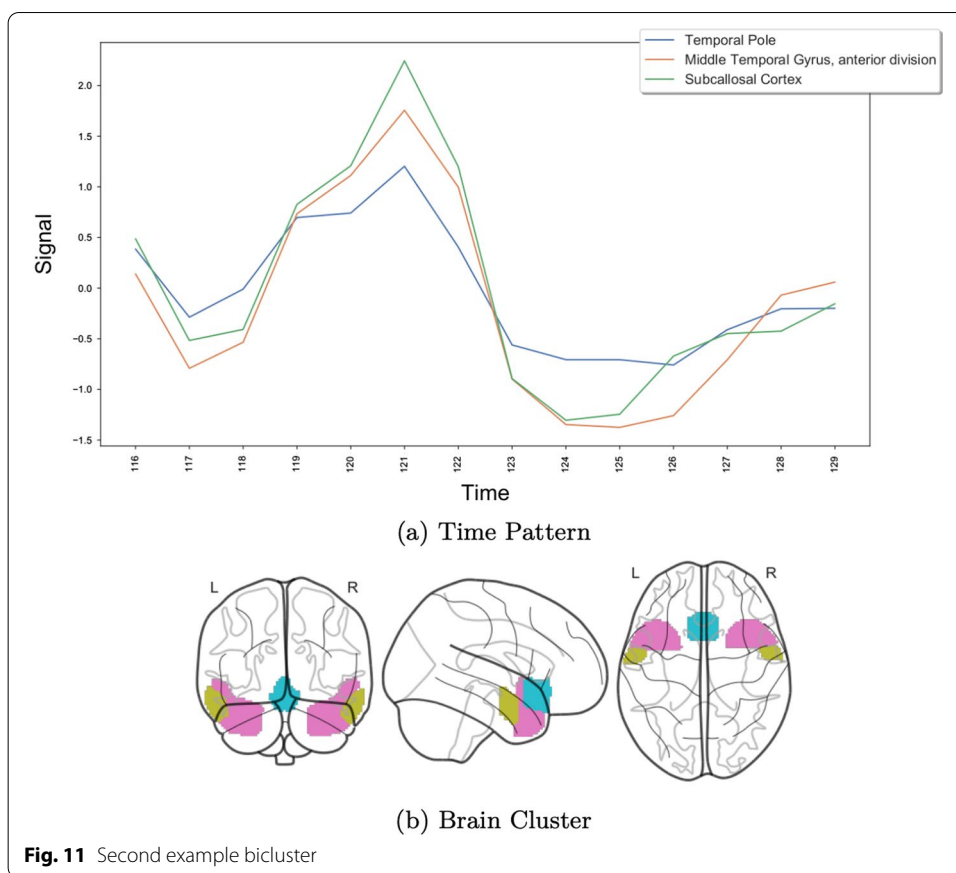
Independent component analysis

Independent Component Analysis (ICA) is a dimensionality reduction technique that separates a multivariate signal into a number of components [72]. Often used for exploratory data analysis, in fMRI is particular popular since it separates the analysis of a **Time × Region** dataset into their temporal (**Time × components**) and spatial (**components × Space**) parts [73]. ICA is used in fMRI to analyse static brain connectivity and has been used to consistent networks across several studies [74–77].

ICA and biclustering are fundamentally different techniques. While ICA transforms a $n \times p$ matrix into matrices $n \times m$ and $m \times p$, where m is the number of components (arbitrary chosen). Biclustering operates on a $n \times p$ matrix and extracts an arbitrary number of sub-matrices with an arbitrary size (Fig. 14 illustrates the differences between biclustering and ICA).

While ICA and biclustering are fundamentally different algorithms, decomposition techniques are used as an intermediate step in the internall bicluster algorithm [35, 37, 38, 78, 79]. An additional approach follows Gupta et al. [29], who proposed a new biclustering approach in the **Subject × Voxel** dimensions to analyse structural MRI data, using an intermediate ICA step.

An interesting application of ICA is its capacity to connect to other analysis techniques such as clustering in an analysis pipeline. ICA is used as a preprocessing step, and a clustering approach operates in the results of ICA [80–82]. The same approach can be

Castanho *et al. BMC Bioinformatics* (2022) 23:192

Page 22 of 30
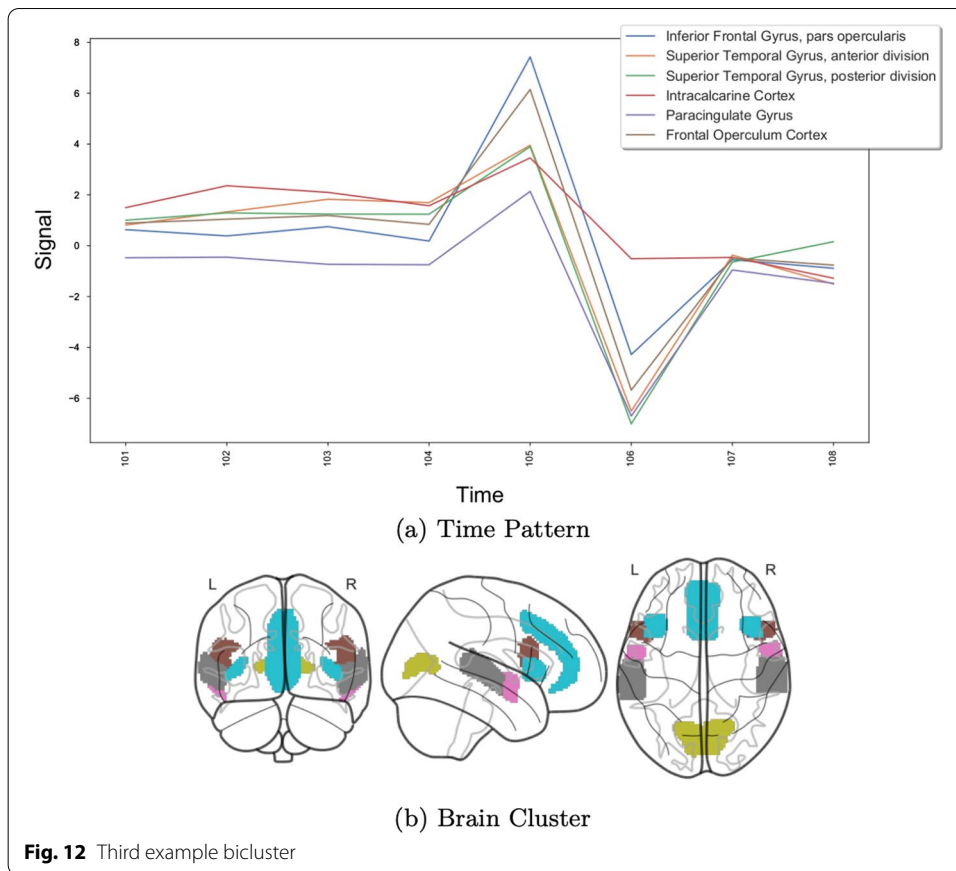
Fig. 12 Third example bicluster

Fig. 13 Diferentes between Clustering and Biclustering. While clustering methods allow to obtain only disjoint strips in the data matrix, biclustering finds more flexible structures (Figure adapted from [6])
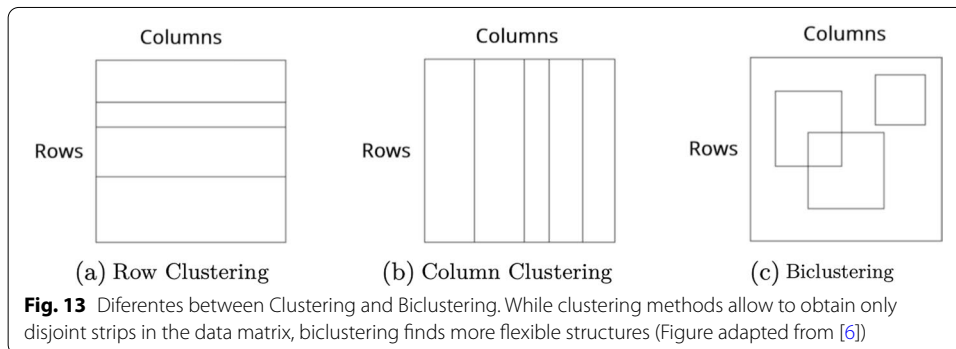
done with biclustering: instead of applying biclustering directly on fMRI time series (the approach of this study), biclustering algorithm can operate on the ICA results [14].

Graph theory

Graph theory is a mathematical field, applied in neurosciences to characterize the network structure, and models the brain as a set of vertices (which represent either ROIs or even single voxels) and the connections between them as edges. Graph techniques can be used either to analyse individual vertices or the graph as a whole [3]. Heuvel et al. [83] and Rubinov et al. [84] discuss the use of graph theory to analyze the brain network, in

Fig. 14 Differences between ICA and Biclustering. While ICA decomposes the original matrix, biclustering generates an arbitrary number of sub-matrices (depending on the algorithm)

Fig. 15 Relation between biclustering and graph theory: a biclustering can be seen as a submodule in a network
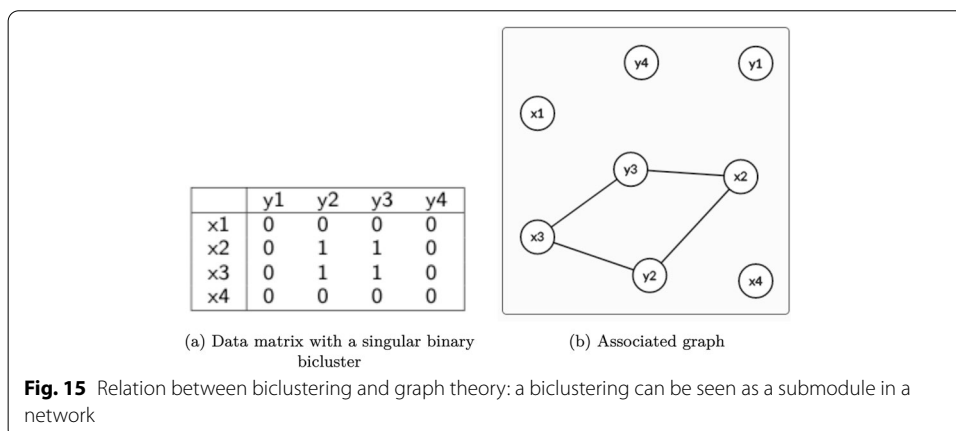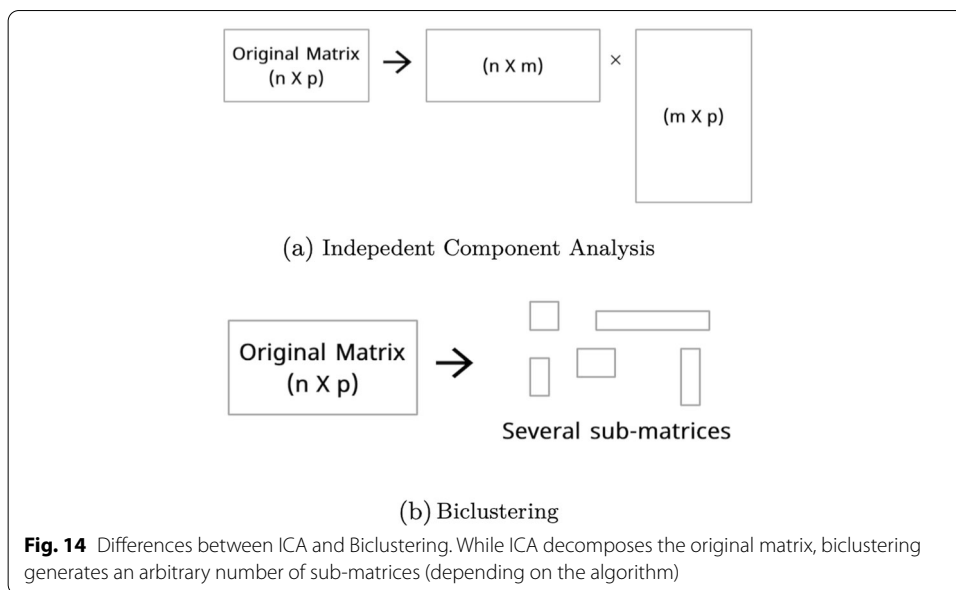
particular the use of fMRI time series to construct the brain network. This approach to analyse fMRI data has been used in the past, not only to cluster the brain [85, 86], but also in tasks such as detecting schizophrenia [87].

Madeira and Oliveira [7], and Henriques et al. [33] established a theoretical connection between biclustering and graph theory. A data matrix can be seen as a weighted bipartite graph, and biclustering of a data matrix is conceptually equivalent to discovering of maximal cliques or other structures from graphs obtained from binary or real-valued matrices. This connection motivates the use of biclustering to find maximal cliques or other structures from graphs obtained from binary or real-valued matrices, where edge values identify connection strength [88, 89]. Several biclustering algorithms use graph concepts internally to obtain biclusters [90–94]. Figure 15 illustrates the conection between biclustering and graph theory.

Biclustering has been extensively used in gene expression data analysis to uncover graph structures such as the discovery of co-expressed gene modules and regulatory

Patient	B1	B2	B3	Class
1	1	0	1	1
2	1	0	0	1
3	0	1	0	0
4	0	1	1	0

Fig. 16 After discovering the biclusters for a group of subjects, a data matrix can be obtained locating the presence of some bicluster in a subject, and then used for classification tasks. The biclusters (sets of features and corresponding representative values) are used as features (Figure adapted from [112])

networks. A review on the application scenarios of biclustering in the discovery of biological networks is provided by Xie et al. [9].

Classification tasks

A group of studies has been exploring fMRI techniques to study neurologic and psychiatric brain disorders, suggesting that neurodegenerative diseases target cortical networks rather than single regions [95]. These approaches have been applied in several diseases such as Alzheimer's [96, 97], depression [98], dementia [99], multiple sclerosis [100], amyotrophic lateral sclerosis [101] and schizophrenia [87, 102–105]. In general, the goal of biomedical research is to establish clinical biomarkers. These biomarkers are a set of characteristics (features) that allows early disease detection and prognostic prediction [106–110].

Biclustering has been used for supervised learning tasks [111–113]. Since a bicluster is representative of a certain characteristic of a subject, the presence of a bicluster can be used as a biomarker to discriminate characteristics of the population. Figure 16 illustrates this approach.

Bicluster based classification have been used previously for both temporal and non-temporal data. Considering temporal data, Carreiro et al. [111] considers biclusters of *Genes × Time* to identify biological processes related to the progression of Multiple Sclerosis patients. This approach was extended, considering not biclusters, but meta-biclusters (clusters of biclusters) [113]. In these approaches, the biclustering algorithm *CCC* was used. Matos et al. [112] used the *BicPAM* biclustering algorithm to analyze non-temporal data, together with the concept of meta-biclusters to characterize amyotrophic lateral sclerosis patients. Recently, Henriques and Madeira [114] showed that the use of biclustering based classification improves the performance of state-of-the-art classifiers.

Conclusions

Our study evaluated the use of biclustering in the context of fMRI data. Seven state-of-art biclustering algorithms were selected, compared among each other and with three traditional clustering algorithms. Our results show that most biclustering methods are not able to clearly surpass the traditional clustering when using the virtual error metric. However, the exhaustive methods (BicPAM and CCC) are able to achieve the best values of coherency of all tested algorithms. Additionally, and independently of the measured homogeneity, we point out that biclustering can be useful, and an improvement in comparison to clustering, due to its ability to consider both

spatial and temporal dimensions to discover the groups. We also observed that the bicluster structures found in this type of data are of a shifting nature.

Focusing on individual performance, BicPAM achieved high homogeneity levels by generating many small biclusters, which is not consistent with the promise of an exhaustive search. The issue is related to the performance of the algorithm, since it needs a prohibitive quantity of time and resources to discover bigger biclusters. This could be solved considering two approaches: the first one is related to the algorithm that can be optimized to generate results faster; the second is related to the parameterization of the algorithm that could be used if the size of the desirable biclusters is known apriori. It must be noticed that these observations are not inconsistent with the obtained results: even if this is considered, BicPAM will still be able to achieve good levels of homogeneity in its best biclusters.

Another interesting result comes from the greedy biclustering solutions, ISA and XMotifs. While not being able to achieve results as impressive as the exhaustive ones, they are still able to find some good ones using a fraction of the time BicPAM needs to operate. Additionally, while ISA works by using directly the real data, XMotifs requires a discretization step, which allows different approaches.

Considering the trade-off between the number of generated biclusters, their quality and execution time, Bimax is an interesting choice since it finds a huge number of biclusters very fast. This is achieved by doing a binarization of the data. While not being able to detect the best biclusters, its velocity could simply mean that the algorithm could be used to get some insights on the expected biclustering structures before running other algorithms.

FABIA and Spectral Biclustering are not able to achieve interesting results. Despite being both based on the same search type (Distribution Parameter identification) we believe that the reasons for this are different. The search strategy of FABIA means that the number of solutions will be limited to the number of columns of the dataset, which will strongly limit its capacity to generate solutions with many biclusters. Spectral Biclustering on the other hand is limited due to the generated checkerboard bicluster structure.

The last algorithm considered is CCC which uses notions of temporal contiguity during the exhaustive search for biclusters. This means that CCC is able to find multidimensional time series motifs. This leads to the generation of a high number of biclusters with an easy interpretation which could be possibly ideal for these types of analyses. Since time is contiguous in these biclusters, a new set of quality measures based on statistical significance can further be used for an improvement when filtering results. Furthermore, the high coherency levels observed in this study motivate the use of specific temporal biclustering methods to study fMRI data.

Biclustering is a tool to search for local patterns in data well established and recognised in the gene expression application context. Our study shows that biclustering is equally promising in fMRI data. While comparative studies provide guidance over the selection of the methods, the choice of the biclustering algorithm to analyse must be guided by the study objectives.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References

1. Atluri G, MacDonald A III, Lim KO, Kumar V. The brain-network paradigm: using functional imaging data to study how the brain works. Computer. 2016;49(10):65–71. https://doi.org/10.1109/mc.2016.307.
2. Lindquist MA. The statistical analysis of FMRI data. Stat Sci. 2008;23(4):439–64. https://doi.org/10.1214/09-sts282.
3. Margulies DS, Böttger J, Long X, Lv Y, Kelly C, Schäfer A, Goldhahn D, Abbushi A, Milham MP, Lohmann Gea. Resting developments: a review of FMRI post-processing methodologies for spontaneous brain activity. Magn Reson Mater Phys Biol Med. 2010;23(5–6):289–307. https://doi.org/10.1007/s10334-010-0228-5.
4. Atluri G, Karpatne A, Kumar V. Spatio-temporal data mining: a survey of problems and methods. ACM Comput Surv. 2018;51(4):83–18341. https://doi.org/10.1145/3161602.
5. Pourahmadi M, Noorbaloochi S. Multivariate time series analysis of neuroscience data: some challenges and opportunities. Curr Opin Neurobiol. 2016;37:12–5. https://doi.org/10.1016/j.conb.2015.12.006.
6. Tanay A, Sharan R, Shamir R. Biclustering algorithms: a survey. In: Aluru S, editor. Handbook of computational molecular biology. Computer and information science series. Boca Raton: Chapman & Hall/CRC; 2005.
7. Madeira SC, Oliveira AL. Biclustering algorithms for biological data analysis: a survey. IEEE/ACM Trans Comput Biol Bioinform. 2004;1(1):24–45. https://doi.org/10.1109/TCBB.2004.2.
8. Cheng Y, Church GM. Biclustering of expression data. In: Proceedings of the eighth international conference on intelligent systems for molecular biology. AAAI Press; 2000. p. 93–103. http://dl.acm.org/citation.cfm?id=645635.660833.
9. Xie J, Ma A, Fennell A, Ma Q, Zhao J. It is time to apply biclustering: a comprehensive review of biclustering applications in biological and biomedical data. Brief Bioinform. 2018;20:1450–65. https://doi.org/10.1093/bib/bby014.
10. Busygin S, Prokopyev O, Pardalos PM. Biclustering in data mining. Comput Oper Res. 2008;35(9):2964–87. https://doi.org/10.1016/j.cor.2007.01.005.
11. Padilha VA, Campello RJGB. A systematic comparative evaluation of biclustering techniques. BMC Bioinform. 2017;18(1):1–25. https://doi.org/10.1186/s12859-017-1487-1.
12. Busygin S, Boyko N, Pardalos PM, Bewernitz M, Ghacibeh G, Seref O, Kundakcioglu OE, Pardalos P. Biclustering EEG data from epileptic patients treated with vagus nerve stimulation. AIP Conf Proc. 2007. https://doi.org/10.1063/1.2817345.
13. Fan N, Boyko N, Pardalos PM. Recent advances of data biclustering with application in computational neuroscience. In: Chaovalitwongse W, Pardalos PM, Xanthopoulos P, editors. Computational neuroscience. New York: Springer; 2010. p. 85–112. https://doi.org/10.1007/978-0-387-88630-5_6.
14. Rahaman MA, Mathalon D, Lee HJ, Jiang W, Mueller BA, Andreassen O, Agartz I, Sponheim SR, Mayer AR, Stephen Jea. N-BIC: a method for multi-component and symptom biclustering of structural MRI data: application to schizophrenia. IEEE Trans Biomed Eng. 2020;67(1):110–21. https://doi.org/10.1109/tbme.2019.2908815.

15. Yu Q, Chen J, Du Y, Sui J, Damaraju E, Turner JA, van Erp TGM, Macciardi F, Belger A, Ford JMea. A method for building a genome–connectome bipartite graph model. J Neurosci Methods. 2019;320:64–71. https://doi.org/10.1016/j.jneumeth.2019.03.011.

16. Amar D, Yekutieli D, Maron-Katz A, Hendler T, Shamir R. A hierarchical Bayesian model for flexible module discovery in three-way time-series data. Bioinformatics. 2015;31(12):17–26. https://doi.org/10.1093/bioinformatics/btv228.

17. Wu X, Cheng C, Qiao C, Song C. Spatio-temporal differentiation of spring phenology in china driven by temperatures and photoperiod from 1979 to 2018. Sci China Earth Sci. 2020;63(10):1485–98. https://doi.org/10.1007/s11430-019-9577-5.

18. Shen S, Cheng C, Song C, Yang J, Yang S, Su K, Yuan L, Chen X. Spatial distribution patterns of global natural disasters based on biclustering. Nat Hazards. 2018;92(3):1809–20. https://doi.org/10.1007/s11069-018-3279-y.

19. Kaban PA, Kurniawan R, Caraka RE, Pardamean B, Yuniarto B. Sukim: biclustering method to capture the spatial pattern and to identify the causes of social vulnerability in indonesia: a new recommendation for disaster mitigation policy. Procedia Comput Sci. 2019;157:31–7. https://doi.org/10.1016/j.procs.2019.08.138.

20. Galvani M, Torti A, Menafoglio A, Vantini S (2020) A novel spatio-temporal clustering technique to study the bike sharing system in lyon. In: EDBT/ICDT workshops.

21. Izenman AJ, Harris PW, Mennis J, Jupin J, Obradovic Z. Local spatial biclustering and prediction of urban juvenile delinquency and recidivism. Stat Anal Data Min. 2011;4(3):259–75. https://doi.org/10.1002/sam.10123.

22. Martínez-Álvarez F, Gutiérrez-Avilés D, Morales-Esteban A, Reyes J, Amaro-Mellado J-L, Rubio-Escudero C. A novel method for seismogenic zoning based on triclustering: application to the Iberian Peninsula. Entropy. 2015;17:5000–21. https://doi.org/10.3390/e17075000.

23. Guigourès R, Boullé M, Rossi F. Discovering patterns in time-varying graphs: a triclustering approach. Adv Data Anal Classif. 2018;12(3):509–36. https://doi.org/10.1007/s11634-015-0218-6.

24. Wu X, Zurita-Milla R, Verdiguier EI, Kraak M-J. Triclustering georeferenced time series for analyzing patterns of intra-annual variability in temperature. Ann Am Assoc Geogr. 2018;108(1):71–87. https://doi.org/10.1080/24694452.2017.1325725.

25. Wu X, Zurita-Milla R, Kraak M-J, Izquierdo-Verdiguier E. Clustering-based approaches to the exploration of spatio-temporal data. Int Arch Photogramm Remote Sens Spatial Inf Sci. 2017;XLII–2/W7:1387–91. https://doi.org/10.5194/isprs-archives-XLII-2-W7-1387-2017.

26. Melgar-García L, Godinho MT, Espada R, Gutiérrez-Avilés D, Brito IS, Martínez-Álvarez F, Troncoso A, Rubio-Escudero C. Discovering spatio-temporal patterns in precision agriculture based on triclustering. In: Herrero Á, Cambra C, Urda D, Sedano J, Quintián H, Corchado E, editors. 15th international conference on soft computing models in industrial and environmental applications (SOCO 2020). Cham: Springer; 2021. p. 226–36.

27. Dhamodharavadhani S, Rathipriya R. Biclustering analysis of countries using COVID-19 epidemiological data. In: García Márquez FP, Lev B, editors. Internet of things. International series in operations research and management science, vol. 305. Cham: Springer; 2021. p. 93–114. https://doi.org/10.1007/978-3-030-70478-0_6.

28. Neves F, Finamore AC, Madeira SC, Henriques R. Mining actionable patterns of road mobility from heterogeneous traffic data using biclustering. IEEE Trans Intell Transp Syst. 2021. https://doi.org/10.1109/TITS.2021.3057240.

29. Gupta CN, Castro E, Rachkonda S, van Erp TGM, Potkin S, Ford JM, Mathalon D, Lee HJ, Mueller BA, Greve DN, Andreassen OA, Agartz I, Mayer AR, Stephen J, Jung RE, Bustillo J, Calhoun VD, Turner JA. Biclustered independent component analysis for complex biomarker and subtype identification from structural magnetic resonance images in schizophrenia. Front Psychiatry. 2017;8:179. https://doi.org/10.3389/fpsyt.2017.00179.

30. Prelić A, Bleuler S, Zimmermann P, Wille A, Bühlmann P, Gruissem W, Hennig L, Thiele L, Zitzler E. A systematic comparison and evaluation of biclustering methods for gene expression data. Bioinformatics. 2006;22(9):1122–9. https://doi.org/10.1093/bioinformatics/btl060.

31. Bozdaǧ D, Kumar AS, Catalyurek UV. Comparative analysis of biclustering algorithms. In: Proceedings of the first ACM international conference on bioinformatics and computational biology. BCB '10. New York: Association for Computing Machinery; 2010. p. 265–274. https://doi.org/10.1145/1854776.1854814.

32. Eren K, Deveci M, Küçüktunç O, Çatalyürek ÜV. A comparative analysis of biclustering algorithms for gene expression data. Brief Bioinform. 2012;14(3):279–92. https://doi.org/10.1093/bib/bbs032.

33. Henriques R, Antunes C, Madeira SC. A structured view on pattern mining-based biclustering. Pattern Recogn. 2015;48(12):3941–58. https://doi.org/10.1016/j.patcog.2015.06.018.

34. Aguilar-Ruiz JS. Shifting and scaling patterns from gene expression data. Bioinformatics. 2005;21(20):3840–5. https://doi.org/10.1093/bioinformatics/bti641.

35. Bergmann S, Ihmels J, Barkai N. Iterative signature algorithm for the analysis of large-scale gene expression data. Phys Rev E. 2003;67(3): 031902. https://doi.org/10.1103/physreve.67.031902.

36. Murali T, Kasif S. Extracting conserved gene expression motifs from gene expression data. Pac Symp Biocomput. 2003;8:77–88. https://doi.org/10.1142/9789812776303_0008.

37. Kluger Y. Spectral biclustering of microarray data: coclustering genes and conditions. Genome Res. 2003;13(4):703–16. https://doi.org/10.1101/gr.648603.

38. Hochreiter S, Bodenhofer U, Heusel M, Mayr A, Mitterecker A, Kasim A, Khamiakova T, Van Sanden S, Lin D, Talloen Wea. Fabia: factor analysis for bicluster acquisition. Bioinformatics. 2010;26(12):1520–7. https://doi.org/10.1093/bioinformatics/btq227.

39. Henriques R, Madeira SC. Bicpam: pattern-based biclustering for biomedical data analysis. Algorithms Mol Biol. 2014;9(1):1–30. https://doi.org/10.1186/s13015-014-0027-z.

40. Madeira SC, Teixeira MC, Sa-Correia I, Oliveira AL. Identification of regulatory modules in time series gene expression data using a linear time biclustering algorithm. IEEE/ACM Trans Comput Biol Bioinform. 2010;7(1):153–65. https://doi.org/10.1109/TCBB.2008.34.

41. Horta D, Campello RJGB. Similarity measures for comparing biclusterings. IEEE/ACM Trans Comput Biol Bioinf. 2014;11(5):942–54. https://doi.org/10.1109/tcbb.2014.2325016.

42. Santamaría R, Quintales L, Therón R. Methods to bicluster validation and comparison in microarray data. In: Yin H, Tino P, Corchado E, Byrne W, Yao X, editors. Intelligent data engineering and automated learning—IDEAL 2007. Berlin: Springer; 2007. p. 780–9.

43. Pontes B, Girldez R, Aguilar-Ruiz JS. Quality measures for gene expression biclusters. PLoS ONE. 2015;10(3):1–24. https://doi.org/10.1371/journal.pone.0115497.

44. Pontes B, Giráldez R, Aguilar-Ruiz JS. Biclustering on expression data: a review. J Biomed Inform. 2015;57:163–80. https://doi.org/10.1016/j.jbi.2015.06.028.

45. Henriques R, Madeira SC. BSig: evaluating the statistical significance of biclustering solutions. Data Min Knowl Discov. 2018;32(1):124–61. https://doi.org/10.1007/s10618-017-0521-2.

46. Henriques R, Madeira SC. Triclustering algorithms for three-dimensional data analysis: a comprehensive survey. ACM Comput Surv. 2018;51(5):95–19543. https://doi.org/10.1145/3195833.

47. Zhang Y, Zha H, Chu C-H (2005) A time-series biclustering algorithm for revealing co-regulated genes. In: International conference on information technology: coding and computing (ITCC'05), vol. II. Las Vegas: IEEE, p. 32–371. https://doi.org/10.1109/ITCC.2005.46. http://ieeexplore.ieee.org/document/1428433/. Accessed 2021-08-05.

48. Madeira SC, Oliveira AL. A polynomial time biclustering algorithm for finding approximate expression patterns in gene expression time series. Algorithms Mol Biol. 2009;4(1):8. https://doi.org/10.1186/1748-7188-4-8.

49. Madeira SC, Oliveira AL. Efficient biclustering algorithms for time series gene expression data analysis. In: Omatu S, Rocha MP, Bravo J, Fernández F, Corchado E, Bustillo A, Corchado JM, editors. Distributed computing, artificial intelligence, bioinformatics, soft computing, and ambient assisted living. Lecture notes in computer science, vol. 5518. Berlin: Springer; 2009. p. 1013–9. https://doi.org/10.1007/978-3-642-02481-8_154.

50. Erhardt EB, Allen EA, Wei Y, Eichele T, Calhoun VD. SimTB, a simulation toolbox for FMRI data under a model of spatiotemporal separability. Neuroimage. 2012;59(4):4160–7. https://doi.org/10.1016/j.neuroimage.2011.11.088.

51. Allen EA, Erhardt EB, Wei Y, Eichele T, Calhoun VD. Capturing inter-subject variability with group independent component analysis of FMRI data: a simulation study. Neuroimage. 2012;59(4):4141–59. https://doi.org/10.1016/j.neuroimage.2011.10.010.

52. Vaisvaser S, Lin T, Admon R, Podlipsky I, Greenman Y, Stern N, Fruchter E, Wald I, Pine DS, Rea Tarrasch. Neural traces of stress: cortisol related sustained enhancement of amygdala-hippocampal functional connectivity. Front Hum Neurosci. 2013;7:313. https://doi.org/10.3389/fnhum.2013.00313.

53. Kelly AMC, Uddin LQ, Biswal BB, Castellanos FX, Milham MP. Competition between functional brain networks mediates behavioral variability. Neuroimage. 2008;39(1):527–37. https://doi.org/10.1016/j.neuroimage.2007.08.008.

54. Mennes M, Kelly C, Zuo X-N, Di Martino A, Biswal BB, Castellanos FX, Milham MP. Inter-individual differences in resting-state functional connectivity predict task-induced bold activity. Neuroimage. 2010;50(4):1690–701. https://doi.org/10.1016/j.neuroimage.2010.01.002.

55. Mennes M, Zuo X-N, Kelly C, Di Martino A, Zang Y-F, Biswal B, Castellanos FX, Milham MP. Linking inter-individual differences in neural activation and behavior to intrinsic brain dynamics. Neuroimage. 2011;54(4):2950–9. https://doi.org/10.1016/j.neuroimage.2010.10.046.

56. Kiehl K, Laurens K, Duty T, Forster B, Liddle P. An event-related fMRI study of visual and auditory oddball tasks. J Psychophysiol. 2000;15:221. https://doi.org/10.1027//0269-8803.15.4.221.

57. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. NeuroImage. 2012;62(2):782–90. https://doi.org/10.1016/j.neuroimage.2011.09.015.

58. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DEea. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 2004;23:208–19. https://doi.org/10.1016/j.neuroimage.2004.07.051.

59. Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M, Smith SM. Bayesian analysis of neuroimaging data in FSL. Neuroimage. 2009;45(1):173–86. https://doi.org/10.1016/j.neuroimage.2008.10.055.

60. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage. 2006;31(3):968–80. https://doi.org/10.1016/j.neuroimage.2006.01.021.

61. Gonçalves JP, Madeira SC, Oliveira AL. Biggests: integrated environment for biclustering analysis of time series gene expression data. BMC Res Notes. 2009;2(124):1–11.

62. Henriques R, Ferreira FL, Madeira SC. Bicpams: software for biological data analysis with pattern-based biclustering. BMC Bioinform. 2017;18(1):1–16. https://doi.org/10.1186/s12859-017-1493-3.

63. Kaiser S, Leisch F. A toolbox for bicluster analysis in R. 2008. http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-3293-7.

64. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.

65. Goutte C, Toft P, Rostrup E, Nielsen FÜ, Hansen LK. On clustering FMRI time series. Neuroimage. 1999;9(3):298–310. https://doi.org/10.1006/nimg.1998.0391.

66. Thirion B, Varoquaux G, Dohmatob E, Poline J-B. Which fMRI clustering gives good brain parcellations? Front Neurosci. 2014;8:167. https://doi.org/10.3389/fnins.2014.00167.

67. Worsley KJ, Friston KJ. Analysis of fMRI time-series revisited-again. Neuroimage. 1995;2(3):173–81. https://doi.org/10.1006/nimg.1995.1023.

68. Hartigan JA. Direct clustering of a data matrix. J Am Stat Assoc. 1972;67(337):123–9. https://doi.org/10.1080/01621459.1972.10481214.

69. Mukhopadhyay A, Maulik U, Bandyopadhyay S. A novel coherence measure for discovering scaling biclusters from gene expression data. J Bioinform Comput Biol. 2009;07(05):853–68. https://doi.org/10.1142/s0219720009004370.

70. Divina F, Pontes B, Giráldez R, Aguilar-Ruiz JS. An effective measure for assessing the quality of biclusters. Comput Biol Med. 2012;42(2):245–56. https://doi.org/10.1016/j.compbiomed.2011.11.015.

71. Padilha VA, de Carvalho ACPLF (2018) A study of biclustering coherence measures for gene expression data. In: 2018 7th Brazilian conference on intelligent systems (BRACIS). https://doi.org/10.1109/bracis.2018.00100.
72. Hyvärinen A. Independent component analysis: recent advances. Philos Trans R Soc A Math Phys Eng Sci. 2013;371(1984):20110534. https://doi.org/10.1098/rsta.2011.0534.
73. Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imaging. 2004;23(2):137–52. https://doi.org/10.1109/TMI.2003.822821.
74. Beckmann CF, DeLuca M, Devlin JT, Smith SM. Investigations into resting-state connectivity using independent component analysis. Philos Trans R Soc B Biol Sci. 2005;360(1457):1001–13. https://doi.org/10.1098/rstb.2005.1634.
75. Damoiseaux JS, Rombouts SARB, Barkhof F, Scheltens P, Stam CJ, Smith SM, Beckmann CF. Consistent resting-state networks across healthy subjects. Proc Natl Acad Sci. 2006;103(37):13848–53. https://doi.org/10.1073/pnas.0601417103.
76. De Luca M, Beckmann CF, De Stefano N, Matthews PM, Smith SM. fMRI resting state networks define distinct modes of long-distance interactions in the human brain. Neuroimage. 2006;29(4):1359–67. https://doi.org/10.1016/j.neuroimage.2005.08.035.
77. Kiviniemi V, Starck T, Remes J, Long X, Nikkinen J, Haapea M, Veijola J, Moilanen I, Isohanni M, Zang Y-F, Tervonen O. Functional segmentation of the brain cortex using high model order group PICA. Hum Brain Mapp. 2009;30(12):3865–86. https://doi.org/10.1002/hbm.20813.
78. Dhillon IS. Co-clustering documents and words using bipartite spectral graph partitioning. In: Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining—KDD '01. San Francisco: ACM Press; 2001. p. 269–274. https://doi.org/10.1145/502512.502550. http://portal.acm.org/citation.cfm?doid=502512.502550. Accessed 2021-08-06.
79. Carmona-Saez P, Pascual-Marqui RD, Tirado F, Carazo JM, Pascual-Montano A. Biclustering of gene expression data by non-smooth non-negative matrix factorization. BMC Bioinform. 2006;7(1):78. https://doi.org/10.1186/1471-2105-7-78.
80. Sai Ma, Correa NM, Xi-Lin Li, Eichele T, Calhoun VD, Adali T. Automatic identification of functional clusters in fMRI data using spatial dependence. IEEE Trans Biomed Eng. 2011;58(12):3406–17. https://doi.org/10.1109/TBME.2011.2167149.
81. Esposito F, Scarabino T, Hyvarinen A, Himberg J, Formisano E, Comani S, Tedeschi G, Goebel R, Seifritz E, Di Salle F. Independent component analysis of fMRI group studies by self-organizing clustering. Neuroimage. 2005;25(1):193–205. https://doi.org/10.1016/j.neuroimage.2004.10.042.
82. Gultepe E, Makrehchi M. Improving clustering performance using independent component analysis and unsupervised feature learning. HCIS. 2018;8(1):25. https://doi.org/10.1186/s13673-018-0148-3.
83. van den Heuvel MP, Hulshoff Pol HE. Exploring the brain network: a review on resting-state fMRI functional connectivity. Eur Neuropsychopharmacol. 2010;20(8):519–34. https://doi.org/10.1016/j.euroneuro.2010.03.008.
84. Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. Neuroimage. 2010;52(3):1059–69. https://doi.org/10.1016/j.neuroimage.2009.10.003.
85. Shen X, Papademetris X, Constable RT. Graph-theory based parcellation of functional subunits in the brain from resting-state fMRI data. Neuroimage. 2010;50(3):1027–35. https://doi.org/10.1016/j.neuroimage.2009.12.119.
86. Mumford JA, Horvath S, Oldham MC, Langfelder P, Geschwind DH, Poldrack RA. Detecting network modules in fMRI time series: a weighted network analysis approach. Neuroimage. 2010;52(4):1465–76. https://doi.org/10.1016/j.neuroimage.2010.05.047.
87. Atluri G, Steinbach M, Lim KO, Kumar V, MacDonald A. Connectivity cluster analysis for discovering discriminative subnetworks in schizophrenia. Hum Brain Mapp. 2014;36(2):756–67. https://doi.org/10.1002/hbm.22662.
88. Charrad M, Ben Ahmed M. Simultaneous clustering: a survey. In: Kuznetsov SO, Mandal DP, Kundu MK, Pal SK, editors. Pattern recognition and machine intelligence. Berlin: Springer; 2011. p. 370–5.
89. Makino K, Uno T. New algorithms for enumerating all maximal cliques. In: Hagerup T, Katajainen J, editors. Algorithm theory—SWAT 2004. Berlin: Springer; 2004. p. 260–72.
90. Tanay A, Sharan R, Shamir R. Discovering statistically significant biclusters in gene expression data. Bioinformatics. 2002;18(Suppl 1):136–44. https://doi.org/10.1093/bioinformatics/18.suppl_1.S136.
91. Zhao L, Zaki MJ. Microcluster: efficient deterministic biclustering of microarray data. IEEE Intell Syst. 2005;20(6):40–9 (**cited By 34**).
92. Roy S, Bhattacharyya DK, Kalita JK. CoBi: pattern based co-regulated biclustering of gene expression data. Pattern Recogn Lett. 2013;34(14):1669–78. https://doi.org/10.1016/j.patrec.2013.03.018.
93. Li G, Ma Q, Tang H, Paterson AH, Xu Y. QUBIC: a qualitative biclustering algorithm for analyses of gene expression data. Nucleic Acids Res. 2009;37(15):101–101. https://doi.org/10.1093/nar/gkp491.
94. Xie J, Ma A, Zhang Y, Liu B, Cao S, Wang C, Xu J, Zhang C, Ma Q. QUBIC2: a novel and robust biclustering algorithm for analyses and interpretation of large-scale RNA-Seq data. Bioinformatics. 2020;36(4):1143–9. https://doi.org/10.1093/bioinformatics/btz692.
95. Seeley WW, Crawford RK, Zhou J, Miller BL, Greicius MD. Neurodegenerative diseases target large-scale human brain networks. Neuron. 2009;62(1):42–52. https://doi.org/10.1016/j.neuron.2009.03.024.
96. Greicius MD, Srivastava G, Reiss AL, Menon V. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. Proc Natl Acad Sci. 2004;101(13):4637–42. https://doi.org/10.1073/pnas.0308627101.
97. Rombouts SARB, Barkhof F, Goekoop R, Stam CJ, Scheltens P. Altered resting state networks in mild cognitive impairment and mild Alzheimer's disease: an fMRI study. Hum Brain Mapp. 2005;26(4):231–9. https://doi.org/10.1002/hbm.20160.
98. Greicius MD, Flores BH, Menon V, Glover GH, Solvason HB, Kenna H, Reiss AL, Schatzberg AF. Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. Biol Psychiat. 2007;62(5):429–37. https://doi.org/10.1016/j.biopsych.2006.09.020.

99. Rombouts SARB, Damoiseaux JS, Goekoop R, Barkhof F, Scheltens P, Smith SM, Beckmann CF. Model-free group analysis shows altered bold fMRI networks in dementia. Hum Brain Mapp. 2009;30(1):256–66. https://doi.org/10.1002/hbm.20505.

100. Lowe MJ, Beall EB, Sakaie KE, Koenig KA, Stone L, Marrie RA, Phillips MD. Resting state sensorimotor functional connectivity in multiple sclerosis inversely correlates with transcallosal motor pathway transverse diffusivity. Hum Brain Mapp. 2008;29(7):818–27. https://doi.org/10.1002/hbm.20576.

101. Mohammadi B, Kollewe K, Samii A, Krampfl K, Dengler R, Münte TF. Changes of resting state brain networks in amyotrophic lateral sclerosis. Exp Neurol. 2009;217(1):147–53. https://doi.org/10.1016/j.expneurol.2009.01.025.

102. Liu Y, Liang M, Zhou Y, He Y, Hao Y, Song M, Yu C, Liu H, Liu Z, Jiang T. Disrupted small-world networks in schizophrenia. Brain. 2008;131(4):945–61. https://doi.org/10.1093/brain/awn018.

103. Whitfield-Gabrieli S, Thermenos HW, Milanovic S, Tsuang MT, Faraone SV, McCarley RW, Shenton ME, Green AI, Nieto-Castanon A, LaViolette Pea. Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. Proc Natl Acad Sci. 2009;106(4):1279–84. https://doi.org/10.1073/pnas.0809141106.

104. Bluhm RL, Miller J, Lanius RA, Osuch EA, Boksman K, Neufeld R, Thééberge J, Schaefer B, Williamson P. Spontaneous low-frequency fluctuations in the BOLD signal in schizophrenic patients: anomalies in the default network. Schizophr Bull. 2007;33(4):1004–12. https://doi.org/10.1093/schbul/sbm052.

105. Garrity AG, Pearlson GD, McKiernan K, Lloyd D, Kiehl KA, Calhoun VD. Aberrant "default mode" functional connectivity in schizophrenia. Am J Psychiatry. 2007;164(3):450–7. https://doi.org/10.1176/ajp.2007.164.3.450.

106. Márquez F, Yassa MA. Neuroimaging biomarkers for Alzheimer's disease. Mol Neurodegener. 2019;14(1):1–14. https://doi.org/10.1186/s13024-019-0325-5.

107. Weickert CS, Weickert TW, Pillai A, Buckley PF. Biomarkers in schizophrenia: a brief conceptual consideration. Dis Markers. 2013;35:3–9. https://doi.org/10.1155/2013/510402.

108. Atluri G, Padmanabhan K, Fang G, Steinbach M, Petrella JR, Lim K, MacDonald A, Samatova NF, Doraiswamy PM, Kumar V. Complex biomarker discovery in neuroimaging data: finding a needle in a haystack. NeuroImage Clin. 2013;3:123–31. https://doi.org/10.1016/j.nicl.2013.07.004.

109. Chen C-M, Yang P, Wu M-T, Chuang T-C, Huang T-Y. Deriving and validating biomarkers associated with autism spectrum disorders from a large-scale resting-state database. Sci Rep. 2019;9(1):1–10. https://doi.org/10.1038/s41598-019-45465-9.

110. Abraham A, Milham MP, Di Martino A, Craddock RC, Samaras D, Thirion B, Varoquaux G. Deriving reproducible biomarkers from multi-site resting-state data: an autism-based example. Neuroimage. 2017;147:736–45. https://doi.org/10.1016/j.neuroimage.2016.10.045.

111. Carreiro AV, Anunciação O, Carriço JA, Madeira SC. Prognostic prediction through biclustering-based classification of clinical gene expression time series. J Integr Bioinform. 2011;8(3):73–89. https://doi.org/10.1515/jib-2011-175.

112. Matos J, Pires S, Aidos H, Gromicho M, Pinto S, de Carvalho M, Madeira SC. Unravelling disease presentation patterns in ALS using biclustering for discriminative meta-features discovery. In: Rojas I, Valenzuela O, Rojas F, Herrera LJ, Ortuño F, editors. Bioinformatics and biomedical engineering. Cham: Springer; 2020. p. 517–28.

113. Carreiro AV, Ferreira AJ, Figueiredo MAT, Madeira SC. Towards a classification approach using meta-biclustering: impact of discretization in the analysis of expression time series. J Integr Bioinform. 2012;9(3):105–20. https://doi.org/10.1515/jib-2012-207.

114. Henriques R, Madeira SC. Flebic: learning classifiers from high-dimensional biomedical data using discriminative biclusters with non-constant patterns. Pattern Recogn. 2021;115: 107900. https://doi.org/10.1016/j.patcog.2021.107900.

Publisher's Note