

# An Autism Spectrum Disorder Adaptive Identification Based on the Elimination of Brain Connections: A Proof of Long-Range Underconnectivity

Fatima zahra Benabdallah (✉ [f.zahra.benabdallah@gmail.com](mailto:f.zahra.benabdallah@gmail.com))

Universite Mohammed V de Rabat <https://orcid.org/0000-0003-2067-8321>

Ahmed Drissi El Maliani

Universite Mohammed V de Rabat

Dounia Lotfi

Universite Mohammed V de Rabat

Rachid Jennane

Universite d'Orleans

Mohammed El hassouni

Universite Mohammed V de Rabat

---

## Research Article

**Keywords:** Autism spectrum disorder, long-range connectivity, under-connectivity, ABIDE, Rs-fMRI

**Posted Date:** October 20th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-982098/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Soft Computing on March 18th, 2022. See the published version at <https://doi.org/10.1007/s00500-022-06890-7>.

# An Autism Spectrum Disorder adaptive identification based on the Elimination of brain connections: a proof of long-range underconnectivity

Fatima Zahra Benabdallah\* · Ahmed  
Drissi El Maliani · Dounia Lotfi · Rachid  
Jennane · Mohammed El Hassouni

Received: date / Accepted: date

**Abstract** Autism spectrum disorder (ASD) is theoretically characterized by alterations in functional connectivity between brain regions. Many works presented approaches to determine informative patterns that help to predict autism from typical development. However, most of the proposed pipelines are not specifically designed for the autism problem, i.e they do not corroborate with autism theories about functional connectivity. In this paper, we propose a framework that takes into account the properties of local connectivity and long range under-connectivity in the autistic brain. The originality of the proposed approach is to adopt elimination as a technique in order to well emerge the autistic brain connectivity alterations, and show how they contribute to differentiate ASD from controls. Experimental results conducted on the large multi-site Autism Brain Imaging Data Exchange (ABIDE) show that our approach provides accurate prediction up to 70% and succeeds to prove the existence of deficits in the long-range connectivity in the ASD subjects brains.

---

F. Benabdallah \*Corresponding author  
LRIT - CNRST URAC 29, Rabat IT Center, FS, Mohammed V University in Rabat, Morocco  
E-mail: Fatimazahra.benabdallah@um5.ac.ma

A. Drissi El Maliani  
LRIT - CNRST URAC 29, Rabat IT Center, FS, Mohammed V University in Rabat, Morocco

D. Lotfi  
LRIT - CNRST URAC 29, Rabat IT Center, FS, Mohammed V University in Rabat, Morocco

R. Jennane  
I3MTO, EA 4708, Orleans University, BP 6744, F-45072 Orleans, France

M. El Hassouni  
FLSH, Mohammed V University in Rabat, Rabat, Morocco

**Keywords** Autism spectrum disorder · long-range connectivity · under-connectivity · ABIDE · Rs-fMRI

## Declarations

Funding

Not applicable

Conflict of interest

The authors declare that they have no conflict of interest.

Availability of data and material

The resting state fMRI images data used to support the findings of this study are from the ABIDE I dataset that have been cited. The preprocessed version of the data is available at:

<http://preprocessed-connectomes-project.org/abide/Pipelines.html>.

Code availability

The custom code of this study is available on request from the corresponding author.

## 1 Introduction

Autism spectrum disorder (ASD) is one of the biggest challenges of modern medicine. From being considered a rare disorder affecting approximately 4 to 5 in every 10,000 children [1], it is now considered a common disorder with an increased occurrence that reached 1 in 54 children identified with ASD according to the estimates from the Centers for Disease Control and Prevention CDC's Autism and Developmental Disabilities Monitoring Network [2], with predominance in boys over girls. Its impact does not only touch the children, but also their parents, as high levels of distress were measured in parents with autistic children [3].

Autism symptoms and severity are heterogeneous and vary throughout the development. In general, ASD is characterized by impaired social interaction, language and communication abnormalities, and stereotypical behavior. Moreover, an investigation of Predictive Adaptive Behavior Skill [4] showed that autistic children also presented individual lower raw scores in community use, functional preacademics, home living, health and safety, leisure, self-care and

self-direction.

According to [5], two important processes in the Brain Development involve substantial loss of neural elements. The first is naturally occurring cell death, which involves a loss of more than 50% of the neurons; and the second is synaptic exuberance and pruning in which there is massive production of connections followed by the systematic elimination of up to 50% of those connections. However, [6] states that males with autism had an abnormal excess number of prefrontal neurons (a mean of 67% more) than those in the control group. This clearly indicates that the autistic brain is mainly characterized by some connectivity problems.

The brain is a complex network composed of regions that communicate through anatomical and functional connections with information processing that is near-optimal. These networks can be obtained through the anatomical and functional neuroimaging techniques MRI, DTI, EEG, MEG, fMRI, PET, SPECT [7], [8], [9]. Among these, the MRI have played a pivotal role in providing in vivo information in a noninvasive manner to identify aberration in brain development [10]. fMRI (functional MRI) goes farther by detecting changes in the MRI signal that arise when changes in neuronal activity occur with a change in brain state. This method remains the most used Neuroimaging Technique [11], thanks to its availability and safety, especially on children. In fMRI, the various neural connections are identified by measuring the temporal correlation between fluctuations in the Blood Oxygen Level Dependent (BOLD) signal of discrete anatomical regions [12]. The final results are 4-dimensional images that enable the visualization of brain structure and the study of the functional activities of the brain regions. fMRI can be conducted in two settings, resting-state and task-based. Even though experimental tasks fMRI, may be more suitable to achieve high classification accuracies. But for task-based fMRI data, it is difficult to acquire large datasets. Usually the sizes of the task-based datasets are composed of less than 50 subjects [11]. Moreover, it can be ill-suited for many individuals with ASD, particularly those who would be considered low-functioning or young children [13]. Furthermore, the resting-state fMRI requires neither a constrained experimental setup nor the active and focused participation of the subjects. It also has been proven to capture interactions between brain regions that may lead to neuropathology diagnostic biomarkers [14].

Many methods have been used to study the connectivity networks contained in resting state fMRI images, such as, the multi-echo independent component analysis (multi-echo ICA) [15], the Group-ICA [16] and the neighborhood one-class SVM [17]. Those later allow to come up with descriptors in order to classify subjects. The big issue of the above cited works is the small size of the data that prohibits the generalization of descriptors, as addressed by [18], [19]. This made the authors of [20], [21] and [22] use a large database, ABIDE, to overcome the handicap of reduced samples of previous autism research. In [20], 964 subjects fMRI images' BOLD signal were used to compute 7266 x 7266 connectivity matrices, by calculating the pair-wise correlation between each ROI. The 7266 ROIs were formed by clustering seed voxels into

Euclidean-close regions separated by at least 5 mm. A leave-one-out approach was then conducted reaching 60% of classification accuracy. Later in [21], the authors approach consisted of a specific intra and inter site classification using the cross-validation method. They built functional connectivity matrices using multiple atlases for the extraction of the ROIs and three methods for computing correlation. This approach showed good classification results achieving an accuracy of 67%. However, the pipeline did not consider autism connections specifically, since the connectomes were used with no intermediate processing before the classification. On the other hand, Bolte et al. [24] concluded that research needs to consider autistic traits to achieve better results in the autism identification. The results achieved in [22] support this point, since authors were able to observe differences in patterns between healthy and autistic children when local and long-distance deficits were considered.

Theoretically, ASD is characterized by an over-connectivity between local brain regions, and an under-connectivity between long-regions [23]. In this paper, we propose an approach that consists of eliminating brain connections in order to prove the existence of ASD theories. Our main idea is to extract connections that correspond to a given theory before eliminating them from the initial connectivity matrices. The resultant connectivity matrices are then used to classify the subjects. The impact of the elimination will indicate whether the studied theory exists or not, reposing on the fact that when removal of some specific connections impacts the classification performances, it means that there are descriptors of the disorder hidden in those connections.

In the present study, we focus on the long-range underconnectivity of autistic brains. For this, we use tools that are technically adaptive to this theory, namely, Hierarchical clustering (HC) [25] and Minimum spanning tree (MST) [26]. HC permits to extract local-range and long-range connections, while MST is used to identify the under-connectivity between the long-range regions.

It is worth mentioning that we already investigated the under-connectivity deficit in [27]. Here, we extend the investigation to long-range (inter-cluster) under-connectivity, which is a more specific property of the autistic brain.

To generalize our results, experiments are conducted on the ABIDE (Autism Brain Imaging Data Exchange) database. ABIDE was created to gather a large base of data to facilitate the autism research, and overcome the lack of generalization due to the small size of testing data. This database provides fMRI images of a big number of subjects from multiple international sites with no prior coordination and with different ranges of age, IQ and gender among other characteristics. It is to note that our approach permits to deal with the computational issues related to the large size of the ABIDE database, thanks to the elimination process which helps in alleviating the network.

Hence, in the present work we propose the elimination of brain connectivities as an approach to address three major concerns:

- Investigation of long-range underconnectivity theory related to ASD.
- Enhancement of the classification of ASD subjects.

- Generalization of the investigation using large datasets, and dealing with the computational demands of such datasets.

The remainder of this paper is organized as follows. Section 2 describes the data and the experimental framework used to classify the subjects and verify the autism theories. Section 3 presents the results of the classification after the application of every method from the approach. In section 4, we discuss the results achieved and the performance of the framework. Finally, a conclusion sums up and present what was achieved in this paper.

## 2 Materiel and methods

### 2.1 Data

ABIDE database is partitioned in two datasets, ABIDE I and ABIDE II. We decided to work with ABIDE I because a preprocessed version is proposed [28]. This later offers a base for comparison that diminishes the impact of the preprocessing pipeline between research approaches. Since according to [29] and [23] the final results of any analysis is highly correlated with the choice of the preprocessing pipline. In concordance with this, the used preprocessing pipeline is the C-PAC [30], which was also used in [21].

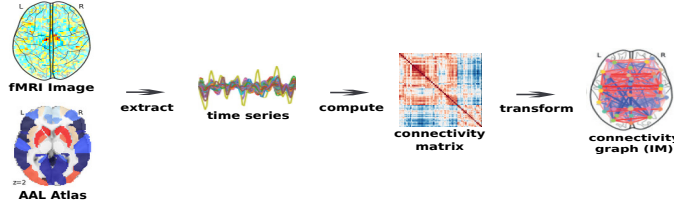
The ABIDE I dataset involved 17 international sites, sharing previously collected resting state functional magnetic resonance imaging (Rs-fMRI), anatomical and phenotypic datasets made available for data sharing with the broader scientific community. This dataset was the first available for sharing and yields 1112 subjects (539 individuals with ASD and 573 typical controls) with ages between 7 and 64 years. All the information included are anonymous. The imaging acquisition machines are from different brands, and most fMRI images in the ABIDE database are not treated or corrected in any way. Therefore, the Nilearn [31] project experts tested the data and eliminated all damaged images [21]. Hence, the new database contains 870 subject with 468 controls and 402 autistics.

### 2.2 FMRI images to connectivity graphs

The brain is a complex system with many interactions. These latter can be visualized in fMRI through the capture of changing blood oxygen level that are related to neurons' activities [33]. However, the brain is composed of billions of neurons, hence, clustering them into ROIs is an important step. From a neuroscience perspective, using ROIs instead of voxels has been proved more informative [32]. Moreover, having the same set of ROIs for all subjects allows easier inter-subject comparisons and reduces the individual anatomical variability that exists in human brains.

As can be seen in Figure.1, the Automated Anatomical Labeling (AAL) atlas [34] is used for brain parcellation. Time series of these parcels (AAL ROIs)

are extracted using an ordinary least square approach (OLS) from Nilearn (a library which patterns are described in [31]). Then, these latter are used to



**Fig. 1** Steps to construct the connectivity graphs.

pull out the correlation matrices between the ROIs of the brain. Here, the correlation matrices are computed using the tangent, as a method that led to the best classification results in [21]. Finally, the correlation matrices are transformed into graphs to visualize the connections between the ROIs of the atlas.

A graph  $g = (V, E)$  consists of vertices  $V$  that are connected by a set of weighted edges  $E \subseteq V \times V$ . In this work, the vertices are the atlas ROIs and the edges are the connections computed in the correlation matrices.

### 2.3 Minimum spanning tree

The spanning tree is an unbiased graph theory tool that permits to reduce the complexity of high-order network structure while preserving its core framework [35], [36]. Its potential as a powerful method has been brought to light in the analysis of schizophrenia[37]. Moreover, its application greatly improved the diagnostic accuracy for Alzheimer's disease [36].

A spanning tree  $T$  of a graph  $G$  is a subgraph that includes all vertices of  $G$ , and connects them without any cycles. To extract a specific spanning tree a condition is added. For example, to extract the minimum spanning tree the condition is to only extract the vertices with the weakest weights.

In this paper, this process is done through a greedy algorithm, the Kruskal algorithm [38]. The greedy choice is to pick the smallest weights while constructing the tree to result into the minimum spanning tree. For that, the algorithm sorts the edges based on their weights. Then, it builds the tree by adding increasing cost arcs at each step while keeping the total weight of all the edges to the minimum.

### 2.4 Hierarchical clustering

Hierarchical clustering (HC) builds a cluster hierarchy that is commonly displayed as a tree diagram called a dendrogram. In the context of brain connectivity, each ROI represents a separate cluster in a first time. Then step by

step, the distances of dissimilarity between every two clusters are computed and the two most similar clusters are joined into a single new cluster till the construction of the final cluster that contains all the ROIs.

The idea behind the clustering came from [39] where a Multi-Layer graph clustering method permitted to detect connectivity problems in the autistic brain. The hierarchical clustering was used to investigate the Genotype to phenotype relationships in autism spectrum disorders [40] using the strength of interactions in the phenotypic network as a metric. However, this latter does not guarantee spatially homogeneous clusters, a requirement that has been found to be advantageous [41]. In this work, clusters are constructed using the Euclidean distance since it provides enforced spatial contiguity and also because we are dealing with ROIs that are defined in a 3-dimensional space. The Euclidean distance between two points  $a = (a_1, a_2, \dots, a_n)$  and  $b = (b_1, b_2, \dots, b_n)$  of an Euclidean n-space is given by:

$$d(a, b) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (1)$$

In addition, to compute the dissimilarity of sets, we use the wards method since according to [42], it offers higher accuracy compared to other methods and minimizes the variance between the elements. In the wards method, the choice of the clusters to merge at each step is based on the optimal value of an objective function.

$$\begin{aligned} d(C_i \cup C_j, C_k) = & \frac{n_i + n_k}{n_i + n_j + n_k} d(C_i, C_k) \\ & + \frac{n_j + n_k}{n_i + n_j + n_k} d(C_j, C_k) \\ & - \frac{n_k}{n_i + n_j + n_k} d(C_i, C_j) \end{aligned} \quad (2)$$

Where,  $C_i$ ,  $C_j$  and  $C_k$  are disjoint clusters with sizes  $n_i$ ,  $n_j$  and  $n_k$  respectively.

## 2.5 ASD classification

An ASD classifier categorizes the input subjects into a defined number of groups. Here, the first group contains the neurotypical subjects. Its label is 0. The second group is for ASD subjects with label 1. The objective of this classification is to decide the class label of every subject using feature vectors extracted from the correlation matrices computed using the methods presented above. The classifier used is an l2-penalized support vector classifier SVC with a p-value of 0.01 and a 10 folds cross-validation. SVC is an implementation of SVM by scikit-learn library [43]. It returns a "best fit" hyperplane that divides the learning data into a fixed number of classes and then predicts the test data.

To measure the performance of this classification, three metrics are computed,



namely, the accuracy, the sensitivity and the specificity [44].

The accuracy shows the classifier ability to categorize ASD and TDC subjects correctly. It is defined by:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

The sensitivity computes the ability of the classifier in classifying the ASD subjects correctly. It is defined by:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

Finally, the specificity determines the accuracy of the TDC subjects classification. It is defined by:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (5)$$

with:

True positive (TP) refers to the number of cases correctly categorized in the ASD group.

False positive (FP) is the number of cases incorrectly classified as ASD.

True negative (TN) equals the number of cases correctly categorized in the TDC group.

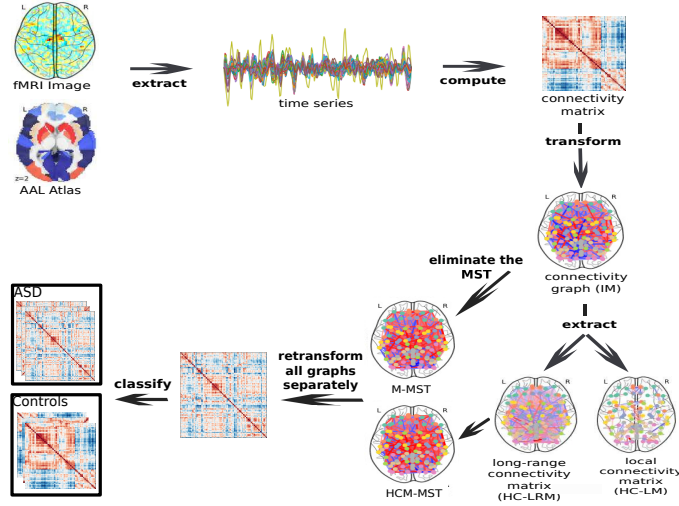
False negative (FN) is the number of cases incorrectly classified as TDC.

## 2.6 Proposed approach

As mentioned in the introduction, the originality of the proposed approach is that it is based on the elimination of brain connectivities. The elimination takes its root from the Mathematical law of double negation [45] where one can negate certain information in order to prove its veracity. Here, the elimination process suppresses connections extracted using the tools described in 2.3 and 2.4. It is also to note, that the approach is ASD specific since the used tools are technically similar to autism theories. It first eliminates weak connections to test the under-connectivity of the autistic brain. Then, it separates the brain connectivity map into long-range and local connectivity. Finally, weak connections are eliminated from the long-range connectivity matrices to test the long-range-under-connectivity. Figure.2 summarizes all the steps of the proposed approach.

### 2.6.1 Elimination of the weakest connections of the brain

The under-connectivity theory stipulates that some regions in the autistic brain are under-connected, or in other words, can not communicate efficiently. Therefore, to verify this theory, weak connections are first extracted from the initial connectivity graphs using the algorithm described in subsection 2.3. Then, the constructed tree is eliminated by suppressing all its edges from its original graph from which we extracted it. Finally, the obtained graphs weights are retransformed into new correlation matrices for the classification. Figure.3 presents all the important steps.

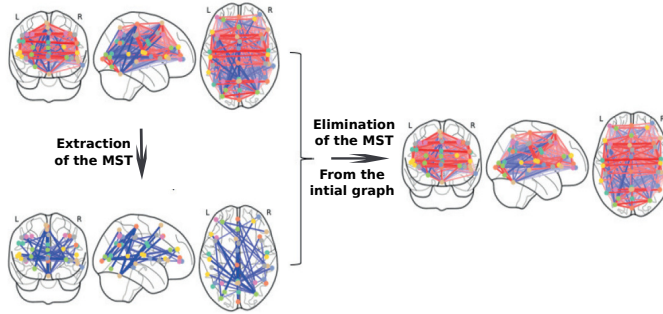


**Fig. 2** Summary of the proposed approach.

### 2.6.2 Long-range and local connectivity matrices

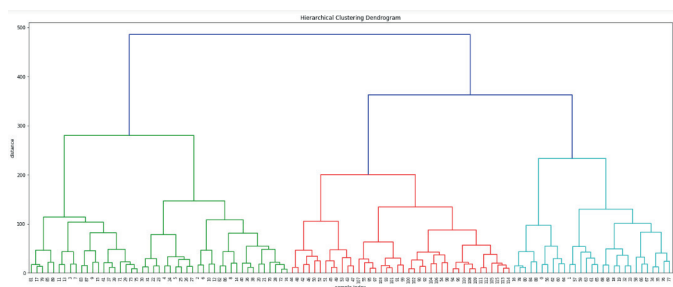
In this step, we firstly create clusters of the AAL atlas ROIs using Scipy (a Python-based ecosystem of open-source software for mathematics) [46]. This first allocates each ROI to a separate cluster, then merges them using the equations described in 2.4 till the creation of the final cluster that contains all the ROIs (Figure.4 represents the dendrogram created in this step). Then, a dissimilarity value is fixed to extract the clusters, before extracting the connections between and outside these clusters. By eliminating these extracted connections we construct the long-range and the local-range connectivity matrices that we use for classification.

Depending on the value of the dissimilarity measure, the number of ROIs in the clusters changes. This can be seen in Figure.4. Therefore, multiple dis-



**Fig. 3** Elimination of the minimum spanning tree. The blue connections' weights are inferior to 0 and the red's superior to 0. The strong degrees of color represent the extremes

similarity measures have been tested to bring out their impact on the results. Their values and results are reported below in the experimental results section.



**Fig. 4** Dendrogram of the AAL Atlas ROIs.

### 2.6.3 Long-range under-connectivity

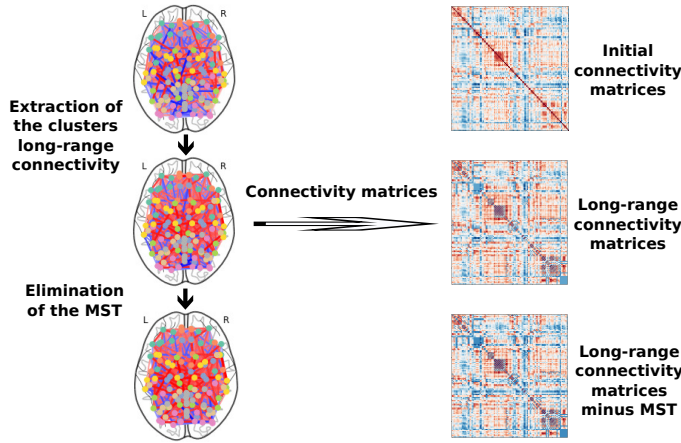
As mentioned in the introduction, the main goal of our study is to verify the under-connectivity between long regions, by eliminating the weak connections from the inter-cluster connectivity matrices. For this, we first extract the long-range connections from the initial graphs using the clusters constructed by the hierarchical clustering. Then, we eliminate the minimum spanning trees from the resultant graphs. Finally, we transform the newly composed graphs into matrices for the classification. All these steps are presented in Figure.5, where a decrease of the number of edges after the extraction of the clusters long-range connections is notable ((a) to (b)). The same decrease is also visible in the matrices (fewer connections in the long-range connectivity matrix than in the initial matrix thanks to the clustering). From (b) to (c), the elimination of the MST leads to a decrease in the number of the weakest long-range connections (blue edges) which can also be seen in the associated connectivity matrix where less blue points are represented compared to the other matrices.

## 3 Results and Discussion

### 3.1 Experimental results

In this work, the connectivity of autistic brains is studied in a highly heterogeneous setting and on large test datasets. Methods of our approach are first applied to the whole ABIDE I dataset. Then, to verify the impact of different criteria (age, gender, handedness) on the results, the same methods are also applied on different sub-datasets. Information on these later are stated in Table.1.

As mentioned before, the proposed approach starts by computing the cor-



**Fig. 5** Elimination of the long-range under-connectivity: extraction of the long-range connections and elimination of the weakest connection.

**Table 1** Sub-datasets of ABIDE used in the evaluation process

Sub-datasets	cardinal	Selection criteria
D1	870	All subjects
D2	726	Males subjects
D3	719	Right handed subjects
D4	556	Subjects with age between 9 and 18
D5	56	Subjects with age inferior to 9

relation matrices of the AAL atlas ROIs. Then, our approach's methods are applied to construct the new connectivity matrices as reported in section 2.6. Later, every group of matrices is classified separately. The performance measures of these classifications are stated in the tables below. Finally, to highlight the performance of our approach, we compare it with the most influential state-of-the-art approaches, namely Nielsen and al. [20] and Abraham and al. [21]. A brief description of those approaches and the proposed one is given as follows:

- Nielsen and al. [20]: This method uses a leave-one-subject-out cross validation for classification on the ABIDE subjects' correlation matrices.
- Abraham and al. [21]: This method uses a leave-one-site-out cross-validation on matrices of the Multi-subject Dictionary Learning (MSDL) atlas ROIs [47].
- IM (Initial Matrices)[27]: This method is comparable to the one used in [21], however, the ROIs used to compute the Initial Matrices are from the AAL atlas.
- M-MST (Minus Minimum Spanning Tree) [27]: In this method the MST is removed from the initial matrices (IM method). The resultant matrices are used for classification.

- HC-LRM (Hierarchical Clustering- Long Range connectivity Matrices) [proposed]: In this method, HC is applied to the AAL ROIs to construct clusters. Then long-range connections are extracted from the IM matrices to construct the Long-Range connectivity matrices for the classification.
- HC-LM (Hierarchical Clustering- Local connectivity Matrices) [proposed]: This method is comparable to HC-LRM, however, here it extracts clusters local connections. Then it constructs Local connectivity matrices for the classification.
- HCM-MST (Hierarchical Clustering Minus MST) [proposed]: Same to M-MST, in this method MST is removed from the HC-LRM matrices before classification.

The first method of the proposed approach (M-MST) eliminates the MST's extracted connections from the initial connectivity matrices. After this elimination, we remark a slight increase in the accuracy values of sub-datasets D1, D2, and D5 but in D3 and D4 the accuracy decreases. However, an important result shows up, which is the decrease of sensitivity in all datasets no matter the value of accuracy. This same result was also found in [27] with different criteria sub-datasets. Which means that the eliminated connections contain information about the autistic brain connectivity deficit. These results are contained in Table.2 that shows the results of the initial matrices classification compared to classification after the elimination of MST values.

In the second method (HC-LRM) we divide the brain into several clusters.

**Table 2** Classification results of the subjects before and after eliminating the minimum spanning tree. IM here refers to the initial matrices and M-MST the matrices after elimination of the MST

		Accuracy	Sensitivity	Specificity
D1	IM	68.13	61.90	73.42
	M-MST	68.78	60.69	75.60
D2	IM	66.33	62.82	69.70
	M-MST	67.48	62.45	72.30
D3	IM	66.21	53.86	75.80
	M-MST	65.93	50.71	77.89
D4	IM	67.31	61.58	72.62
	M-MST	66.76	60.03	72.83
D5	IM	53.63	75.42	37.00
	M-MST	67.62	65.42	65.50

The hierarchical clustering permits to create different groups of clusters depending on the dissimilarity measure value. This latter is chosen empirically by testing different values to extract different clusters of ROIs. Then using these clusters, we construct the long-range and local connectivity matrices of the first dataset D1 and classify them separately. The metrics of this classification are stored in Table.3, where we notice an increased accuracy of the inter-connectivity matrices classification using dissimilarity values of 80, 90 and 100 with a mean of 6-8 ROIs per cluster. However, after tests on the other

datasets, 80 remains the suitable tuned dissimilarity value.

Another point that jumps to attention in Table.3 is that the inter-connectivity

**Table 3** Accuracy results of inter- and intra-connectivity on D1 using different values of the dissimilarity distance. HC-LRM are the long-range connectivity matrices and HC-LM the local connectivity matrices

d	50	80	90	100	150	250	300
HC-LRM	68.90	69.39	69.34	69.61	67.51	66.94	64.86
HC-LM	61.09	60.96	61.43	62.01	63.09	63.04	65.70

(HC-LRM) leads to better accuracy results than the intra-connectivity (HC-LM). The highest intra-connectivity classification result in the whole dataset (D1) is 65.09%. It was achieved with a 300 dissimilarity distance value (3 clusters of ROIs). But this value is still lower than the accuracy achieved without clustering (68.13%). That is why we focus on the inter-connectivity in the remainder of this work.

The results of Table.4 stems from classifying inter-connectivity matrices of clusters constructed using a dissimilarity value of 80. They show an improvement of the three classification metrics (accuracy, sensitivity and specificity) of the inter-connectivity matrices when compared to the initial matrices in all datasets. In other words, the elimination of the clusters' local connections guarantees better results even when using different criterions datasets. This supports the brain's communication deficit in autism, and precisely the long-range deficit.

From the results presented in Table.2 and Table.4, we can conclude that both

**Table 4** Results of the inter-connectivity matrices classification compared to the initial matrices classification. IM are the initial matrices and HC-LRM the long-range matrices

		Accuracy	Sensitivity	Specificity
D1	IM	68.13	61.90	73.42
	HC-LRM	69.40	63.07	74.78
D2	IM	66.33	62.82	69.70
	HC-LRM	67.24	63.65	70.73
D3	IM	66.21	53.86	75.80
	HC-LRM	68.81	56.63	78.34
D4	IM	67.31	61.58	72.62
	HC-LRM	69.88	63.45	75.71
D5	IM	53.63	75.42	37.00
	HC-LRM	56.41	78.33	39.00

the elimination of the minimum spanning tree and the elimination of the HC clusters local connections had an impact on the classification. When combined (eliminating the local connections, then removing the MST from the resulting inter clusters), they lead to the same decrease in sensitivity as reported

**Table 5** Classification results of the inter-connectivity matrices before and after eliminating the spanning tree. Here HC-LRM are the long range connectivity matrices and HCM-MST the long range matrices after eliminating the MST

		Accuracy	Sensitivity	Specificity
D1	HC-LRM	69.40	63.07	74.78
	HCM-MST	69.76	61.06	77.11
D2	HC-LRM	67.24	63.65	70.73
	HCM-MST	67.80	62.71	72.64
D3	HC-LRM	68.81	56.63	78.34
	HCM-MST	68.42	54.94	78.98
D4	HC-LRM	69.88	63.45	75.71
	HCM-MST	69.88	60.46	78.45
D5	HC-LRM	56.41	78.33	39.00
	HCM-MST	61.28	64.17	59.50

**Table 6** Classification performance of the proposed approach compared to state-of-the-art methods applied to ABIDE I

	Accuracy	Sensitivity	Specificity
HCM-MST	69.76	61.06	77.11
HC-LRM	69.40	63.07	74.78
M-MST [27]	68.78	60.69	75.60
Abraham and al. [21]	66.8	61.0	72.3
Nielsen and al. [20]	60.0	62.0	58.0

in Table.2 after removing MST from the initial matrices. In other words, the elimination of the minimum spanning tree has a direct impact on the sensitivity (the ability of ASD prediction), which means that the weakest connections extracted using the MST contain information about autism. Therefore, after their elimination the sensitivity always drops. Table.5 summarizes those results. Comparing our approach's methods with the state-of-art approaches applied to ABIDE I, reported in Table.6, we can see an increase in the accuracy value, reaching 69.76% with a 3% raise compared to [21] where the best result was 66.80%. We also remark an increase in sensitivity in the long-range classification compared to the highest result achieved by [20]. Finally, the impact of eliminating the MST is highlighted when comparing the sensitivity values of all methods.

### 3.2 Discussion

In this work, we studied the autistic brain connectivity in a highly heterogeneous setting and on large test datasets from ABIDE I. The problem of using ABIDE is the uncontrolled heterogeneity due to the multi-site fMRI data acquisition that influences the results. The uncontrolled heterogeneity poses great challenges to develop brain-based classifiers for psychiatric illnesses in general. Without forgetting the sources of uncontrolled variation that can arise

across sites as the scanner type, pulse sequence and sample composition. But since our objective is to simplify the process of autism detection and since the heterogeneity is part of the real setting, we used the datasets with little changes. Only the images with quality issues were eliminated. That is why using sub-datasets with certain criteria helps to meld those challenges and might even give more weight to findings from the whole dataset.

Our objective was to verify the ASD theories and also to improve the detection process of autism. Our approach consisted of suppressing brain connections that are related to a given theory. For this, we used and combined specific tools to extract those connections. We taught about the minimum spanning tree because it permits to simplify a complex network into a simple one based on weights. Moreover, MST is concordant with the under-connectivity theory of the autistic brain, since it extracts the weakest connections.

Besides, we used hierarchical clustering to divide the brain regions into clusters. By doing so, we verified the intra- and inter-connectivity of the brain as in[39] where the use of clustering led to the discovery of deficits in ASD brain connectivity.

Moreover, since Autism affects children in the early years of childhood, we searched for the most common and simple imagery acquisition method that can extract the brain connectivity without causing any damage to the brain and has the advantage of availability and cheap cost compared to other imagery methods. Then, we tested multiple ASD specific methods and combined them to improve the accuracy of ASD detection. The accuracy obtained is 69.76% on the whole dataset D1, as can be seen in Table.5, which exceeds previously published ABIDE findings [21], [20].

The elimination of the minimum spanning tree led to a diminution in the sensitivity in all datasets. Since the sensitivity mirrors the classification of the autistic subjects, its decrease implies that the eliminated weak connections contain biomarkers relative to autism, and supports the autism under-connectivity theories. The result of this elimination was reported in Table.2.

On the other hand, the elimination of intra-connectivity (the clusters local connectivity) using the clusters constructed with the hierarchical clustering improved accuracy and sensitivity. This latter increased from 61.9% to 63.07% which indicates a long-range connectivity deficit in ASD. This can be seen in Table.4 that reports the elimination of intra-connectivity in all datasets. The importance of the choice of the dissimilarity distance used to construct the clusters was also highlighted by testing multiple values. A good choice can strongly improve the prediction accuracy and sensitivity, as seen in Table.3.

Later, when the minimum spanning tree was again eliminated from the long-range connectivity matrices, the above findings of Table.2 were confirmed in Table.5. Where, again, the elimination of the weak connections extracted from the inter-connectivity matrices showed a decrease in the sensitivity. Which gives more weight to the under-connectivity deficit in Autism, especially between the long-range regions. Hence, we conclude the existence of the long-range under-connectivity deficit in the autistic brain.

In this study, the results showed evidence of a deficit in the long-range con-



nectivity between the regions of the autistic brain that is independent of age, gender and handedness. This was confirmed by the classification results of the sub-datasets that showed the same pattern as the classification applied on the whole dataset. An accuracy of 70% proves that the proposed approach is very promising and leads to accurate prediction. As the elimination of inter-connectivity enhances the accuracy while the elimination of the MST impacts the sensitivity. Their combined use highlighted the long-range under-connectivity and permitted better results compared to the state-of-art, as can be seen in Table.6. However, there are still other theories that we could not verify as the over connectivity reported in [39], [48] and [49]. Such connectivity might be between specific ROIs of the brain and should be studied separately.

## 4 Conclusion

In this paper, we suggested a novel approach based on ASD adaptive methods to verify autism theories about brain connectivity. Our aim was to improve the accuracy of the detection of this disorder while keeping up with the advance in the autism research. We used ABIDE I a large dataset with data from different world sites. We applied the minimum spanning tree and the hierarchical clustering to extract the weak connections, and the local/long-range connections, respectively. A combination of these two methods, was used to construct connectivity matrices to verify the long-range underconnectivity. From the results, we were able to detect a long-range connectivity deficit and proved that it is independent of age, gender and other criteria as the handedness. Furthermore, an under-connectivity problem was highlighted by the pattern found in the specificity and the sensitivity values. This holds promise to future detection of decisive biomarkers of autism that will help in understanding and diagnosing this disorder.

## References

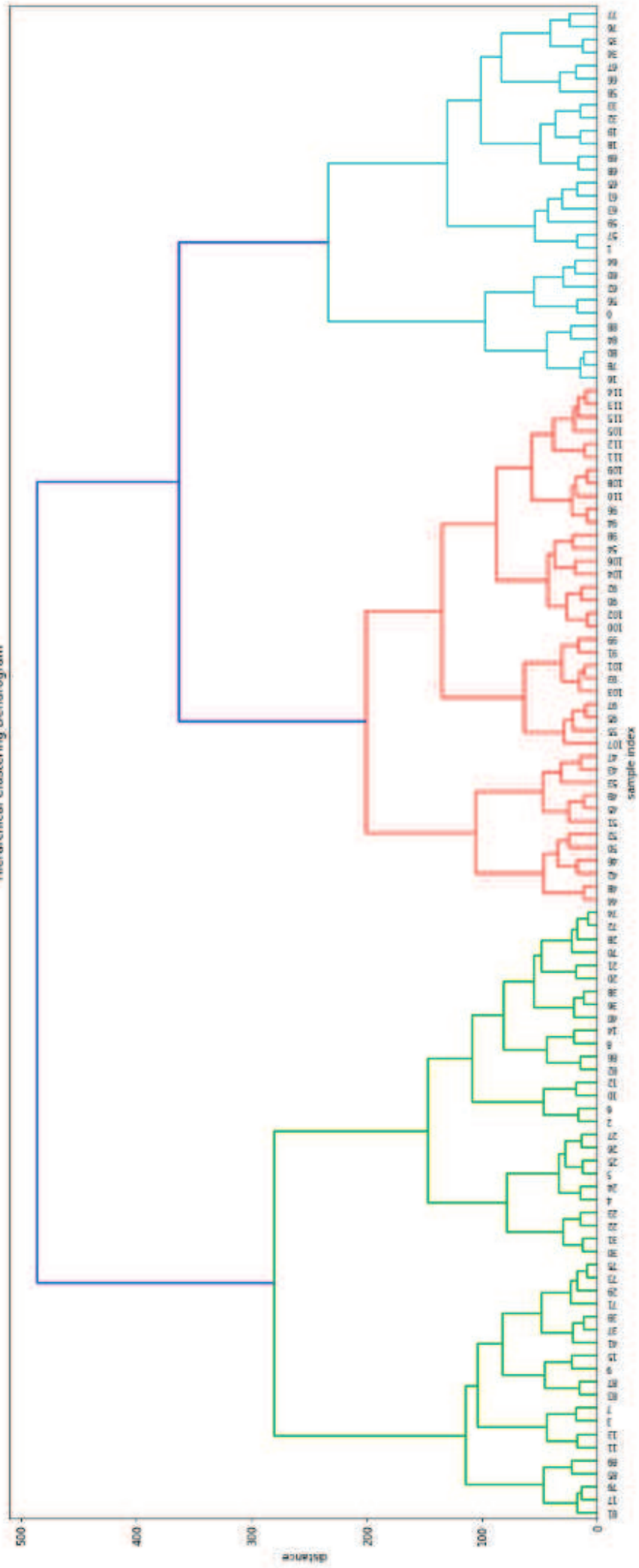
1. Harris JC, Journal of the American Academy of Child & Adolescent Psychiatry 55(8):729 – 735, 2016. see <http://www.sciencedirect.com/science/article/pii/S0890856716302805>
2. Maenner MJ, Shaw KA, Baio J, et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. MMWR Surveill Summ 2020;69(No. SS-4):1–12. DOI: <http://dx.doi.org/10.15585/mmwr.ss6904a1>
3. Salomone E, Settanni M, Ferrara F, Salandin A; CST Italy Team. The Interplay of Communication Skills, Emotional and Behavioural Problems and Parental Psychological Distress. J Autism Dev Disord. 2019 Nov;49(11):4365-4374. doi: 10.1007/s10803-019-04142-6.
4. Emma Feige, Rhonda Mattingly, Teresa Pitts, Alan F. Smith, "Autism Spectrum Disorder: Investigating Predictive Adaptive Behavior Skill Deficits in Young Children", Autism Research and Treatment, vol. 2021, Article ID 8870461, 9 pages, 2021. <https://doi.org/10.1155/2021/8870461>
5. Stiles J, Jernigan TL The basics of brain development. Neuropsychology Review 20(4):327–348, 2010. doi:10.1007/s11065-010-9148-4

6. Courchesne E, Mouton PR, Calhoun ME, Semendeferi K, Ahrens-Barbeau C, Hallet MJ, Barnes CC, Pierce KJ, Neuron number and size in prefrontal cortex of children with autism. *JAMA* 306 18:2001–10. 2011.
7. Camprodon J, Stern T, Selecting neuroimaging techniques: A review for the clinician. *The primary care companion to CNS disorders* 15, 2013. doi:10.4088/PCC.12f01490
8. Alessandro, G., Roma, S., and Remo, J. Editorial: advanced neuroimaging methods for studying autism disorder. *Front. Neurosci.* 11:533. 2017. doi: 10.3389/fnins.2017.00533
9. Du, Y., Fu, Z., and Calhoun, V. D. Classification and prediction of brain disorders using functional connectivity: promising but challenging. *Front. Neurosci.* 12:525. 2018. doi: 10.3389/fnins.2018.00525
10. Hiremath, C.S., Sagar, K.J.V., Yamini, B.K. et al. Emerging behavioral and neuroimaging biomarkers for early and accurate characterization of autism spectrum disorders: a systematic review. *Transl Psychiatry* 11, 42. 2021. <https://doi.org/10.1038/s41398-020-01178-6>
11. Liu Meijie, Li Baojuan, Hu Dewen, Autism Spectrum Disorder Studies Using fMRI Data and Machine Learning: A Review, *Frontiers in Neuroscience* 15, p 1111, 2021 DOI=10.3389/fnins.2021.697870
12. Nierhaus T, Margulies D, Long X, Villringer A, fMRI for the Assessment of Functional Connectivity. 2012. doi:10.5772/23864
13. Lau, W. K. W., Leung, M. K. Lau, B. W. M. Resting-state abnormalities in autism spectrum disorders: a meta-analysis. *Sci. Rep.* 9, 1–8. 2019.
14. Greicius M, Resting-state functional connectivity in neuropsychiatric disorders. *Current opinion in neurology* 21:424–30, 2008. doi:10.1097/WCO.0b013e328306f2c5
15. Kundu P, Voon V, Balchandani P, Lombardo M, Poser B, Bandettini P, Multi-echo fmri: A review of applications in fmri denoising and analysis of bold signals. *NeuroImage* 154, 2017. doi:10.1016/j.neuroimage.2017.03.033
16. Tang XY, Zeng WM, Wang N, Shi YH, Zhao L, A novel layered data reduction mechanism for clustering fmri data. *Biomedical Signal Processing and Control* 33, 2017. doi:10.1016/j.bspc.2016.11.014
17. Yang J, Zhong N, Liang P, Wang J, Yao Y, Lu S, Brain activation detection by neighborhood one-class svm. *Cognitive Systems Research* 11:16–24, 2007. doi:10.1016/j.cogsys.2008.08.001
18. Button K, Ioannidis J, Mokrysz C, Nosek B, Flint J, Robinson E, Munafò M, Power failure: Why small sample size undermines the reliability of neuroscience. *Nature reviews Neuroscience* 14, 2013. doi:10.1038/nrn3475
19. Mumford J, A power calculation guide for fmri studies. *Social cognitive and affective neuroscience* 7:738–42, 2012. doi:10.1093/scan/nss059
20. Nielsen J, Zielinski B, Fletcher P, Alexander A, Lange N, Bigler E, Lainhart J, Anderson J, Multisite functional connectivity mri classification of autism: Abide results. *Frontiers in human neuroscience* 7:599, 2013. doi:10.3389/fnhum.2013.00599
21. Alexandre Abraham, Michael P. Milham, Adriana Di Martino, R. Cameron Craddock, Dimitris Samaras, Bertrand Thirion, Gael Varoquaux, Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example, *NeuroImage*, Volume 147, 2017, Pages 736–745, ISSN 1053-8119, <https://doi.org/10.1016/j.neuroimage.2016.10.045>.
22. Chang Liu, Jie Xue, Xu Cheng, Weiwei Zhan, Xin Xiong, Bin Wang, "Tracking the Brain State Transition Process of Dynamic Function Connectivity Based on Resting State fMRI", *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 9027803, 12 pages, 2019. <https://doi.org/10.1155/2019/9027803>
23. Hull Jocelyn V., Dokovna Lisa B., Jacokes Zachary J., Torgerson Carinna M., Irimia Andrei, Van Horn John Darrell, Resting-State Functional Connectivity in Autism Spectrum Disorders: A Review, *Frontiers in Psychiatry* 7, p 205, 2017, DOI=10.3389/fpsy.2016.00205
24. S. Bolte, S. Girdler, and P. B. Marschik, "The contribution of environmental exposure to the etiology of autism spectrum disorder," *Cellular and Molecular Life Sciences: CMLS*, vol. 76, no. 7, pp. 1275–1297, 2019.
25. Zhaoyuan Fang, *The Methods and Tools for Clustering Analysis*, Editor(s): Olaf Wolkenhauer, *Systems Medicine*, Academic Press, 2021, Pages 9–13, ISBN 9780128160787, <https://doi.org/10.1016/B978-0-12-801238-3.11463-1>.

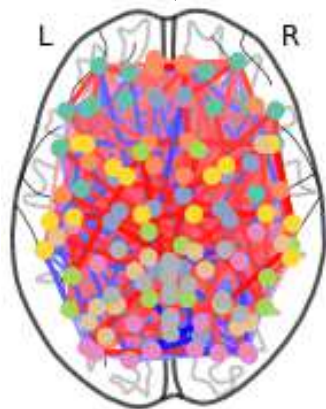
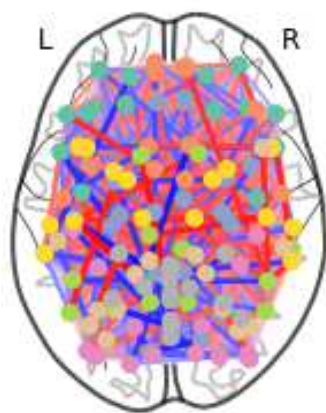
26. Daintith, John, and Edmund Wright. A Dictionary of Computing. : Oxford University Press, 2008. DOI: 10.1093/acref/9780199234004.001.0001
27. Benabdallah F, Drissi El Malinai A, Lotfi D, Jennane R, El Hassouni M, Analysis of under-connectivity in autism using the minimum spanning tree: application on large multi-site dataset. pp 296–299, 2018 doi:10.1109/ISIVC.2018.8709213
28. Craddock C, Benhajali Y, Carlton C, Francois C, Evans A, Jakab A, Khundrakpam B, Lewis J, Qingyang I, Michael M, Chaogan Y, Bellec P, The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. Frontiers in Neuroinformatics 7, 2013. doi:10.3389/conf.fninf.2013.09.00041
29. Strother S, Evaluating fmri preprocessing pipelines. IEEE engineering in medicine and biology magazine : the quarterly magazine of the Engineering in Medicine & Biology Society 25:27–41, 2006. doi:10.1109/MEMB.2006.1607667
30. Craddock C, Sharad S, Brian C, Ranjeet K, Satrajit G, Chaogan Y, Qingyang I, Daniel L, Vogelstein J, Burns R, Stanley C, Mennes M, Clare K, Adriana D, Castellanos F, Michael M, Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac). Frontiers in Neuroinformatics 7, 2013. doi:10.3389/conf.fninf.2013.09.00042
31. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, Gramfort A, Thirion B, Varoquaux G, Machine learning for neuroimaging with scikit-learn. Frontiers in neuroinformatics 8:14, 2014. doi:10.3389/fninf.2014.00014
32. Zafar R, Malik A, Kamel N, Dass S, Role of voxel selection and roi in fmri data analysis. pp 1–6, 2016. doi:10.1109/MeMeA.2016.7533739
33. W. Li and P. C. van Zijl, “Quantitative theory for the transverse relaxation time of blood water,” NMR in Biomedicine, vol. 33, no. 5, p. 427, 2020. <https://doi.org/10.1002/nbm.4207>
34. Tzourio-Mazoyer N, Landeau B, DF P, Crivello F, Etard O, Delcroix N, Mazoyer B, Marc J, Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. NeuroImage 15:273–89, 2002. doi:10.1006/nimg.2001.0978
35. Li X, Jing Z, hu B, Jing Z, Zhong N, Li M, Ding Z, Yang J, Zhang L, Feng L, Majoe D, A resting-state brain functional network study in mdd based on minimum spanning tree analysis and the hierarchical clustering. Complexity 2017:1–11, 2017. doi:10.1155/2017/9514369
36. Guo H, Liu L, Chen J, Xu Y, Jie X, Alzheimer classification using a minimum spanning tree of high-order functional network on fmri dataset. Frontiers in Neuroscience 11:639, 2017. doi:10.3389/fnins.2017.00639
37. Anjomshoa A, Dolatshahi M, Amirkhani F, Rahmani F, Mirbagheri M, Aarabi M, Structural brain network analysis in schizophrenia using minimum spanning tree. vol 2016, pp 4075–4078, 2016. doi:10.1109/EMBC.2016.7591622
38. Horowitz E, Sahni E, Sahni S, Fundamentals of computer algorithms. xiv, 626 p : ill Potomac, Maryland: Computer Science Press Inc, 1978 include bibliography and index. 2019.
39. Ghanbari Y, Bloy L, Tunc B, Shankar V, Roberts T, Edgar J, Schultz R, Verma R, On characterizing population commonalities and subject variations in brain networks. Medical Image Analysis 38, 2015. doi:10.1016/j.media.2015.10.009
40. Chang J, Gilman S, Chiang A, Sanders S, Vitkup D, Genotype to phenotype relationships in autism spectrum disorders. Nature neuroscience 18, 2014. doi:10.1038/nn.3907
41. Blumensath T, Jbabdi S, Glasser M, Essen D, Ugurbil K, Behrens T, Smith S, Spatially constrained hierarchical parcellation of the brain with resting-state fmri. NeuroImage 76, 2013. doi:10.1016/j.neuroimage.2013.03.024
42. Esztergár-Kiss D, Caesar B, Definition of user groups applying ward’s method. Transportation Research Procedia 22:25–34, 2017. doi:10.1016/j.trpro.2017.03.004
43. 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, Louppe G, Scikit-learn: Machine learning in python. Journal of Machine Learning Research 12, 2012.
44. Baratloo A, Hosseini M, Negida A, El Ashal G, Evidence-based emergency medicine; part 1: Simple definition and calculation of accuracy, sensitivity and specificity. Emergency (Tehran, Iran) 3:48–9, 2015.

45. Hazewinkel M, Encyclopaedia of Mathematics: Volume 10. Encyclopaedia of Mathematics, Springer Netherlands, 2013. En-see [https://books.google.co.ma/books?id=\\_YPtCAAAQBAJ](https://books.google.co.ma/books?id=_YPtCAAAQBAJ)
46. Oliphant T, Scipy: Open source scientific tools for python 9:10–20, 2007
47. Varoquaux G, Gramfort A, Pedregosa F, Michel V, Thirion B, Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. vol-22, pp 562–73, 2011
48. Supekar K, Uddin LQ, Khouzam A, Phillips J, Gaillard WD, Kenworthy LE, Yerys BE, Vaidya CJ, Menon V. Brain hyperconnectivity in children with autism and its links to social deficits. *Cell Rep.* 2013 Nov 14;5(3):738-47. doi: 10.1016/j.celrep.2013.10.001.
49. Keown CL, Shih P, Nair A, Peterson N, Mulvey ME, Müller RA. Local functional overconnectivity in posterior brain regions is associated with symptom severity in autism spectrum disorders. *Cell Rep.* 2013 Nov 14;5(3):567-72. doi: 10.1016/j.celrep.2013.10.003.

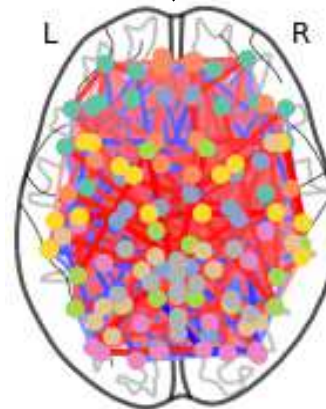
Hierarchical Clustering Dendrogram



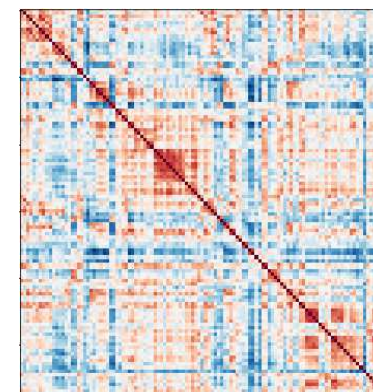
**Extraction of  
the clusters  
long-range  
connectivity**



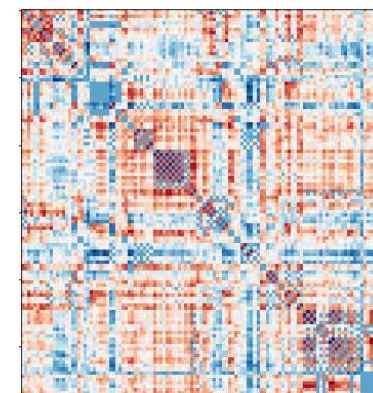
**Elimination  
of the MST**



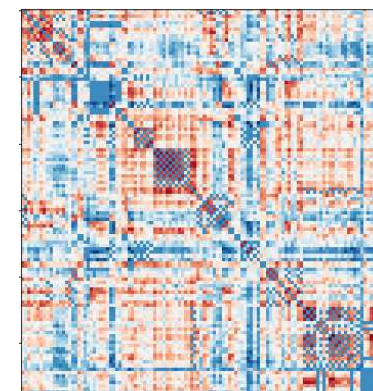
**Connectivity matrices**



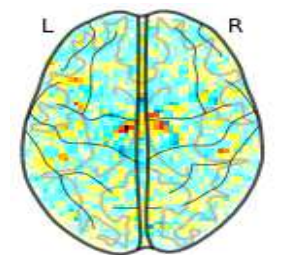
**Initial  
connectivity  
matrices**



**Long-range  
connectivity  
matrices**



**Long-range  
connectivity  
matrices  
minus MST**



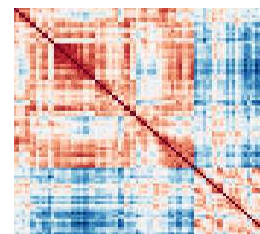
**fMRI Image**

extract



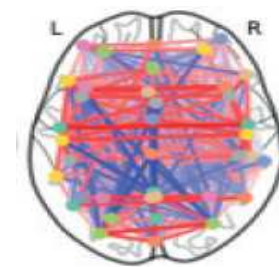
**time series**

compute

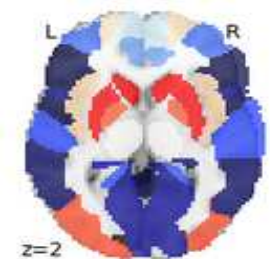


**connectivity  
matrix**

transform

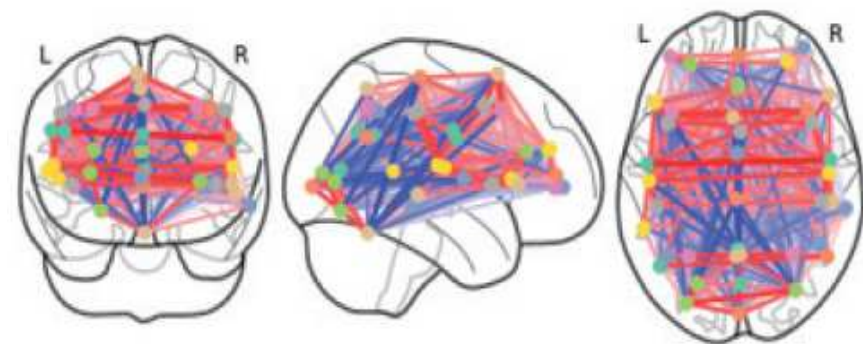


**connectivity  
graph (IM)**

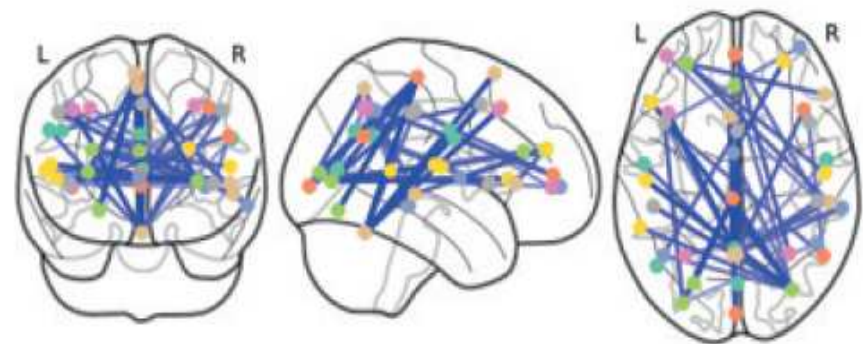


**AAL Atlas**





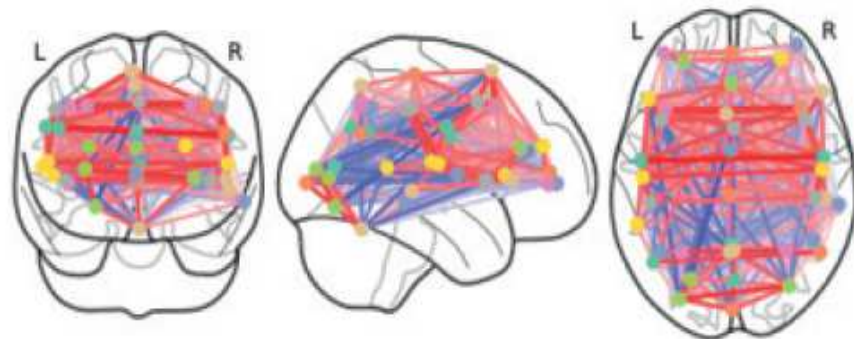
**Extraction  
of the MST**



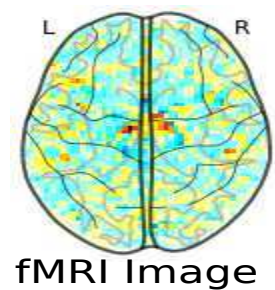
**Elimination  
of the MST**



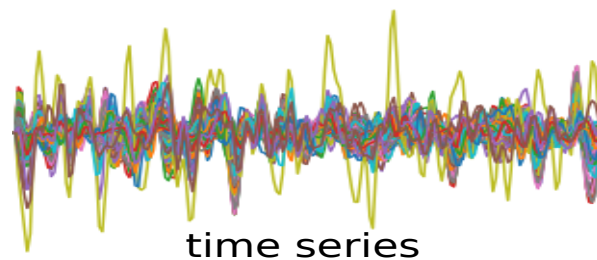
**From the  
initial graph**



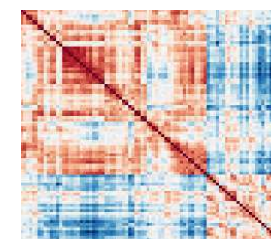




**extract**

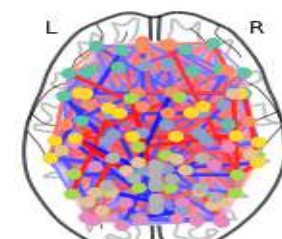


**compute**



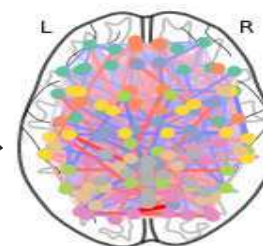
connectivity matrix

**transform**

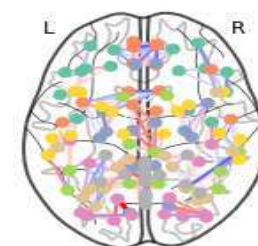


connectivity graph (IM)

**extract**

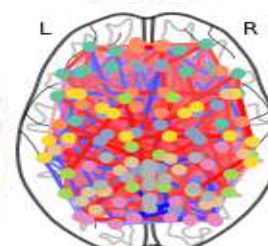


long-range connectivity matrix (HC-LRM)

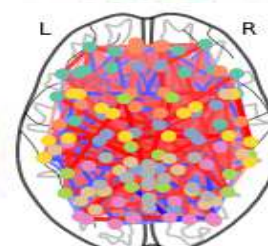


local connectivity matrix (HC-LM)

**eliminate the MST**

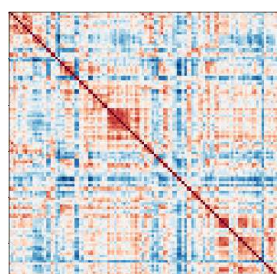


M-MST



HCM-MST

**retransform all graphs separately**



**classify**

