Pattern Recognition 101 (2020) 107183

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Identifying the best data-driven feature selection method for boosting reproducibility in classification tasks

Nicolas Georges[a,b], Islem Mhiri[a,c], Islem Rekik[a,b,1,*], the Alzheimer's Disease Neuroimaging Initiative

[a] BASIRA lab, Faculty of Computer and Informatics, Istanbul Technical University, Istanbul, Turkey
[b] School of Science and Engineering, Computing, University of Dundee, UK
[c] University of Sousse, National Engineering School of Sousse (ENISo), Sousse, Tunisia

ARTICLE INFO

Article history:
Received 24 October 2018
Revised 10 November 2019
Accepted 24 December 2019
Available online 9 January 2020

Keywords:
Feature selection methods
Multi-graph topological analysis
Feature reproducibility
Biomarker discovery
Morphological brain network
Neurological disorders
Connectomics
Cross-validation

ABSTRACT

Considering the proliferation of extremely high-dimensional data in many domains including computer vision and healthcare applications such as computer-aided diagnosis (CAD), advanced techniques for reducing data dimensionality and identifying the most relevant features for a given classification task such as distinguishing between healthy and disordered brain states are needed. Despite the existence of many works on boosting the classification accuracy using a particular feature selection (FS) method, choosing *the best* one from a large pool of existing FS techniques for boosting *feature reproducibility* within a dataset of interest remains a formidable challenge to tackle. Notably, a good performance of a particular FS method does not necessarily imply that the experiment is *reproducible* and that the features identified are optimal for the entirety of the samples. Essentially, this paper presents the first attempt to address the following challenge: "Given a set of different feature selection methods $\{FS_1, \ldots, FS_K\}$, and a dataset of interest, how to identify the *most reproducible and 'trustworthy'* connectomic features that would produce reliable biomarkers capable of accurately differentiate between two specific conditions?" To this aim, we propose FS-Select framework which explores the relationships among the different FS methods using a multi-graph architecture based on feature *reproducibility* power, *average accuracy*, and feature *stability* of each FS method. By extracting the 'central' graph node, we identify the *most reliable and reproducible FS method* for the target brain state classification task along with the most discriminative features fingerprinting these brain states. To evaluate the reproducibility power of FS-Select, we perturbed the training set by using different cross-validation strategies on a multi-view small-scale connectomic dataset (late mild cognitive impairment vs Alzheimer's disease) and large-scale dataset including autistic vs healthy subjects. Our experiments revealed *reproducible* connectional features fingerprinting disordered brain states.

1. Introduction

Recent studies have shown that neurological disorders such as Alzheimer's disease (AD) [1], autism spectrum disorder (ASD) [2,3] or mild cognitive impairment (MCI) affect the connectional morphology of the human brain [4–6]. Unraveling the morphological connectomics of these neurological and neuropsychiatric disorders [7] can help improve the diagnosis and prognosis of these conditions. To this aim, various studies leveraged machine learning techniques [4,8,9] as well as graph analysis [2,10] to spot connections between the healthy and disordered brain [11]. Once these disordered connections (or features) are identified, they may serve as biomarkers which can be targeted to improve the detection of the disease and for effective treatment [12].

In bioinformatics, researchers generally use a small sample size where each sample has a high dimensionality, which might cause issues (such as bias) for the target learning task [13,14]. Feature Selection (FS) methods have been proposed as a potential solution to this issue [15], where a subset of highly relevant features is extracted from the dataset of interest to both reduce the dimensionality of the data samples and improve the overall performance of the classifier [16]. Learning how to effectively and reliably select a subset of features with high discrimination power is one of the fundamental quests of pattern recognition since its early

* Corresponding author at: BASIRA lab, Faculty of Computer and Informatics, Istanbul Technical University, Istanbul, Turkey.
E-mail address: irekik@itu.edu.tr (I. Rekik).
[1] http://basira-lab.com/, GitHub: https://github.com/basiralab/FS-Select, YouTube video: https://www.youtube.com/watch?v=9HbLxNef2t8&feature=youtu.be.

https://doi.org/10.1016/j.patcog.2019.107183
0031-3203/

foundation [17–20]. Feature selection from high-dimensional data has been extensively studied with a wide spectrum of applications [21–23]. A growing number of works continue to investigate existing FS methods in an attempt to select the best FS techique for their target application [22,24,25]. These showed that the performances of FS methods largely varied with the the input datasets and thus the produced results are influenced by the method chosen [26,27]. On the other hand, developing a new approach that would produce the best classification result and identify the most reliable features for *all* data types seems to be an intractable problem. Furthermore, the ongoing proliferation of multi-source medical data, including structural and functional magnetic resonance imaging (MRI) data collected for the human brain connectome project [28] presents unprecedented challenges to devising feature selection methods that generate reproducible biomarkers across different data sources. This is because each data source has its unique characteristics and statistical distribution that might not match that of another data source. Hence, identifying the best feature selection method that unravels the inherent traits of a particular dataset remains a major challenge.

However, besides the improvement achieved in the past years in devising robust and precise FS methods [29] to identify reliable biomarkers for neurological disorders [30–32], new challenges have arisen including instance stability and scalability [33]. Operating on small datasets induces an inevitable variability in the results [20]. To address this issue, several studies have investigated the stability of FS algorithms [34], which measures the robustness of the selected features to perturbations in the data [35]. A better resistance to perturbations leads to a better consistency in the results and thus an improved reproducibility. It explains why stability is now even considered of the same importance as accuracy by some papers [36]. Undeniably and especially in bioinformatics, the results need to be reproducible across patients sharing the same condition. Each discovered biomarker needs to be reproducible and stable. Being able to rely on a stable FS method that is 'optimal' for a specific dataset and could detect robust, reproducible biomarkers would constitute a radical change for detecting disordered brain changes through connectomic data. Our hypothesis is that the best performing FS method for a dataset of interest might not be optimal for a different dataset in *both* classification accuracy and feature reproducibility. Basically, the question that we aim to address in this work is: "Given a series of different feature selection methods $\{FS_1, \ldots, FS_K\}$, and a dataset of interest, how to identify the *most reproducible and 'trustworthy'* connectomic features that would produce reliable biomarkers capable of accurately differentiate between two specific conditions?" (Fig. 1).

In contrast to methods focusing on boosting the accuracy (or improving solely the stability [37]) of FS methods [38] in classifying different brain states, our primary goal is not to maximize exclusively the performance of the classifier but to identify the best FS method that will produce reproducible brain features associated with a specific brain disorder (i.e., potential biomarkers) for a dataset of interest. To this aim, we propose FS-Select framework which models the relationships among the different FS methods using a multi-graph architecture to identify the most trustworthy FS method that finds the most reproducible features for a dataset of interest. In particular, we propose three graphs, modeling respectively, the relationship between FS methods in *reproducibility, similarity in average accuracy* and feature *stability* of each FS method for a number of best ranked features (i.e., a 'feature threshold' *K*). Ultimately, by integrating all reproducibility, accuracy similarity and stability graphs, we generate a holistic graph which allows to identify the central FS method with most reproducible features *in relation* to other FS methods in the graph. The weight of an edge connecting two FS nodes in the final graph represents the overlap in top *K* ranked features balanced by accuracy and stability. This allows to identify, for a dataset of interest, the 'central' node (the node with the highest strength), which will be used to identify the most meaningful and reproducible connectomic features for a brain disorder of interest.

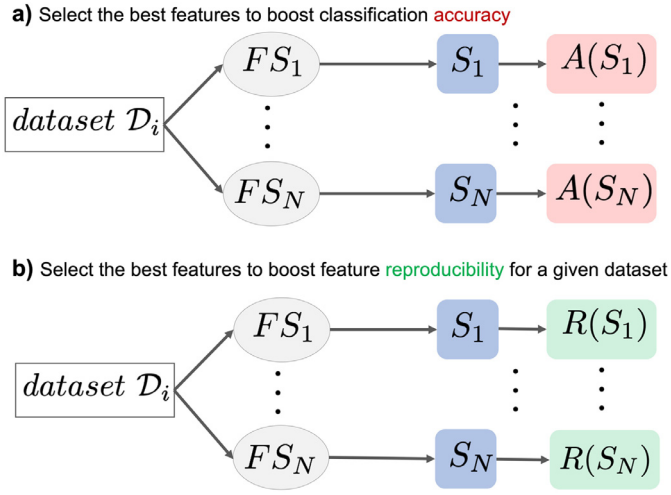

Fig. 1. Diagram of the widely used protocol for identifying the best feature selection method for a dataset of interest. (a) Given a dataset of interest $\mathcal{D}_i$ and a pool of feature selection methods $\{F_1, \ldots, F_N\}$, typical protocols rely on finding the method which selects the optimal subset of features *S*, producing the best classification accuracy *A*. However, this overlooks the issue of feature reproducibility, which is fundamental for identifying trustworthy biomarkers in biological and clinical applications. (b) Proposed diagram of data-driven protocol for identifying the best feature selection method with the most reproducible selected set of features.

Our framework is simple, intuitive, and presents the first attempt to tackle the challenging problem of identifying the most *reproducible* biomarkers for different neurological conditions. It is also generic and can be applied to any dataset for identifying *reproducible patterns* in the data. The contributions of this paper are the following:

- It unprecedentedly solves the problem of identifying the most reproducible FS method for a dataset of interest by devising a simple but effective graph-based analysis framework to model the multifaceted relationships between a set of FS methods.
- We bring up the importance of investigating *the relationship* between different FS methods –an aspect generally neglected in the quest of the best FS method for a particular dataset of interest.
- It introduces the centrality concept rooted in the field of social sciences into the best data-driven FS identification problem.
- It is able to identify the most reproducible FS method for both small and large datasets of interest and discover disordered brain connectivity biomarkers.

2. Perspective on the issue of feature selection methods and reproducibility

2.1. A diverse pool of feature selection methods

For classification problems, the presence of a huge number of features may lead to an overfitting of the learning model. Hence, FS methods aim to select only highly discriminative features. Depending on the availability of training labels, FS methods can be grouped into three categories: unsupervised, semi-supervised and supervised techniques [39]. Unsupervised FS methods may exploit data distribution or data variance to evaluate the relevance of features without labels [40–42]. The common drawback of these approaches is the neglect of correlation between different features. Semi-supervised methods generally use a small number of labeled

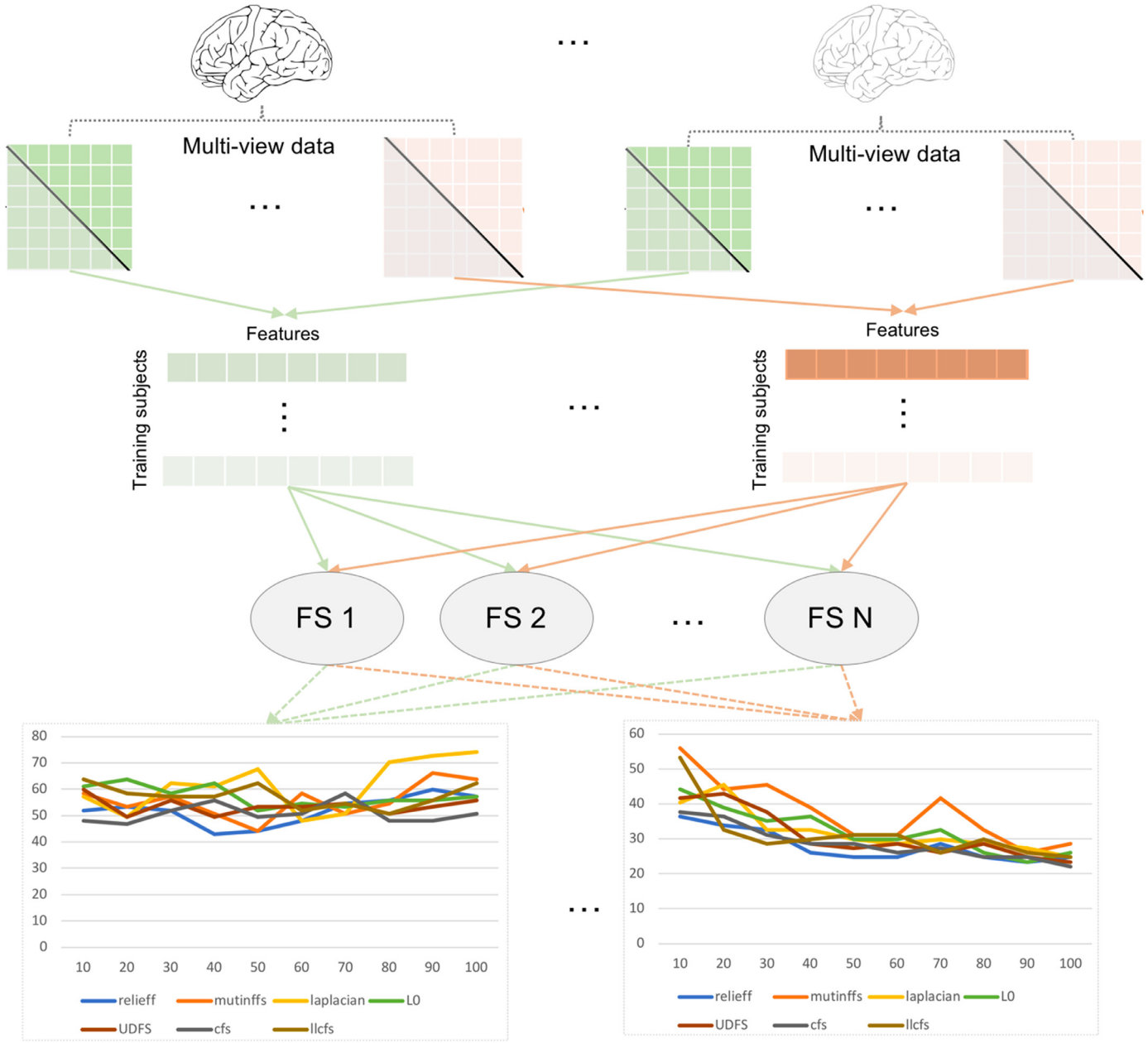

Fig. 2. Fluctuation of feature selection methods's performance across different datasets. For each subject, we define connectomic feature vectors, each derived from a particular brain view. Since each brain connectivity matrix is symmetric (i.e., connectivity) and self-connections are irrelevant, we only vectorize the off-diagonal upper triangular part of each matrix for feature extraction. We train a support vector machine (SVM) classifier using Leave-One-Out (LOO) cross-validation and seven Feature Selection (FS) methods on different datasets, each derived from a particular representation (or view) of brain connectivity. The right graph plots the classification accuracy of 7 FS methods against different numbers of selected features for brain connectivity dataset derived from view 1 (maximum principal curvature brain view), while the left graph plots classification accuracy using the same FS methods for a second dataset derived from view 2 (the mean cortical thickness brain view). We note that the performance of different FS methods varies with data types.

data to guide the feature selection process [43–45]. The most discussed techniques are the supervised feature selection algorithms [46–48], which use the labeled data to select the most discriminative features. These techniques are grouped into three categories: wrapper, filter and embedded models, depending on the utility of the selected features [15]. The first category uses the performance of a learned model (often a classifier) to identify most discriminative features [49,50]. Despite their simplicity, wrapper methods have a high risk of overfitting. The second category looks only at the properties of the data such as correlation, dependency and distance [51,52]. They are more scalable than wrapper methods on computational complexity. However, they do not consider the interaction with the classifier, which might worsen classification accuracies. To overcome the high computational training cost of these two categories, embedded methods were introduced [53,54]. These methods utilize the selected features as part of the training process and model fitting. They are less computationally intensive than wrapper methods and are able to model the interaction with the classification model.

Given this large pool of FS methods, the best performing method on a particular dataset of interest may not show the same performances for a different dataset as demonstrated in Fig. 2. Although classification accuracy induced by a particular FS method allows to evaluate the discriminative power of the selected features, it does not allow to measure the reproducibility power of the selected method, which is paramount for biomedical applications

for developing effective treatments based on the identified potential biomarkers.

2.2. Cherry-picking feature selection methods for classification tasks

Typically, feature selection algorithms for classification tasks are evaluated through classification accuracy. However, for the same dataset, different subsets of features can be identified by the same feature selection algorithm when slightly perturbing the training dataset (e.g., changing the cross-validation strategy) or other FS methods that achieve similar predictive accuracy [36]. For applications which rely on the interpretability of the selected features, one cannot trust such FS algorithms as they are non-reproducible and robust against training dataset perturbation. This issue motivated researchers to seek other evaluation metrics. The closest measure related to reproducibility is feature selection stability. As mentioned in [55], unstable feature selection leads to unstable feature subsets. Feature selection stability is defined as the sensitivity of the feature selection process to a small data perturbation in the training set [56]. Even underlying parameters such as feature dimensionality and sample size can greatly affect the stability of an algorithm [57], which might decrease our confidence in the results. More importantly, given a particular dataset of interest and a pool of FS methods, it remains challenging how to identify 'the best' feature selection method for a target classification task. Typically, such selection criteria are based on simply comparing the classification accuracy of different FS methods on the input dataset, then picking the one with the highest accuracy. This 'cherry-picking' might not be effective since the accuracy generally rises and drops with different cross-validation schemes [27]. On the other hand, such a widely used protocol might fail to identify reproducible feature sets across different cross-validations (Figs. 1–3). In this paper, we propose the first automated framework for automatically identifying the best FS method in terms of *reproducibility*.

3. Proposed FS-Select framework

In this section, we detail the key steps that constitute FS-Select framework illustrated in Fig. 3. FS-Select aims to identify the FS method that produces 'the most agreed upon' features for distinguishing between two groups drawn from a particular data of interest, when perturbing the training set. For easy reference, Table 1 presents the main mathematical notations used to design FS-Select.

3.1. FS-to-FS multi-graph construction

Given a particular dataset of interest, we aim to identify the best feature selection method that gives the most reproducible and reliable features allowing to tease apart two classes (e.g., healthy and disordered brain states). We hypothesize that *the most reliable FS method is able to reproduce the top most discriminative features identified by other methods, thereby achieving the highest consensus with other FS methods*. The most appealing characteristic of the proposed approach is that it evaluates the importance of a given FS method while considering a set of FS methods at a given cut-off threshold K representing the number of top K ranked features selected to train the classifier (e.g., support vector machine –SVM) [58]. Given a set of N FS methods, we construct an undirected fully-connected graph composed of N nodes, where each node represents a FS method, and each edge connecting two nodes captures their relationship in a particular trait (reproducibility, accuracy, similarity, or stability). Each graph is represented as a similarity matrix (Fig. 3). Ultimately, by averaging the similarity matrices of the constructed three graphs, we get the final FS-to-FS similarity matrix $\mathbf{S}$.

3.1.1. FS-to-FS feature reproducibility matrix construction

Given a set of N FS methods $\mathcal{F} = \{FS_1, \ldots, FS_N\}$, we construct a graph $G_K = (\mathcal{V}_K, E_K)$. V_K denotes the set of nodes, each nesting a FS method in $\mathcal{F}$, while E_K represents weighted edges, which models the pairwise overlap in top K features among FS methods. Each graph G_K is represented as a similarity matrix $\mathbf{S}_K$ (Fig. 3). By varying the cut-off values K, we define a set of graphs $\mathcal{G}$ (or multi-graph) that model the overlap between FS methods at different values. Next, for easily merging the generated multiple graphs, we represent each G_K as a similarity matrix $\mathbf{S}_K$ (Fig. 3), where each element $\mathbf{S}_K(i, j)$ denotes the overlap in top K ranked features between FS methods i and j. We generate an average similarity matrix $\bar{\mathbf{S}}$ by merging all similarity matrices across all thresholds, thereby capturing the *average FS method consensus* with other methods (Fig. 3).

3.1.2. FS-to-FS accuracy similarity matrix construction

Since classification accuracy influences the credibility of the produced distinctive features, we propose to model the relationship between FS methods in terms of similarity in average classification accuracy. Hence, we define an average accuracy similarity matrix $\bar{\mathbf{A}}$, where the cost $\bar{\mathbf{A}}(i, j)$ of an edge connecting two nodes i and j is defined as $\bar{\mathbf{A}}(i, j) = \exp(-|\bar{a}_i - \bar{a}_j|/\sigma_A)$, where $\bar{a}_i$ represents the average accuracy of FS method i at different cut-off thresholds. In our experiments, σ_A is set to 10 for range normalization.

3.1.3. FS-to-FS stability matrix construction

Having a performant classifier and an overall good accuracy in the classification results is important; however when dealing with biomarkers, reproducibility is crucial. Results need to be valid for every subject. A few studies have been carried out [36] highlighting the importance of a FS method's stability for the reproducibility of the results for a specific pair of FS methods. One way to better identify reproducible features is to further leverage the stability score which models the robustness of the features selected by a FS method. Similarly to building $\bar{\mathbf{S}}$ by averaging multi-graphs at different feature numbers, we introduce a third graph, represented by a matrix $\bar{\mathbf{K}}$. This is an average of stability matrices produced at different numbers of top ranked features. Each element (i, j) in a stability matrix denotes the normalized Kuncheva stability score [59]) of two FS methods FS_i and FS_j.

Finally, we integrate all $\bar{\mathbf{A}}$, $\bar{\mathbf{S}}$ and $\bar{\mathbf{K}}$ using element-wise multiplication to output the final FS similarity matrix $\mathbf{S} = \bar{\mathbf{A}} \times \bar{\mathbf{S}} \times \bar{\mathbf{K}}$ (Fig. 3).

3.2. Identifying most reproducible FS method

In graph theory, one can determine the importance of a node in a graph using centrality measure [60]. The concept of node centrality aims to quantify node importance within a graph [61]. Interestingly, such a concept has not been widely explored outside the field of social network analysis [62,63]. Node centrality presents a powerful tool in measuring the relevance of a node in a graph. To solve our problem, we bring the so-called graph centralities into the identification process of the most reproducible FS method. Specifically, we propose to use centrality measures on the estimated FS adjacency graph matrix $\mathbf{S}$, taking into account the significance of FS methods in reproducibility with respect to each other. To the best of our knowledge, our approach is the first to explicitly adopt centralities for 'best' FS method selection. As highlighted in Section 2.2, existing methods tend to rely on FS cherry-picking by comparing their performances without modeling or exploring their intrinsic topological relationships.

We formalize the definition of feature selection method reproducibility below.

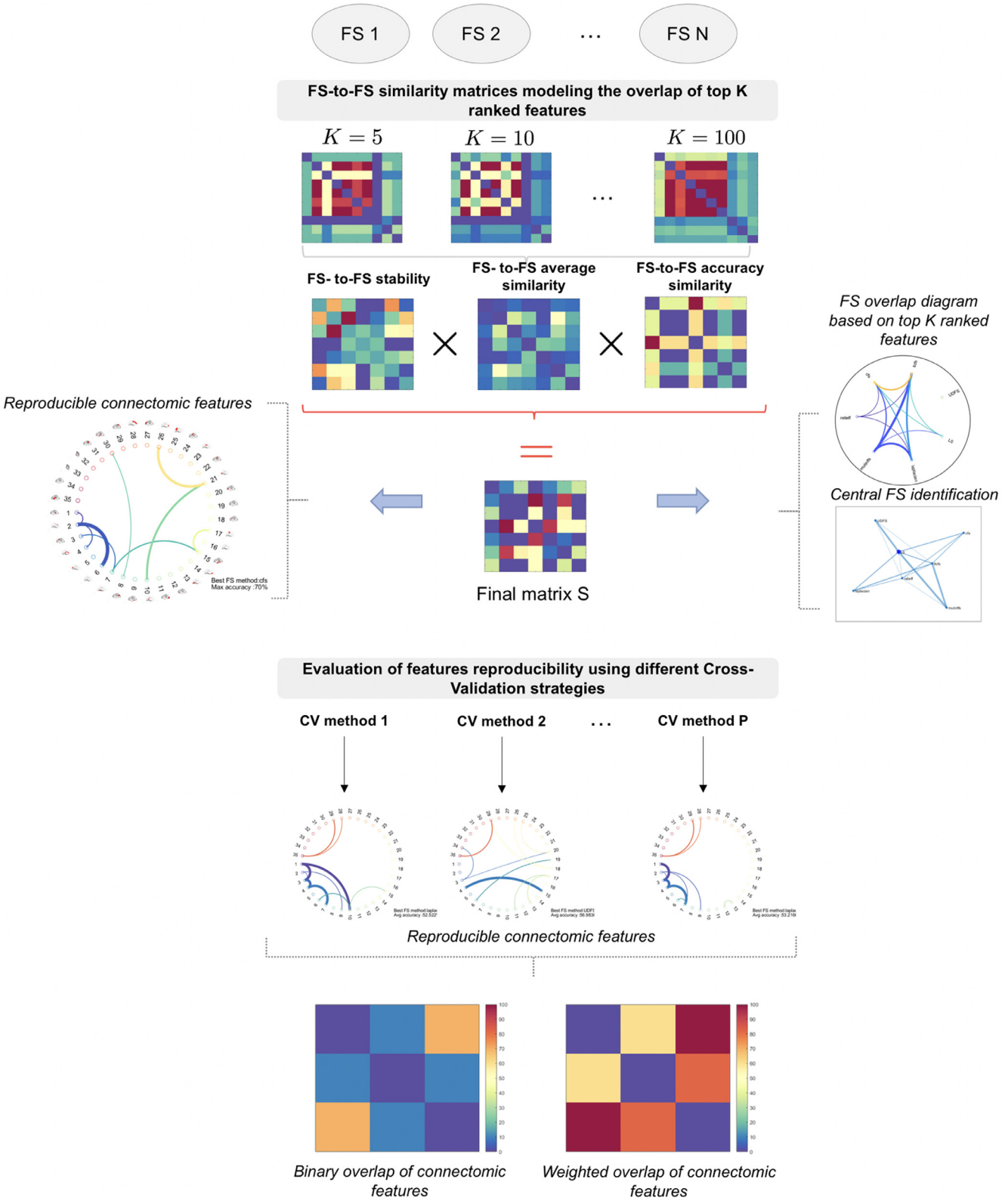

Fig. 3. Proposed FS-Select pipeline for data-specific feature selection method identification. Given a particular data view, we define multiple graphs, each represented as a similarity matrix modeling the consensus in top K ranked features among other selection methods. Next, we define an accuracy similarity matrix measuring the pairwise difference in average accuracy between FS methods and a stability matrix where each element denotes the Kuncheva stability score between two FS methods to boost the reproducibility of the selected features. By merging the reproducibility, accuracy and stability matrices, we generate a final matrix **S**. The best FS method for the dataset of interest is identified as the node with the highest centrality in **S**, thereby allowing to identify the most reproducible features distinguishing between two brain states (e.g., healthy vs disordered states). To evaluate the reproducibility power of FS-Select, we assess the binary and weighted overlaps in identified features by FS-Select using different cross-validation (CV) strategies. The circular graphs display the top 10 brain connectivities (features) selected by each CV strategy. A circular edge connects two brain regions and its strength represents the discriminative power of the selected brain connectivity. The three rows in the bottom matrices respectively represent the binary and weighted overlap in selected features using LOO, 5-fold and 10-fold CV strategies, respectively. Both matrices give insights into the reproducibility of a given FS method.

Table 1
Major mathematical notations used in this paper.

Mathematical notation	Definition
n_v	Number of views
$\mathbf{V}$	Brain network (single view) in $\mathbb{R}^{n\times n}$
$\mathbf{v}_k$	Feature vector for each brain network view k
K	Cut-off threshold representing the top ranked features
N	Number of FS methods
G_K	Graph representing feature overlap across FS methods for top K ranked features
$\mathcal{V}_K$	Set of nodes
E_K	Weighted edges
$\mathcal{G}$	Multi-graph representing feature overlap across FS methods at different cut-off thresholds K
$\mathbf{S}_K$	Similarity matrix
$\bar{\mathbf{S}}$	Average feature overlap similarity matrix
$\bar{\mathbf{A}}$	Average accuracy similarity matrix
$\mathbf{r}_K$	Features' ranking for FS methods K
$\bar{\mathbf{K}}$	Average stability similarity matrix
$\mathbf{S}$	Holistic FS graph adjacency matrix
(v_i)	Node corresponding to FS method FS_i
$d(v_i, v_k)$	The shortest distance between nodes v_i and v_k
c_i	Centrality measure for FS method FS_i
$(v^\star)$	Node with the highest centrality in $\mathcal{G}$
$\mathbf{M}_b$	Binary FS reproducibility matrix
$\mathbf{M}_w$	Weighted FS reproducibility matrix
n_f	Number of selected features
P	Number of cross-validation strategies
$\mathbf{w}_p^K$	Ranking score for threshold K and CV strategy p

Definition of FS method reproducibility at threshold K: We define the reproducibility of a feature selection method FS_i at threshold K as the average overlap ratio of its shared top K ranked features with other FS_j methods in $\mathcal{F}$.

Definition of average FS method reproducibility: We define the average reproducibility of a feature selection method FS_i as its average reproducibility ratio when varying the threshold K within a pre-defined internal.

In what follows, we use 'average reproducibility' to quantify the reproducible power of a given FS method.

Using closeness centrality, the most central nodes are strongly connected to their neighbors. In our case, this means that the FS method which shares the largest number of features (i.e., strongest connections) with other node graphs (i.e., FS methods) has the highest reproducible power. It naturally follows from the definition of FS reproducibility that the most reproducible FS methods are nodes in $\bar{\mathbf{S}}$ with highest closeness centrality in weighted graph.

Hence, to identify the most reproducible FS method, we identify the node $v^\star$ with the highest closeness centrality $c(v^\star)$ [64] in the holistic graph $\mathbf{S}$ defined as follows:

$$c(v^\star) = \max_i c(v_i) = \max_i \sum_{v_k \in V - \{v_i\}} \frac{N-1}{d(v_i, v_k)} \tag{1}$$

where $d(v_i, v_k)$ denotes the shortest distance between nodes v_i and v_k. By inverting the similarity between two nodes, we intuitively measure their distance.

Inspired by graph analysis theory, we define c_i as the closeness centrality measure, indicating to which degree a node is able to spread information to other nodes in a relatively short time. Specifically, we assign a score c_i for each FS_i in $\mathbf{S}$, that quantifies the consensus in reproducibility, stability, and accuracy among other methods. The final FS method is selected as the one with the highest closeness centrality in $\mathbf{S}$ (i.e. the one which is closest to other FS techniques). It is marked with a $\star$ in the $\mathbf{S}$ graph displayed in Fig. 3.

3.3. Identifying most reproducible connectomic features

Once the most reliable FS method is identified, we train an SVM classifier using the top K selected features to reveal the most discriminative ones. We then investigate more deeply the reproducible features by plotting the top n_f most relevant connectomic features using a circular graph which also displays the name of the best FS method and its average accuracy for this particular data set (Fig. 3).

3.4. Evaluation of FS-Select using different cross-validation strategies

In order to evaluate the reproducibility of FS-Select and have a better assessment of its effectiveness, we train a linear SVM classifier using P Cross-Validation (CV) strategies. To illustrate the similarity between FS methods in terms of the three landmark traits (i.e., reproducibility, accuracy, and stability), we created both a binary $\mathbf{M}_b$ and a weighted matrix $\mathbf{M}_w$ (Fig. 3). Each element in the first matrix simply includes the top K feature overlap (in %) between two different CV strategies p and p': $\mathbf{M}_b(p, p') = \frac{(\sum \mathbf{r}_p^K \cap \mathbf{r}_{p'}^K) \times 100}{K}$, where $\mathbf{r}_p^K$ denotes the ranking vector for top K features using the pth CV strategy. To generate the weighted stability matrix $\mathbf{M}_w$, we first identify the top K ranked features between CV strategies p and p', then we average their corresponding ranking scores $\mathbf{w}_p^K$ and $\mathbf{w}_{p'}^K$ produced by CV p and p', respectively, to produce $\mathbf{M}_w(p, p')$.

4. Results and discussion

4.1. Evaluation datasets

We evaluated FS-Select on a multi-view small-scale connectomic dataset (late mild cognitive impairment vs Alzheimer's disease) and large-scale dataset including autistic vs healthy subjects as follows.

4.1.1. Multi-view connectomic feature extraction

Each brain is represented by a set of n_v networks $\{\mathbf{V}_i\}_{i=1}^{n_v}$, each encoding a particular view of the connectional brain construct. To train our classification model based on the identified FS method, we define a feature vector $\mathbf{v}_k$ for each brain network view $\mathbf{V}_k$, whose elements belong to the off-diagonal upper triangular part of the corresponding connectivity matrix (Fig. 3).

4.1.2. Small-scale dataset

To distinguish between patients diagnosed with Alzheimer's disease (AD) and those with late mild cognitive impairment (LMCI), we used leave-one-out (LOO) cross validation on 77 subjects (41 AD and 36 LMCI) from ADNI data,[2] each with structural T1-w MR image [65]. We reconstructed both right and left cortical hemispheres for each subject from T1-w MRI using FreeSurfer software [66]. Next, we parcellated each cortical hemisphere into 35 cortical regions using Desikan-Killiany Atlas [66,67]. We generated two morphological brain networks [1,4] derived from $M = 2$ cortical views: *maximum principal curvature brain view and the mean cortical thickness brain view*. For each cortical attribute, we compute the strength of the morphological brain network connection linking ith region of interest (ROI) to the jth ROI as the absolute difference between the averaged attribute values in both ROIs [1,4]. Then, we extract a feature vector from the off-diagonal triangular part of each brain view matrix.

[2] http://adni.loni.usc.edu/.

4.1.3. Large-scale dataset

For generalizability and scalability, we evaluated FS-Select on a large-scale multi-view connectomic dataset comprising 341 subjects including 155 diagnosed with autism spectrum disorder (ASD) and 186 normal control (NC) subjects from ABIDE dataset.[3] Multi-view morphological brain networks and corresponding feature vectors were extracted using the aforementioned strategy as also in [3,68].

4.2. Results and Discussion

4.2.1. FS methods and training

For building our FS pool, we used the Feature Selection Library [69] provided by Matlab. We selected 7 FS methods: relieff [70], MutInfFS [71], laplacian [72], L0 [73], UDFS [74], llcFS [75], and cFS [76]. We adopted a leave-one-out cross-validation (CV) strategy to train each FS in combination with an SVM classifier. For FS methods that required parameter tuning, we used nested CV (relieff, UDFS). For each FS method, we evaluated the performance of the SVM classifier on different numbers of top K selected features varying from 10 to 100 (with a step size of 10 features). At first sight, it seems that the graphs tend to confirm our initial hypothesis claiming that depending on the data, the quality of one specific FS method varies (Fig. 2). In the next step, we explore the difference of rankings between the FS methods and identify the one with the most reproducible features and an overall satisfactory accuracy and stability.

4.3. FS-Select performance

4.3.1. Small-scale dataset (LMCI vs AD)

Fig. 4 illustrates the weighted FS similarity matrix and its corresponding graph as well as the reproducible features identified by FS-Select. This figure confirms our hypothesis that the best FS method for one data type might not be optimal for another one. For instance, relieff was identified for view 1 LH connectomic data with a classification accuracy of 61.03%, while L0 was identified for view 2 LH connectomic data with a classification accuracy of 70.3%. Furthermore, we note that there is a significant difference in the accuracy between the hemispheres ($\approx$ 70% vs $\approx$ 40%). The most discriminative morphological connectional features included the morphological connections between (i) [Superior parietal cortex (29) $\leftrightarrow$ the Insula cortex (35)] and (ii) [Caudal anterior-cingulate cortex (2) $\leftrightarrow$ Unmeasured corpus callosum (4)]. The pair of ROIs [Caudal middle frontal gyrus (3) $\leftrightarrow$ the Unmeasured corpus callosum (4)] and [Unmeasured corpus callosum (4) $\leftrightarrow$ Cuneus cortex (5)] were also regularly selected. Regions 1 (Bank of the superior temporal sulcus), 2 (Caudal anterior-cingulate cortex) and 35 (Insula cortex) were also identified as morphological hubs.

4.3.2. Large-scale dataset (ASD vs NC)

Fig. 5 illustrates the results obtained for the larger dataset (ASD vs NC). We observe some fundamental differences and similarities between Figs. 4 and 5 that help better investigate the behaviour of FS-Select. First, we note that the four selected FS methods are different (laplacian, relieff, cfs, mutinffs) and thus do not seem to depend of the brain connectomic view. Likewise, we do not notice a significant difference in the accuracy across views ($\approx$ 52% for all view). When examining the top reproducible features for this dataset, the morphological connection [Superior parietal cortex (29) $\leftrightarrow$ Insula cortex (35)] is always selected while [Caudal anterior-cingulate cortex (2) $\leftrightarrow$ Unmeasured corpus callosum (4)] and [Bank of the superior temporal sulcus (1) $\leftrightarrow$ Entorhinal cortex (6)] appear as relevant features. We also note that the most discriminative features identified for ASD differ from those identified for AD dataset. Overall, this might indicate that FS-Select is capable of selecting relevant connectomic features for a specific disorder.

4.4. Evaluation of FS-Select using multiple CV strategies

FS-Select identifies the best FS method from a given FS pool and is capable of revealing the most reproducible and discriminative features disentangling two classes in a biomedical dataset of interest. However, to the best of our knowledge, there is no consensus in biomedical data analysis state-of-the-art on how to evaluate the reproducibility of features based on machine learning. As a potential evaluation criterion, we leverage different cross-validation strategies to demonstrate feature reproducibility against diverse perturbations of the training set. In particular, we apply FS-Select using three CV strategies: leave-one-out, 5-fold and 10-fold CV. With the results presented in Figs. 6 and 7, we aim to highlight two key aspects of FS-Select:

- The impact of the stability on the results (i.e., the selected FS method and identified connectomic features).
- The reproducibility of the identified features by exploring their overlap across different CV strategies.

4.4.1. Identified most reproducible morphological brain connectivities distinguishing between AD and LMCI brain states.

FS-Select identified cfs as important FS method as it is selected 50% across all experiments (only relieff and L0 were selected for views 1 and 2). The circular graphs display the top most discriminative reproducible morphological connectivities between brain regions differentiating between AD and LMCI brain states. The most reproduced morphological connectional features across the three CV strategies include connections between: [Caudal anterior-cingulate cortex (2) $\leftrightarrow$ Unmeasured corpus callosum (4)] and [Superior parietal cortex (29) $\leftrightarrow$ Insula cortex (35)] which are comparable to the ones discovered earlier. When excluding the stability graph, nodes 1 (Bank of the superior temporal sulcus) and 2 (Caudal anterior-cingulate cortex) lose their weight and the most reproducible connectional feature becomes [Superior parietal cortex (29) $\leftrightarrow$ Insula cortex (35)]. From the CV similarity matrices displayed in Fig. 7, we notice that LOO and 10-fold CV strategies present the highest overlap (100%) in reproducing exactly the same 10 most discriminative features as shown in the weighted CV similarity matrices.

4.4.2. Identified most reproducible morphological brain connectivities distinguishing between ASD and NC brain states

For this dataset, while cfs was only selected once, relieff and laplacian were frequently selected by our method. Fig. 6 identifies [Superior parietal cortex (29) $\leftrightarrow$ Insula cortex (35)] as the top most reproducible connectional feature shared between different CV strategies. When including stability, the connectivities linking the Bank of the superior temporal sulcus (1) with Entorhinal cortex (6 and with Caudal middle frontal gyrus (3) are identified as most discriminative. When stability is not included to produce the final $\mathbf{S}$ matrix, we observe that region 1 (Bank of the superior temporal sulcus) is not frequently selected. On the contrary, region 2 (Caudal anterior-cingulate cortex) appears more regularly. The same pattern is reproduced when stability is not included and more nodes are selected only once and graphs look less similar. Including stability tend to increase the number of commonly selected features across CV strategies. Overall, the displayed circular graphs look more similar and they have more important features in common. From the CV similarity matrices displayed in Fig. 6, we can conclude that the most discriminative morphological connections identified by the 10-fold and LOO CV strategies are most

[3] http://fcon_1000.projects.nitrc.org/indi/abide/.

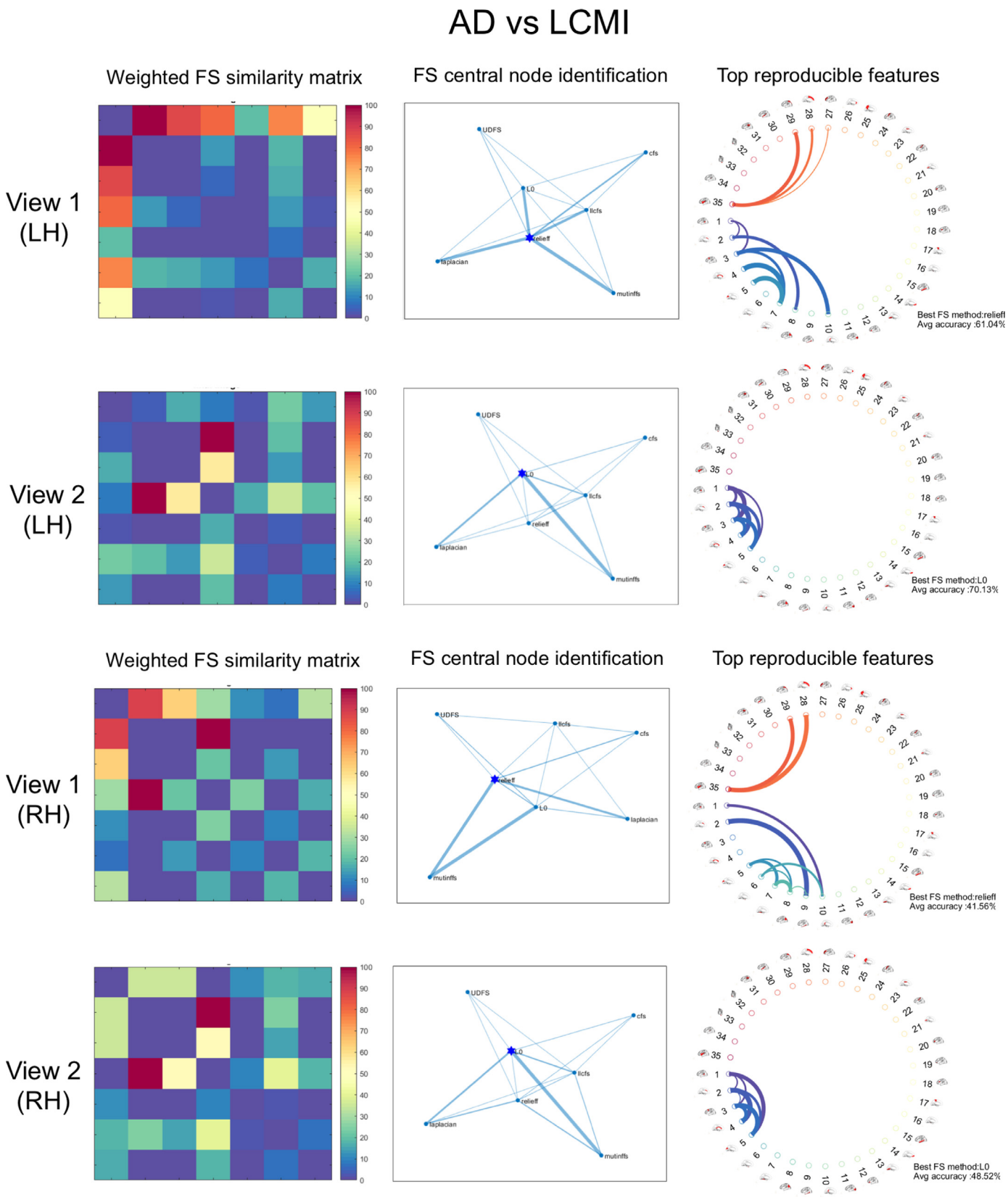

Fig. 4. AD *vs* LMCI - Top 10 reproducible discriminative features identification using the best identified feature selection (FS) method for each network brain view data. Selected FS methods (⋆), corresponding classification accuracy, and top reproducible features varied across data types and right and left hemispheres (RH and LH) for Alzheimer's Disease (AD) vs Late Mild Cognitive Impairment (LMCI) classification task.

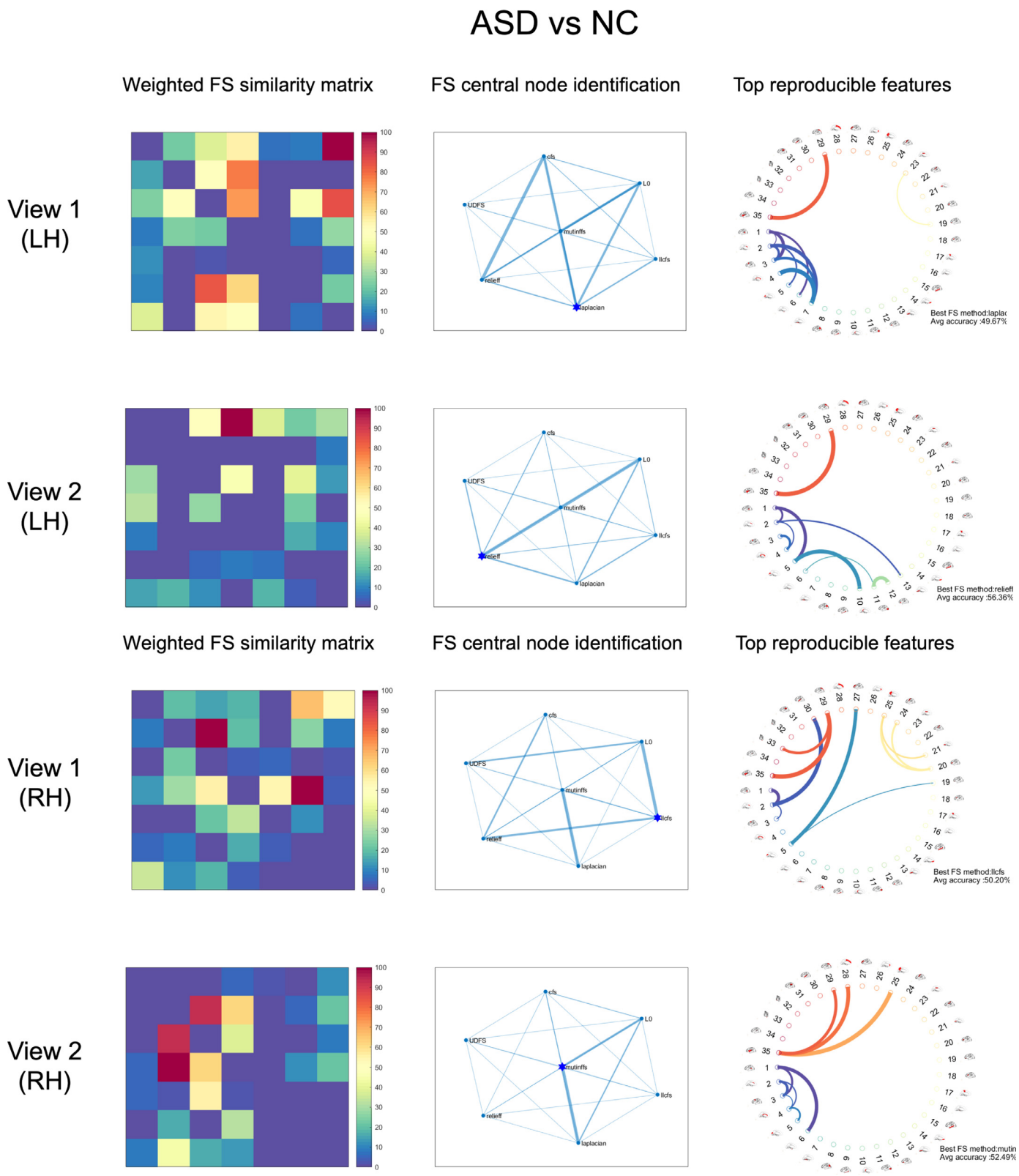

Fig. 5. *ASD vs NC - Top 10 reproducible discriminative features identification using the best identified feature selection (FS) method for each network brain view data.* Selected FS methods (⋆), corresponding classification accuracy, and top reproducible features varied across data types in the right hemisphere (RH) for Autism Spectrum Disorder (ASD) vs Normal Control (NC) classification task.

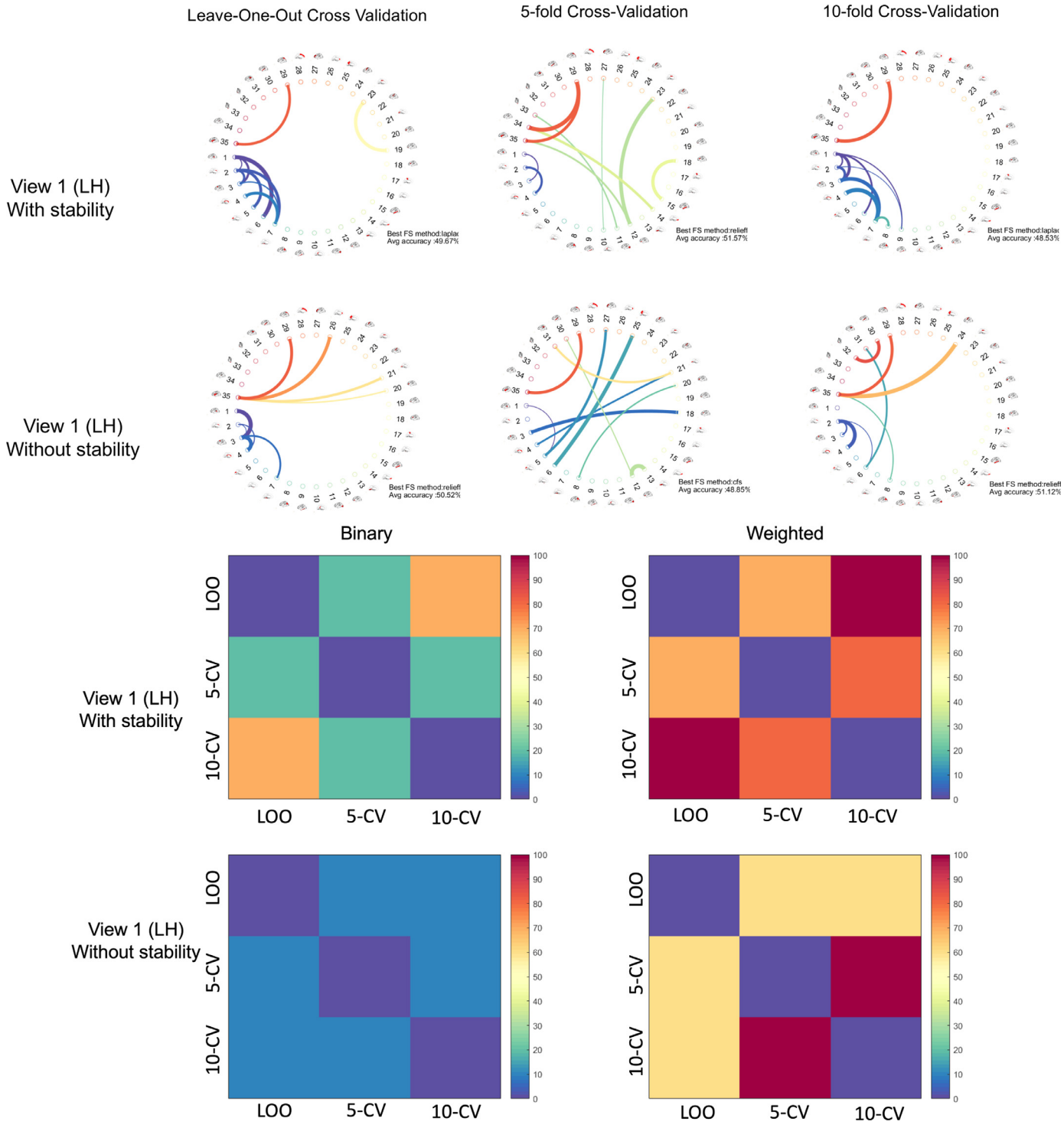

Fig. 6. Circular graphs and CV-to-CV similarity matrices representing the top 10 reproducible discriminative features for ASD/NC dataset datasets and using three cross-validation strategies (leave-one-out, 5-fold, and 10-fold) with and without the stability score. Each CV is tested with and without the inclusion of the stability criteria. The CV similarity matrices on the right present the overlap (in %) between the top 10 features discovered using a pair of CV methods. The CV matrices on the left represent the overlap (in %) between the top 10 features discovered by two CV methods weighted by their ranking scores.

reproducible since (LOO,10-CV) present the highest overlap (100%) between pairs of CV strategies as shown in the weighted CV similarity matrices for brain view 3.

4.5. Clinical findings of FS-Select

Table 2 displays the two most discriminative and reproducible morphological connections identified for each dataset and each brain view. For more visual display, Figs. 4–6 show that regardless of the input dataset and brain view, one connectional feature was consistently selected: [Superior parietal cortex (29) ↔ Insula cortex (35)]. Both cortical regions were reported in previous studies on which AD and ASD disorders [77–80].

For AD vs LMCI dataset, we conclude that connectional features [Caudal anterior-cingulate cortex (2) ↔ Unmeasured corpus callosum (4)] and [Caudal anterior-cingulate cortex (2) ↔ Entorhinal

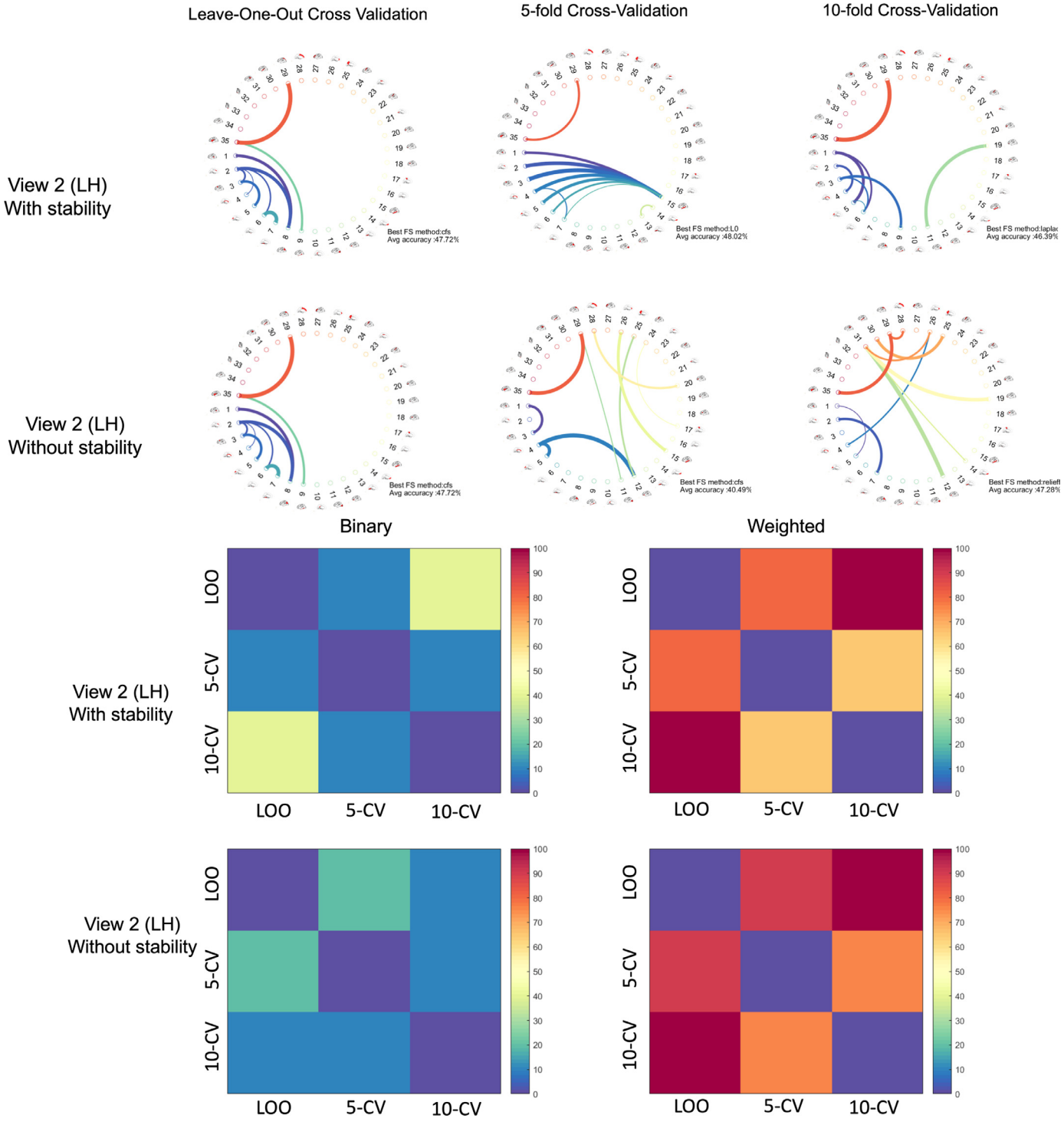

Fig. 7. Circular graphs and CV-to-CV similarity matrices representing the top 10 reproducible discriminative features for LMCI/AD dataset datasets and using three cross-validation strategies (leave-one-out, 5-fold, and 10-fold) with and without the stability score. Each CV is tested with and without the inclusion of the stability criteria. The CV similarity matrices on the right present the overlap (in %) between the top 10 features discovered using a pair of CV methods. The CV matrices on the left represent the overlap (in %) between the top 10 features discovered by two CV methods weighted by their ranking scores.

Table 2
Two most discriminative and reproducible features (i.e., brain connectivities) identified across different cross-validation strategies by the selected FS method.

Dataset	Most discriminative morphological connectivities	
AD vs LMCI	Caudal anterior-cingulate cortex (2) ↔ Unmeasured corpus callosum (4)	Superior parietal cortex (29) ↔ Insula cortex (35)
AD vs LMCI	Caudal anterior-cingulate cortex (2) ↔ Entorhinal cortex (6)	Superior parietal cortex (29) ↔ Insula cortex (35)
ASD vs NC	Bank of the superior temporal sulcus (1) ↔ Entorhinal cortex (6)	Superior parietal cortex (29) ↔ Insula cortex (35)
ASD vs NC	Caudal anterior-cingulate cortex (2) ↔ Caudal middle frontal gyrus (3)	Superior parietal cortex (29) ↔ Insula cortex (35)

cortex (6)] are identified as most discriminative and reproducible. Cortical region 2 corresponding to the Caudal anterior-cingulate cortex is found to be an important hub region, which is in line with other studies investigating AD [81].

For ASD vs NC dataset, connectional features including [Caudal anterior-cingulate cortex (2) ↔ Caudal middle frontal gyrus (3)] and [Bank of the superior temporal sulcus (1) ↔ Entorhinal cortex (6)] were detected as the most reproducible and discriminative. These morphological connections involved cortical regions reported in previous studies on Autism Spectrum Disorder [82–84].

The identified regions could be significant biomarkers and may help in the diagnosis and the treatment of both neurological conditions.

4.6. Performance of FS-Select and limitations

FS-Select achieved our primal objective of identifying the most reproducible and discriminative connectomic features for the detection of a neurological brain disorder of interest with good classification accuracy. We demonstrated the feature reproducibility power of FS-Select against different perturbations of the training set by adopting three different cross-validation strategies. FS-Select selected *the same* connectional biomarkers using at least 2 different CV out of 3. FS-Select revealed the importance of specific brain regions that were *repeatedly* identified as discriminative for all different cross-validation strategies including the bank of the superior temporal sulcus, caudal anterior-cingulate cortex, and cuneus cortex. This might indicate that these landmark regions should be primarily considered when investigating the effect of late dementia on brain morphology.

Although FS-Select has many appealing aspects, it has a few limitations that we intend to address in our future work:

- Among the pool of seven FS methods, only five were regularly selected as the most suitable for the evaluation datasets. UDFS and llcFS were never selected. In this work, we have only tested our framework on two different datasets. One would need to evaluate FS-Select on different datasets to reliably assess the potential of used FS methods.
- When investigating the most reproducible connectomic features, we have only selected the top 10 features. One can explore a larger number of features as neurological disorders might alter brain connections in different numbers depending on the severity of the condition and its stage.
- Each FS method outputs a ranking and weight vectors for features. So far, we have only considered the rank of the features for selecting the most discriminative and reproducible ones. One can also integrate the feature weight into the estimation of the reproducibility graph.
- The computational time of identifying the most reproducible FS method depends on the time complexity of the utilized FS methods as well as the data size. This can be potentially solved by using parallel computing where different FS methods are trained simultaneously, hence time complexity is not a big issue here. Besides, recent state-of-the-art FS methods have quite reasonable time complexity (e.g., time complexity of infinite feature selection of the quadratic order). Ideally, the ultimately selected FS method will be computationally least expensive, but in biological data patterns recognition tasks such as biomarker discovery for effective treatment of neurological disorders, reproducibility tips the balance compared to computation time. This paper does not focus on the time complexity of the utilized FS methods, but rather on the *reproducible power* of each FS method in selecting the most reproducible features.

4.7. Future work and improvements

There are several future directions to explore to further improve our seminal work. *First*, instead of pre-defining a similarity matrix modeling the relationship between FS methods in terms of top ranked feature consensus, we can instead learn these associations in a more generic way. *Second*, we will evaluate FS-Select on multiple connectomic datasets, including functional and structural connectomes. *Third*, ideally, the FS method giving the best classification accuracy would identify the most discriminative and reproducible features. We aim to further improve our framework to identify the data-specific FS method that satisfies both criteria. *Fourth*, in this study, we only focused on using FS-Select to demonstrate feature reproducibility *within* a dataset of interest. In our future work, we will investigate the reproducibility potential of our method across independent datasets for a specific disorder acquired from different medical centers. *Fifth*, how to evaluate the reproducibility of a given feature selection method is an open area of research that requires the development of more advanced mathematical tools for accurate and comprehensive evaluation and comparison.

5. Conclusion

While the majority of feature selection methods focus on boosting prediction accuracy, in this work, we address the issue of selecting the best FS method for a dataset of interest to boost feature reproducibility. Particularly, we introduced FS-Select, a method capable of identifying the best feature selection method to discover the most reproducible and reliable subset of features that distinguish between two groups (e.g., healthy and disorders brains). Using both small-scale and large-scale multi-view brain connectomic datasets, we demonstrated the reproducibility power of the FS method chosen by FS-Select using different cross-validation strategies. We have also discovered different reproducible connectional features fingerprinting the morphology of the autistic and demented brains. Since this is a first initiative in solving the problem of finding the most reproducible FS method for a particular dataset of interest, we only explored the pairwise relationship between different FS methods encoded in a multigraph. In our future work, we will investigate the high-order relationships between different FS methods using hypergraph learning techniques [85], where we *learn* how to model the relationship between subsets of FS methods to boost the reproducibility of discriminative data-driven patterns. Although proving mathematical claims about the behavior of even simple programs appears to be very difficult [86], providing a proof of correctness of FS-Select will lay the foundation for selecting and even designing more rigorously *reproducible* FS methods. One can also investigate alternative FS methods such as efficient and robust feature selection via joint l_{21} norms minimization [87] and more [15].

Declaration of Competing Interest

No conflict of interest to be declared.

Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech;

BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patcog.2019.107183.

References

[1] I. Mahjoub, M.A. Mahjoub, I. Rekik, Brain multiplexes reveal morphological connectional biomarkers fingerprinting late brain dementia states, Sci. Rep. 8 (1) (2018) 4103.
[2] C. Morris, I. Rekik, Autism spectrum disorder diagnosis using sparse graph embedding of morphological brain networks, Graphs Biomed. Image Anal. Comput. Anat. Imaging Genet. (2017) 12–20.
[3] M. Soussia, I. Rekik, Unsupervised manifold learning using high-order morphological brain networks derived from T1-W MRI for autism diagnosis, Front. Neuroinform. 12 (2018) 70.
[4] A. Lisowska, I. Rekik, A.A. AbbVie, A.D.D. Foundation, A. Biotech, I. Bio-Clinica, Biogen, B.-M.S. Company, I. CereSpir, Cogstate, et al., Joint pairing and structured mapping of convolutional brain morphological multiplexes for early dementia diagnosis, Brain Connect. 9 (1) (2019) 22–36.
[5] R. Raeper, A. Lisowska, I. Rekik, Cooperative correlational and discriminative ensemble classifier learning for early dementia diagnosis using morphological brain multiplexes, IEEE Access 6 (2018) 43830–43839.
[6] S. Dhifallah, I. Rekik, for the Alzheimer's Disease Neuroimaging Initiative, Estimation of connectional brain templates using selective multi-view network normalization, Med. Image Anal. 59 (2020) 101567.
[7] A. Fornito, A. Zalesky, M. Breakspear, The connectomics of brain disorders, Nat. Rev. Neurosci. 16 (3) (2015) 159.
[8] I. Mahjoub, M.A. Mahjoub, I. Rekik, Brain multiplexes reveal morphological connectional biomarkers fingerprinting late brain dementia states, Sci. Rep. 8 (1) (2018) 4103.
[9] F. Zhao, H. Zhang, I. Rekik, Z. An, D. Shen, Diagnosis of autism spectrum disorders using multi-level high-order functional networks derived from resting-state functional mri, Front. Hum. Neurosci. 12 (2018) 184. doi: 10.3389/fnhum.
[10] E. Bullmore, O. Sporns, Complex brain networks: graph theoretical analysis of structural and functional systems, Nat. Rev. Neurosci. 10 (3) (2009) 186.
[11] A. Shahrjooihaghighi, H. Frigui, X. Zhang, X. Wei, B. Shi, A. Trabelsi, An Ensemble Feature Selection Method for Biomarker Discovery, IEEE, 2017, pp. 416–421.
[12] C.-W. Woo, L.J. Chang, M.A. Lindquist, T.D. Wager, Building better biomarkers: brain models in translational neuroimaging, Nat. Neurosci. 20 (3) (2017) 365.
[13] K.S. Button, J.P. Ioannidis, C. Mokrysz, B.A. Nosek, J. Flint, E.S. Robinson, M.R. Munafò, Power failure: why small sample size undermines the reliability of neuroscience, Nat. Rev. Neurosci. 14 (5) (2013) 365.
[14] S.J. Raudys, A.K. Jain, Small sample size effects in statistical pattern recognition: Recommendations for practitioners, IEEE Trans. Pattern Anal. Mach.Intell. (3) (1991) 252–264.
[15] Y. Saeys, I. Inza, P. Larrañaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (19) (2007) 2507–2517.
[16] H. Liu, H. Motoda, R. Setiono, Z. Zhao, Feature selection: An ever evolving frontier in data mining, Feature Sel. Data Mining (2010) 4–13.
[17] E.A. Patrick, Fundamentals of Pattern Recognition, 1972.
[18] M. Dash, H. Liu, Feature selection for classification, Intell. Data Anal. 1 (1–4) (1997) 131–156.
[19] I. Guyon, A. Elisseeff, Special issue on variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
[20] L.I. Kuncheva, J.J. Rodríguez, On feature selection protocols for very low-sample-size data, Pattern Recognit. 81 (2018) 660–673.
[21] L. Yu, H. Liu, Feature selection for high-dimensional data: a fast correlation-based filter solution, in: Proceedings of the 20th International Conference on Machine Learning (ICML-03), 2003, pp. 856–863.
[22] J. Hua, W.D. Tembe, E.R. Dougherty, Performance of feature-selection methods in the classification of high-dimension data, Pattern Recognit. 42 (3) (2009) 409–424.
[23] X. Zhu, Z. Huang, Y. Yang, H.T. Shen, C. Xu, J. Luo, Self-taught dimensionality reduction on the high-dimensional small-sized data, Pattern Recognit. 46 (1) (2013) 215–229.
[24] Y. Yang, J.O. Pedersen, A comparative study on feature selection in text categorization, volume 97, 1997, pp. 412–420.
[25] T. Li, C. Zhang, M. Ogihara, A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression, Bioinformatics 20 (15) (2004) 2429–2437.
[26] A.-C. Haury, P. Gestraud, J.-P. Vert, The influence of feature selection methods on accuracy, stability and interpretability of molecular signatures, PloS ONE 6 (12) (2011) e28210.
[27] K. Dadi, M. Rahim, A. Abraham, D. Chyzhyk, M. Milham, B. Thirion, G. Varoquaux, Benchmarking functional connectome-based predictive models for resting-state FMRI (2018).
[28] D.C. Van Essen, M.F. Glasser, The human connectome project: progress and prospects, Cerebrum 2016 (2016).
[29] H. Liu, H. Motoda, Computational Methods of Feature Selection, 2007.
[30] A. Lisowska, I. Rekik, A.D.N. Initiative, et al., Pairing-Based Ensemble Classifier Learning Using Convolutional Brain Multiplexes and Multi-View Brain Networks for Early Dementia Diagnosis, Springer, 2017, pp. 42–50.
[31] M. Soussia, I. Rekik, High-Order Connectomic Manifold Learning for Autistic Brain State Identification, Springer, 2017, pp. 51–59.
[32] H. Wen, Y. Liu, I. Rekik, S. Wang, Z. Chen, J. Zhang, Y. Zhang, Y. Peng, H. He, Combining disrupted and discriminative topological properties of functional connectivity networks as neuroimaging biomarkers for accurate diagnosis of early tourette syndrome children, Mol. Neurobiol. 55 (4) (2018) 3251–3269.
[33] J. Tang, S. Alelyani, H. Liu, Feature selection for classification: a review, in: Data Classification: Algorithms and Applications, 2014, p. 37.
[34] A. Kalousis, J. Prados, M. Hilario, Stability of feature selection algorithms: a study on high-dimensional spaces, Knowl. Inf. Syst. 12 (1) (2007) 95–116.
[35] J.L. Lustgarten, V. Gopalakrishnan, S. Visweswaran, Measuring Stability of Feature Selection in Biomedical Datasets, 2009, American Medical Informatics Association, 2009, p. 406.
[36] Z. He, W. Yu, Stable feature selection for biomarker discovery, Comput. Biol. Chem. 34 (4) (2010) 215–225.
[37] P. Křížek, J. Kittler, V. Hlaváč, Improving Stability of Feature Selection Methods, Springer, 2007, pp. 929–936.
[38] Y.-W. Chen, C.-J. Lin, Combining SVMS with Various Feature Selection Strategies, Springer, 2006, pp. 315–324.
[39] H. Liu, H. Motoda, R. Setiono, Z. Zhao, Feature Selection: An Ever Evolving Frontier in Data Mining, 2010, pp. 4–13.
[40] H. Liu, L. Yu, Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17 (4) (2005) 491–502.
[41] J.R. Adhikary, M.N. Murty, Feature Selection for Unsupervised Learning, 2012, pp. 382–389.
[42] J. Yao, Q. Mao, S. Goodison, V. Mai, Y. Sun, Feature selection for unsupervised learning through local learning, Pattern Recognit. Lett. 53 (2015) 100–107.
[43] X. Kong, P.S. Yu, Semi-supervised feature selection for graph classification (2010) 793–802.
[44] Z. Zeng, X. Wang, J. Zhang, Q. Wu, Semi-supervised feature selection based on local discriminative information, Neurocomputing 173 (2016) 102–109.
[45] R. Sheikhpour, M.A. Sarram, S. Gharaghani, M.A.Z. Chahooki, A survey on semi–supervised feature selection methods, Pattern Recognit. 64 (2017) 141–158.
[46] G. Chandrashekar, F. Sahin, A survey on feature selection methods, Comput. Electr. Eng. 40 (1) (2014) 16–28.
[47] J. Ang, A. Mirzal, H. Haron, H. Hamed, Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection., IEEE/ACM Trans. Comput. Biol.Bioinform. 13 (5) (2016) 971.
[48] J. Novaković, Toward optimal feature selection using ranking methods and classification algorithms, Yugosl. J. Oper. Res. 21 (1) (2016).
[49] K. Chrysostomou, Wrapper feature selection (2009) 2103–2108.
[50] X.-t. Zhang, Y. Zhang, H.-r. Gao, C.-l. He, A wrapper feature selection algorithm based on brain storm optimization (2018) 308–315.
[51] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.
[52] H. Lyu, M. Wan, J. Han, R. Liu, C. Wang, A filter feature selection method based on the maximal information coefficient and gram-schmidt orthogonalization for biomedical data mining, Comput. Biol. Med. 89 (2017) 264–274.
[53] X. Zhang, X. Lu, Q. Shi, X. Xu, E.L. Hon-chiu, L.N. Harris, J.D. Iglehart, A. Miron, J.S. Liu, W.H. Wong, Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data, BMC Bioinform. 7 (1) (2006) 197.
[54] Z. Xiao, E. Dellandrea, W. Dou, L. Chen, ESFS: a new embedded feature selection method based on SFS, Rapp. Rech. (2008).
[55] A. Kalousis, J. Prados, M. Hilario, Stability of feature selection algorithms(2005) 8.
[56] S. Nogueira, G. Brown, Measuring the stability of feature selection(2016) 442–457.
[57] S. Alelyani, H. Liu, L. Wang, The effect of the characteristics of the dataset on the selection stability (2011) 970–977.
[58] S.R. Gunn, et al., Support Vector Machines for Classification and Regression, 1998. ISIS Technical Report 14(1).
[59] L.I. Kuncheva, A Stability Index for Feature Selection., 2007, pp. 421–427.

[60] M. Ashtiani, A. Salehzadeh, Z. Razaghi-Moghadam, H. Hennig, O. Wolkenhauer, M. Mirzaie, M. Jafari, Selection of most relevant centrality measures: asystematic survey on protein-protein interaction networks, (2017) 149492.
[61] S. De Sousa, W.G. Kropatsch, Graph-based point drift: graph centrality on the registration of point-sets, Pattern Recognit. 48 (2) (2015) 368–379.
[62] L.C. Freeman, Centrality in social networks conceptual clarification, Soc. Netw. 1 (3) (1978) 215–239.
[63] K. Okamoto, W. Chen, X.-Y. Li, Ranking of closeness centrality for large-scale social networks, in: International Workshop on Frontiers in Algorithmics, 2008, pp. 186–195.
[64] M.A. Beauchamp, An improved index of centrality, Behavioral science 10 (2) (1965) 161–163.
[65] S.G. Mueller, M.W. Weiner, L.J. Thal, R.C. Petersen, C. Jack, W. Jagust, J.Q. Trojanowski, A.W. Toga, L. Beckett, The alzheimer's disease neuroimaging initiative., Neuroimaging Clin. North Am. 10 (4) (2005) 869–877.
[66] B. Fischl, A. Van Der Kouwe, C. Destrieux, E. Halgren, F. Ségonne, D.H. Salat, E. Busa, L.J. Seidman, J. Goldstein, D. Kennedy, et al., Automatically parcellating the human cerebral cortex, Cereb. Cortex 14 (1) (2004) 11–22.
[67] R.S. Desikan, F. Ségonne, B. Fischl, B.T. Quinn, B.C. Dickerson, D. Blacker, R.L. Buckner, A.M. Dale, R.P. Maguire, B.T. Hyman, et al., An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest, Neuroimage 31 (3) (2006) 968–980.
[68] M. Soussia, I. Rekik, High-order connectomic manifold learning for autistic brain state identification, in: International Workshop on Connectomics in Neuroimaging, 2017, pp. 51–59.
[69] G. Roffo, Feature Selection Library (MATLAB Toolbox), 2016. arXiv: 1607.01327.
[70] I. Kononenko, E. Šimec, M. Robnik-Šikonja, Overcoming the myopia of inductive learning algorithms with RELIEFF, Appl. Intell. 7 (1) (1997) 39–55.
[71] P.A. Estévez, M. Tesmer, C.A. Perez, J.M. Zurada, Normalized mutual information feature selection, IEEE Trans. Neural Netw. 20 (2) (2009) 189–201.
[72] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, Adv. Neural Inf. Process. Syst. (2006) 507–514.
[73] J. Han, Z. Sun, H. Hao, l0-norm based structural sparse least square regression for feature selection, Pattern Recognit. 48 (12) (2015) 3927–3940.
[74] Y. Yang, H.T. Shen, Z. Ma, Z. Huang, X. Zhou, l2, 1-norm regularized discriminative feature selection for unsupervised learning, in: IJCAI Proceedings-International Joint Conference on Artificial Intelligence, 22, 2011, p. 1589.
[75] H. Zeng, Y. Cheung, Feature selection and kernel learning for local learning-based clustering, IEEE Trans. Pattern Anal. Mach.Intell. 33 (8) (2011) 1532–1547.
[76] M.A. Hall, Correlation-Based Feature Selection for Machine Learning, 1999.
[77] T.J. Silk, N. Rinehart, J.L. Bradshaw D Sc, B. Tonge, G. Egan, M.W. OBoyle, R. Cunnington, Visuospatial processing and the function of prefrontal-parietal networks in autism spectrum disorders: a functional MRIstudy, Am. J. Psychiatry 163 (8) (2006) 1440–1443.
[78] P.V. Arriagada, K. Marzloff, B.T. Hyman, Distribution of alzheimer-type pathologic changes in nondemented elderly individuals matches the pattern in alzheimer's disease, Neurology 42 (9) (1992) 1681.
[79] A.L. Foundas, C.M. Leonard, S.M. Mahoney, O.F. Agee, K.M. Heilman, Atrophy of the hippocampus, parietal cortex, and insula in alzheimer's disease: a volumetric magnetic resonance imaging study., Neuropsychiatry Neuropsychol. Behav. Neurol. (1997).
[80] S.J. Ebisch, V. Gallese, R.M. Willems, D. Mantini, W.B. Groen, G.L. Romani, J.K. Buitelaar, H. Bekkering, Altered intrinsic functional connectivity of anterior and posterior insula regions in high-functioning participants with autism spectrum disorder, Hum. Brain Mapp. 32 (7) (2011) 1013–1028.
[81] C.-Y. Wee, P.-T. Yap, D. Shen, A.D.N. Initiative, Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns, Hum. Brain Mapp. 34 (12) (2013) 3411–3425.
[82] J.A. Bastiaansen, M. Thioux, L. Nanetti, C. van der Gaag, C. Ketelaars, R. Minderaa, C. Keysers, Age-related increase in inferior frontal gyrus activity and social functioning in autism spectrum disorder, Biol. Psychiatry 69 (9) (2011) 832–838.
[83] Y. Jiao, R. Chen, X. Ke, K. Chu, Z. Lu, E.H. Herskovits, Predictive models of autism spectrum disorder based on brain regional cortical thickness, Neuroimage 50 (2) (2010) 589–599.
[84] M. Zilbovicius, I. Meresse, N. Chabane, F. Brunelle, Y. Samson, N. Boddaert, Autism, the superior temporal sulcus and social perception, Trends Neurosci. 29 (7) (2006) 359–366.
[85] D. Zhou, J. Huang, B. Schölkopf, Learning with hypergraphs: clustering, classification, and embedding, Adv. Neural Inf. Process. Syst. (2007) 1601–1608.
[86] C. Demetrescu, I. Finocchi, G.F. Italiano, Algorithm engineering, in: Current Trends in Theoretical Computer Science: The Challenge of the New Century Vol 1: Algorithms and Complexity Vol 2: Formal Models and Semantics, World Scientific, 2004, pp. 83–104.
[87] F. Nie, H. Huang, X. Cai, C.H. Ding, Efficient and robust feature selection via joint l2, 1-norms minimization, Adv. Neural Inf. Process. Syst. (2010) 1813–1821.

Nicolas Georges is a member of BASIRA lab and an MSc student within Computing at the University of Dundee. His Master's degree is in Data Engineering. His research interests include machine learning, medical data analysis, and feature selection methods.

Islem Mhiri is a Ph.D. student at the National Engineering School of Sousse (ENISo) in Tunisia and a member of BASIRA lab. Her research work includes machine learning, medical data super-resolution, and brain template estimation.

Islem Rekik is an assistant professor at Istanbul Technical University in the Faculty of Computer and Informatics and honorary researcher within Computing at the University of Dundee. Previously, she was a postdoctoral research scholar at the IDEA lab (University of North Carolina) following a Ph.D. in "Neuroimaging and Computer Sciences" from the University of Edinburgh in 2014. Currently, she is the director of BASIRA (Brain And SIgnal Research & Analysis, www.basira-lab.com) laboratory and a member of the Computer Vision & Image Processing (CVIP) group at the University of Dundee. She is a program committee member of leading international medical image analysis conferences and workshops including MICCAI, Machine Learning in Medical Imaging (MLMI), Connectomics in NeuroImaging (CNI), and Deep Learning in Medical Imaging (DLMI). She co-organized Predictive Intelligence in MEdicine (PRIME) and CNI workshops, in conjunction with MICCAI 2018 and is an area chair for MIDL 2019 Medical Imaging with Deep Learning Conference. Her research work aims to infuse advanced computer vision and machinelearning methods into big neuroimaging and signal data analysis for improving healthcare and wellbeing.