

RESEARCH

Open Access



# Density-based clustering of static and dynamic functional MRI connectivity features obtained from subjects with cognitive impairment

D. Rangaprakash<sup>1,2,3</sup>, Toluwanimi Odemuyiwa<sup>4</sup>, D. Narayana Dutt<sup>5</sup>, Gopikrishna Deshpande<sup>6,7,8,9,10,11,12,13\*</sup>  and Alzheimer's Disease Neuroimaging Initiative

## Abstract

Various machine-learning classification techniques have been employed previously to classify brain states in healthy and disease populations using functional magnetic resonance imaging (fMRI). These methods generally use supervised classifiers that are sensitive to outliers and require labeling of training data to generate a predictive model. Density-based clustering, which overcomes these issues, is a popular unsupervised learning approach whose utility for high-dimensional neuroimaging data has not been previously evaluated. Its advantages include insensitivity to outliers and ability to work with unlabeled data. Unlike the popular *k*-means clustering, the number of clusters need not be specified. In this study, we compare the performance of two popular density-based clustering methods, DBSCAN and OPTICS, in accurately identifying individuals with three stages of cognitive impairment, including Alzheimer's disease. We used static and dynamic functional connectivity features for clustering, which captures the strength and temporal variation of brain connectivity respectively. To assess the robustness of clustering to noise/outliers, we propose a novel method called recursive-clustering using additive-noise (R-CLAN). Results demonstrated that both clustering algorithms were effective, although OPTICS with dynamic connectivity features outperformed in terms of cluster purity (95.46%) and robustness to noise/outliers. This study demonstrates that density-based clustering can accurately and robustly identify diagnostic classes in an unsupervised way using brain connectivity.

**Keywords:** Functional MRI, Brain networks and dynamic connectivity, Cognitive impairment and alzheimer's disease, Unsupervised learning and clustering, DBSCAN, OPTICS

## 1 Introduction

Since the successful emergence of functional neuroimaging, a new barrier has surfaced: can a strong correlation be established between brain activity (as measured by functional magnetic resonance imaging [fMRI]) and the cognitive state of an individual? More specifically, can we accurately classify neurological diseases based on fMRI

data? In response to this, machine learning classifiers have been employed on neuroimaging features to generate models that, within some accuracy, predict the cognitive and disease states to which new data belong [1–11].

There are two major categories of machine learning classification techniques: supervised and unsupervised. Supervised learning, commonly used in fMRI studies, involves splitting the dataset into training and test data. Each member of the training data is given a 'label' as to which class (or group) it belongs to. The classifier is then 'trained' on this data to determine a generalized model (thus 'supervised' learning). Then, the classification

\*Correspondence: gopi@auburn.edu

<sup>6</sup> AU MRI Research Center, Department of Electrical and Computer Engineering, Auburn University, 560 Devall Dr, Suite 266D, Auburn, AL 36849, USA

Full list of author information is available at the end of the article

accuracy is measured by testing the model on the test data with known labels. Conversely, in unsupervised learning, patterns within the entire dataset are used to ‘cluster’ the data without any pre-assigned labels, and cluster purity is measured against the known ground-truth, post hoc, instead of an accuracy. As such, unsupervised learning is agnostic to pre-assigned labels, and thus determines inherent classes instead of fitting a model based on classes provided by us. Although the ‘cluster purity’ given by clustering is, in principle, the same as the ‘classification accuracy’ given by supervised classifiers (both give the percentage of correct classifications as against the known ground truth), we would use the term ‘cluster purity’ in this work, so as to highlight that this metric was obtained through unsupervised clustering, and not through conventional supervised classification.

fMRI studies generally use supervised learning methods to classify disease or cognitive states [1–3, 12, 13]. Unsupervised learning methods are generally used for spatially and temporally clustering voxel signals to localize brain activity [10, 11, 14–20]; for dimensionality reduction or feature selection before applying a supervised learning algorithm to the data [13, 21–24]; or for mapping specific brain patterns to a cognitive state [6, 25]. A few reports have adopted unsupervised learning methods as state classifiers across individuals. For example, one study used the one-class support vector machine (SVM) to determine the boundary around healthy control subjects and to classify outliers based on their distance from this boundary. However, such methods are highly susceptible to noise and do not work well with the high-dimensional data common in fMRI [26].

Research on clustering of subjects using unsupervised learning methods is still limited in fMRI literature, to the best of our knowledge. Diagnostic labels are often predetermined in neuroimaging studies, and most studies often stick to this labeling, because of which supervised classifiers are more popular. However, the main advantage of unsupervised learning is that it requires no a priori knowledge of categorical labels, thus reducing the problem of selection bias described by Demerci et al. [5]. This allows for the learning of more complex models, pattern discrimination and the identification of hidden states [27] among others. Unsupervised learning also does not have the issue of overfitting the data, which largely plagues supervised learning models. The overfitting nature of supervised learning models results in high performances on the data being used, but much lower performance when tested on an independent dataset. Unsupervised learning does not encounter this issue from the very nature of its formulation since learning is done in an entirely blind manner. With unsupervised learning, what you see is what you get. In this work, we

demonstrate the utility and feasibility of the unsupervised learning approach of density-based clustering, applied on resting-state fMRI (rs-fMRI) data to cluster disease states in a noise-robust manner.

Clustering is an important technique of grouping objects that are similar to each other and isolating them from those that are dissimilar. Clusters are formed with features having minimum intra-cluster distance and maximum inter-cluster distance [28]. Clustering has found wide application in many fields including, but not limited to, the recovery of information, recognition of natural patterns, the analysis of digital images, bioinformatics, data mining, taxonomy, DNA microarray analysis, and many others [29–34]. The performance of most clustering algorithms, however, is highly sensitive to their input parameters, such as the number of clusters in *k*-means clustering [35]. For effective performance, one almost needs a priori knowledge of these parameters. Although there are a variety of clustering algorithms, density-based clustering is one of the most computationally effective methods in clustering large-scale databases [36]. Unlike *k*-means clustering, density-based clustering techniques do not need the user to specify the number of clusters since the algorithm determines that by itself. Using a local proximity measure, clusters are automatically formed by points in higher density regions, well separated by lower density regions. In this work, we focus on two of the most popular density-based clustering methods: Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Ordering Points to Identify the Clustering Structure (OPTICS). They require few input parameters, need no a priori user choices, detect outliers, and effectively detect clusters of arbitrary shapes [37, 38].

DBSCAN is a single-scan technique that makes no pre-assumptions regarding the distribution of data. In addition to clustering the data, the algorithm also detects outliers. The OPTICS algorithm was developed by Ankerst et al. [38] to address DBSCAN’s major shortcoming that it cannot detect clusters with different local densities. OPTICS is a non-rational, data-independent representation of the cluster structure that displays important information regarding the distribution of the data. It is capable of identifying all possible clusters of varying shapes. Advantages of OPTICS over DBSCAN are that it eliminates free parameter choices needed in DBSCAN, is relatively insensitive to noise and can detect clusters of different densities [38].

In fMRI data, noise is any signal variation that is not contributed by neuronal activity. It may arise from head movements during scanning, scanner limitations, thermal noise, or other sources [39, 40]. The signal-to-noise ratio (SNR) is a measure of how much signal is present

in the data relative to noise. Higher SNR values indicate higher quality data with low amounts of noise. However, much progress has been made in detecting and minimizing noise artifacts from fMRI data, like high-pass filtering, motion correction and other pre-processing steps [39, 41]. Some noise is always prevalent in the data. It is important that learning methods perform well even when given lower quality data. In this work, we assess the robustness of clustering to such noisy variations.

Resting-state fMRI (rs-fMRI) involves the collection of fMRI data from subjects while at rest in the scanner, with eyes open and mind left to wander. It is characterized by correlations across various regions of the brain, also called functional connectivity (FC). FC estimated from rs-fMRI is extensively used to study brain networks, and prior works have identified alterations in FC in subjects with various psychiatric and neurological conditions (please see [42–45] for a review). We have used functional connectivity features in this work.

Static Functional Connectivity (SFC) refers to the strength of connectivity between two brain regions and is quantified using Pearson's correlation coefficient between pairs of fMRI time series. It evaluates the temporal correlation between fMRI time series of two brain regions, giving one correlation value over the entire duration of the scan. On the other hand, variance of Dynamic Functional Connectivity (DFC) [46–48] captures time-varying connectivity and is obtained using sliding-window Pearson's correlation between pairs of brain regions. Although SFC is the most widely used measure of fMRI connectivity, DFC is emerging as an important measure having certain unique special properties [42]. SFC and DFC provide characteristically different information regarding the relationship between two brain regions. While SFC gives the strength of connectivity or co-activation, DFC gives the variation of connectivity over time (please refer to Hutchison et al. [42] for a review of dynamic connectivity). In this work, in addition to evaluating the performance of DBSCAN and OPTICS clustering approaches, we compared the performance of SFC and DFC features. We explore which of these two popular measures is more suitable for clustering in cognitive impairment.

In this work, SFC and DFC features were obtained from 132 subjects with progressive stages of cognitive impairment (early mild cognitive impairment [MCI], late MCI and Alzheimer's disease), along with matched healthy controls. Cognitive impairment [49] is a spectrum disorder ranging from mild to severe symptoms. There has been enormous interest in identifying subgroups of cognitive impairment representing relatively homogeneous symptoms [49]. In this work, we performed clustering and assessed how well the clustered subject groups matched with clinically diagnosed groups. If this

approach yields impressive and reliable results, it would advance clinical diagnosis and classification of cognitive impairment.

In this work, we assessed the performance of DBSCAN and OPTICS, in combination with SFC and DFC features, in accurately clustering three groups of cognitive impairment (including Alzheimer's disease) and a matched healthy control group. Higher clustering performance need not necessarily imply higher robustness to noisy variations in the data. It is possible that a high cluster purity is obtained, but that high performance is not sustained if noise in the data increases. Noise robustness is very important for two reasons: (i) biomedical data, especially brain imaging data, are vulnerable to a large number of unknown sources of variability and known sources of noise, and (ii) the data used in any study is only a representative sample of the general population, and hence generalizability and inter-subject variability are imminent issues; therefore, if the results have to be applicable in a real-world setting, then it has to be robust to such variability. Thus, to assess robustness, we developed a novel technique called 'recursive clustering using additive noise' (R-CLAN). Additionally, since well-formed clusters are partly defined by how well the clusters are separated in the feature space, we propose the 'separation index' to supplement the R-CLAN outcome, as a measure of OPTICS' performance.

The organization of the paper is as follows: section-2 describes the methods used for grouping the dataset into clusters and our methods to assess the algorithms' robustness; section-3 presents the results with clustering performances and their robustness; section-4 evaluates and discusses the findings, and highlights possible future work; and section-5 provides concluding remarks.

## 2 Methods

### 2.1 Data acquisition and pre-processing

The fMRI data used in this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<https://www.adni-info.org/>). ADNI is a multisite, longitudinal study employing imaging, clinical, bio-specimen and genetic biomarkers in healthy elders as well as in individuals with early mild cognitive impairment (EMCI), late MCI (LMCI) and Alzheimer's disease (AD). The ADNI was launched as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial neuroimaging and other biological markers can be combined with clinical and neuropsychological assessment to measure the progression of MCI and AD. For up-to-date detailed information, please see [www.adni-info.org](http://www.adni-info.org). Table 1 provides the demographics of the subjects used by us.

**Table 1 Basic demographics**

Variable	Control	Early MCI	Late MCI	AD
Age, years				
Mean	74.5	72.2	71.4	73.1
Median	73.8	72.9	72.3	74.5
SD	5.9	5.9	8.6	7.4
Range	20.5	26.8	30.9	30.6
Gender, no. of subjects				
Male	15	18	21	13
Female	21	16	13	16

A total of 132 subjects were considered from phase-1 of the database: 35 control subjects, 34 EMCI, 34 LMCI and 29 AD patients. RS-fMRI data were acquired in 3 T Philips MR scanners using a  $T2^*$  weighted single shot echo-planar imaging (EPI) sequence with 48 slices and the following parameters: repetition time (TR) = 3000 ms, echo time = 30 ms, flip angle =  $80^\circ$ , voxel size =  $3.3125 \times 3.3125 \times 3.3125 \text{ mm}^3$ , and 140 temporal volumes in each run. Field of view parameters were the following: Right–Left (RL) = 212 mm, Anterior–Posterior (AP) = 198.75 mm, and Foot–Head (FH) = 159 mm. Anatomical images were acquired using magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence for overlay and localization (TR = 6.8 ms, echo time = 3.1 ms, voxel size:  $1.11 \times 1.11 \times 1.2 \text{ mm}^3$ , flip angle =  $9^\circ$ , field of view: RL = 204 mm, AP = 253 mm, FH = 270 mm).

Prior to pre-processing, first five volumes of the fMRI time series were discarded to allow for MR scanner equilibration. Standard rs-fMRI data preprocessing steps were performed on the raw data (realignment, normalization to MNI space, detrending, regressing out nuisance covariates [6 head motion parameters, white-matter signal and cerebrospinal fluid signal] using SPM8 [50] and DPARSF [51] toolboxes in the Matlab<sup>®</sup> environment. Mean fMRI time series were then extracted from 200 functionally homogeneous regions-of-interest (ROIs) obtained through spectral clustering (Craddock-200 atlas, [52]).

## 2.2 Obtaining connectivity features

SFC was obtained between all pair-wise ROIs using Pearson's correlation coefficient, giving a  $200 \times 200$  SFC matrix per subject. Multivariate N-way analysis of covariance (MANCOVAN) was used to perform statistical tests between SFC features of all groups, controlling for age, gender and head motion. The top-100 (i.e. lowest  $p$  values) significant features were selected for further analysis, that is, a  $132 \times 100$  (subjects  $\times$  features) SFC matrix was used for clustering. We chose a fixed number of features,

because we did not want the difference in number of features (between SFC and DFC) to impact clustering performance. DFC was obtained using sliding-windowed Pearson's correlation [46–48], giving a time series of correlation values. The variance of these values over time is a measure of how much the connectivity varies over the duration of the scan, which was the measure we used further. The width of the sliding windows was not chosen arbitrarily, like in most studies, but was instead evaluated objectively wherein the window lengths were determined adaptively using the augmented Dickey-Fuller (ADF) test [46–48], an analytical test based on timeseries stationarity. Overlapping windows were used, with successive windows differing by one time point. Similar to SFC, a  $132 \times 100$  variance of DFC matrix was obtained, which was then used in clustering. Please refer to Additional file 1: Tables S1 through S3 for the list of all the top-100 significant SFC and DFC connectivity paths from which the features were extracted for further analysis.

## 2.3 The DBSCAN algorithm

For implementation of DBSCAN and OPTICS, we wrote custom Matlab scripts using Daszykowski's functions (<https://www.chemometria.us.edu.pl/index.php?goto=downloads>). Here, we give a brief overview of the DBSCAN algorithm, as described by Ester et al. [37]. The idea, in its simplest form, is that the density of points within a cluster must exceed a certain global threshold. Regions of density below this threshold are considered as noise or outliers. Two input parameters,  $\epsilon$ , the minimum radius of a cluster, and  $MinPts$ , the minimum number of points required in a cluster, form the global threshold measures. In distinguishing regions of data by density, the following terms are used:

- $\epsilon$ -neighborhood*: For a point  $p$  in the cluster, its  $\epsilon$ -neighborhood is the set of points contained within the radius  $\epsilon$  from that point.
- Core point*: A point is a core point if its  $\epsilon$ -neighborhood contains at least  $MinPts$  points.
- Border point*: A point is a border point if there are less than  $MinPts$  number of points in its  $\epsilon$ -neighborhood.
- Density-reachable*: A point  $q$  is directly density-reachable from point  $p$  if  $p$  is a core point, and  $q$  is within the  $\epsilon$ -neighborhood of  $p$ . In general,  $q$  is density-reachable from  $p$  if there is a chain of points from  $p$  to  $q$  such that each successive point is directly density-reachable from the previous point.
- Density-connected*: Two points,  $p$  and  $q$ , are density-connected if there exists a third point from which both are density-reachable.

Starting at a core point  $p$ , DBSCAN gathers all the density-reachable points of  $p$  into a single cluster  $C$ . The

procedure is then repeated again with all the new core points in  $C$  until the cluster is completely surrounded by border points from which no other point is density-reachable. DBSCAN then repeats the process again for any unclustered points in the database until all points have been processed. Finally, the points within a cluster will all be density-connected with each other, and the points lying completely outside these clusters will be considered as outliers.

#### 2.4 Determination of radius in DBSCAN algorithm

As with the other clustering algorithms, DBSCAN is sensitive to its input parameters, particularly to  $\varepsilon$ . As noted in Tench et al. [53], if  $\varepsilon$  is too large, the algorithm will not be as discriminative and may include outliers in its clusters, and if too small, clusters may not be detected at all. Consequently, Ester et al. [37] proposed the method of  $k$ -distance graphs to determine  $\varepsilon$ . In this method, all the data points are ordered according to their distance  $d$  from their  $k$ th nearest neighbor (setting  $k = \text{MinPts}$ ), and a  $k$ -distance graph is formed by plotting  $d$  against the ordered points in the feature space. By visual inspection,  $\varepsilon$  is chosen to be the value  $d$  at which an ‘elbow’ occurs in the plot. However, this method does not provide dependable  $\varepsilon$  values [54], although it effectively identifies the range of values where one can search for a more optimal  $\varepsilon$  value.

We modified their approach by first obtaining an initial estimate of  $\varepsilon$  using the  $k$ -distance method, and then performing DBSCAN for all  $\varepsilon$  values ranging from zero to twice this initial estimate, resulting in different number of clusters for different  $\varepsilon$  values. Finally, we searched for the number of clusters that spanned the largest range of  $\varepsilon$  values, since this would identify the number of clusters that were most stable among all choices of the parameter  $\varepsilon$ . We assigned the mean  $\varepsilon$  value within this range as the final  $\varepsilon$  value for performing DBSCAN clustering. The rationale behind this approach was that those number of clusters that spanned the largest  $\varepsilon$  range would represent the most stable and true separation between the clusters.

This process can be visualized through what we term as the epsilon plot (Fig. 3, shown later in the results section), which is a graph of the number of clusters identified by DBSCAN against the range of  $\varepsilon$  values. Since the number of clusters is a discrete set of values, the plot is characterized by a series of horizontal steps, where a step corresponds to the set of  $\varepsilon$  values that give rise to that number of clusters. As described earlier, the  $\varepsilon$  value corresponding to the midpoint of the widest step was chosen as the final  $\varepsilon$  value used in the DBSCAN clustering.

#### 2.5 The OPTICS algorithm

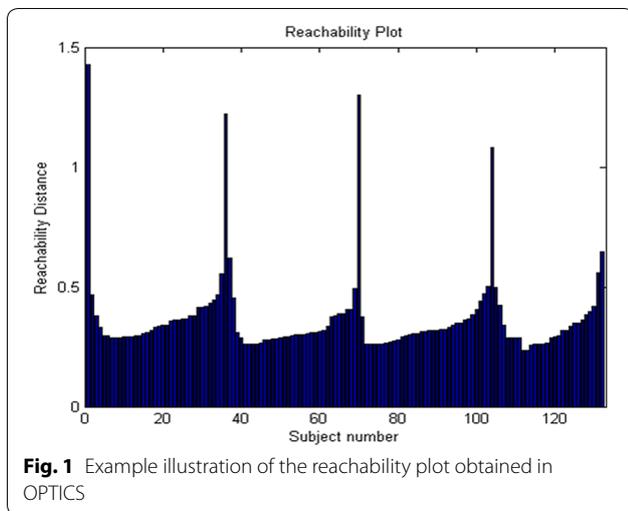
Here, we provide a brief overview of OPTICS as described by Ankerst et al. [38]. OPTICS is effectively an extension of the DBSCAN algorithm that uses a range of  $\varepsilon$  values, rather than a single, global threshold, to identify clusters of different local densities. Akin to DBSCAN, OPTICS uses input parameters *MinPts* and ‘generating distance’  $\varepsilon$ ; however,  $\varepsilon$  is now a maximum threshold value. We use  $\varepsilon'$  to denote the range of radius values used by OPTICS, where  $0 < \varepsilon' < \varepsilon$ . To understand the algorithm, two terms are defined in addition to those used in DBSCAN:

- a. *Core distance*: The core distance of a point  $p$  is the distance  $\varepsilon'$  between  $p$  and its *MinPts*’ neighbor such that  $p$  is a core point with respect to  $\varepsilon'$ . If this  $\varepsilon'$  value is greater than  $\varepsilon$ , then the core-distance is undefined.
- b. *Reachability distance (of  $p$  from  $s$ )*: If  $s$  is not a core point with respect to  $\varepsilon$ , the reachability distance is undefined. Otherwise, the reachability distance of a point  $p$  from a point  $s$  is the maximum of  $s$ ’s core distance and smallest distance such that  $p$  is density-reachable from  $s$ .

OPTICS works on the principle that to accurately distinguish between different densities of clusters; higher density clusters must first be found before lower density clusters are identified. Hence, OPTICS orders data points based on increasing  $\varepsilon'$  value, since smaller  $\varepsilon$  values indicate higher density regions. This ordered list is part of the output of OPTICS. As in DBSCAN, points not belonging to any cluster are considered outliers. Unlike DBSCAN, however, OPTICS does not explicitly assign cluster memberships; rather, points are ordered based on their reachability and core distance values. In addition, OPTICS is relatively insensitive to  $\varepsilon$  and *MinPts*. As long as a large enough value is chosen for  $\varepsilon$ , and a choice of around 10–20 is made for *MinPts* (according to [38]), consistent cluster structures will be identified. This makes implementing OPTICS practically non-arbitrary and choice-free.

#### 2.6 Separation index to evaluate robustness of reachability plot in OPTICS clustering

To assess the clustering structure in OPTICS, a reachability plot is popularly used [38]. The reachability distance is plotted for each data point, ordered according to the ordered-list output of OPTICS. Figure 1 gives an example illustration of the reachability plot defined by four clusters. A series of peaks and valleys characterize the plot, where a valley indicates a cluster. This plot effectively indicates the cluster membership of each point, although



further steps are performed in OPTICS to formally define each cluster. Further details can be found in [38]. Since obtaining the reachability plot is an intermediate step in OPTICS clustering, the success of the method critically depends on the reachability plot. Consequently, assessing the robustness of the plot is indirectly an assessment of the outcome of OPTICS clustering.

To assess the performance of OPTICS on both SFC and DFC features using the reachability plot, we devised a novel measure called the OPTICS separation index. Leaving out the first and last points of the reachability plot, we defined the separation index as the ratio of the mean of peak heights to the mean value of points between the peaks in the valleys. For example, suppose there exists peaks at points  $i$  and  $j$  in the reachability plot, with cluster- $k$  defined by the valley bounded by these peaks, then the separation index for cluster- $k$  is defined as:

$$S_k = \frac{\frac{1}{2}(RP(i) + RP(j))}{\frac{1}{j-i-1} \sum_{m=i+1}^{j-1} RP(m)}, \quad (1)$$

where  $S_k$  is the separation index for the  $k$ th group, and  $RP(i)$ ,  $RP(j)$  are the values of the reachability plot. The final separation index is an average of the separation indices obtained for each cluster. In simple terms, the separation index evaluates the ratio of the peak heights to the average of the baseline heights. A higher separation index value indicates that the peaks are relatively much larger than the baseline; hence, obtaining clusters using a threshold on the reachability plot would then be less prone to noise from outliers. This is an indirect measure of the robustness of OPTICS clustering.

## 2.7 Assessing robustness using additive noise

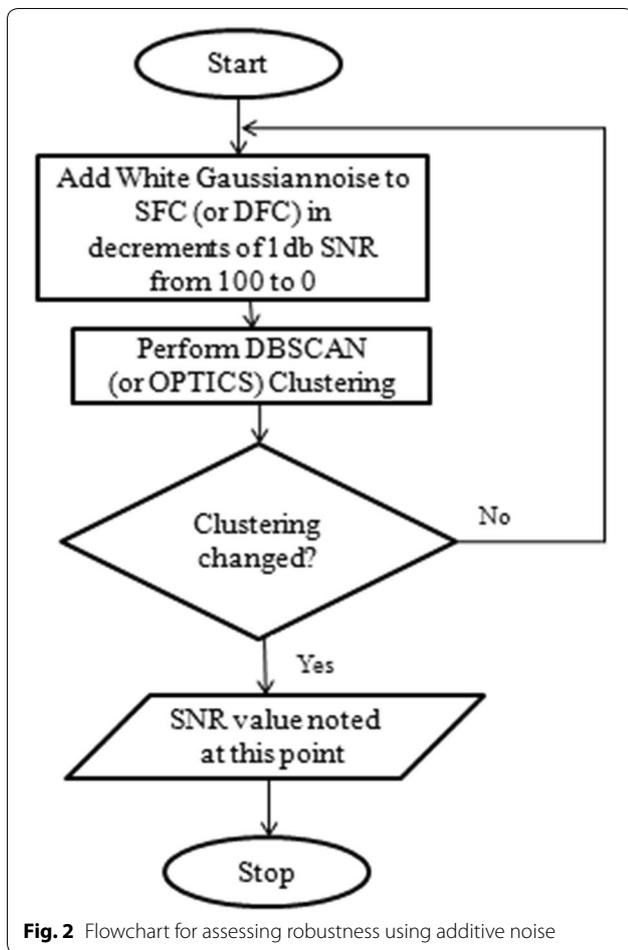
Clustering could be defined as robust if it can maintain its performance in the presence of higher noise levels compared to noiseless data. Thus, robustness is a measure of how well the clustering would be performed if lower quality data were given. A clustering method providing higher cluster purity does not necessarily imply that it is more robust to noisy variations in the data. It is possible for a clustering method to provide high performance yet be more sensitive to outliers and noise.

To assess the robustness of DBSCAN and OPTICS, each in association with SFC and DFC input features, we devised a novel approach called Recursive CLustering using Additive Noise (R-CLAN). Here, white Gaussian noise was successively added to SFC (or DFC) features such that the SNR value of the subsequent corrupted features decreased by 1 dB per iteration, with SNR values starting from 100 dB in the first iteration. In this context, the traditional definition of SNR was used, that is,  $SNR = 10 \times \log(S/N)$ , where  $S$  was the signal power (mean squared value of all the original SFC [or DFC] values),  $N$  was the noise power (mean squared value of additive Gaussian noise), and the logarithm being taken to the base of 10. An SNR value of 100 dB indicates a high amount of signal relative to noise (higher quality data), whereas a value of 0 dB indicates an equal amount of noise and signal in the data, making it difficult to distinguish signal from noise (lower quality data).

Starting with data having an SNR of 100 dB, we performed DBSCAN and OPTICS clustering at each SNR value, decreasing the SNR by 1 dB in each iteration and terminating only when the clustering structure had changed (i.e. the cluster purity value changed from the original at 100 dB). That means, we stopped at that SNR value when at least one subject in one of the groups was clustered into another group, thus changing the groups' structure. At this point, the terminating SNR value was taken as a measure of robustness. The lower the SNR value, the more noise that was present in the data, thus more robust the algorithm would be in combination with the input features (SFC or DFC), since the low SNR indicates a higher tolerance for noise. Figure 2 provides the flowchart for this procedure. R-CLAN and separation-index procedures complement one another and were used to measure robustness objectively.

## 3 Results

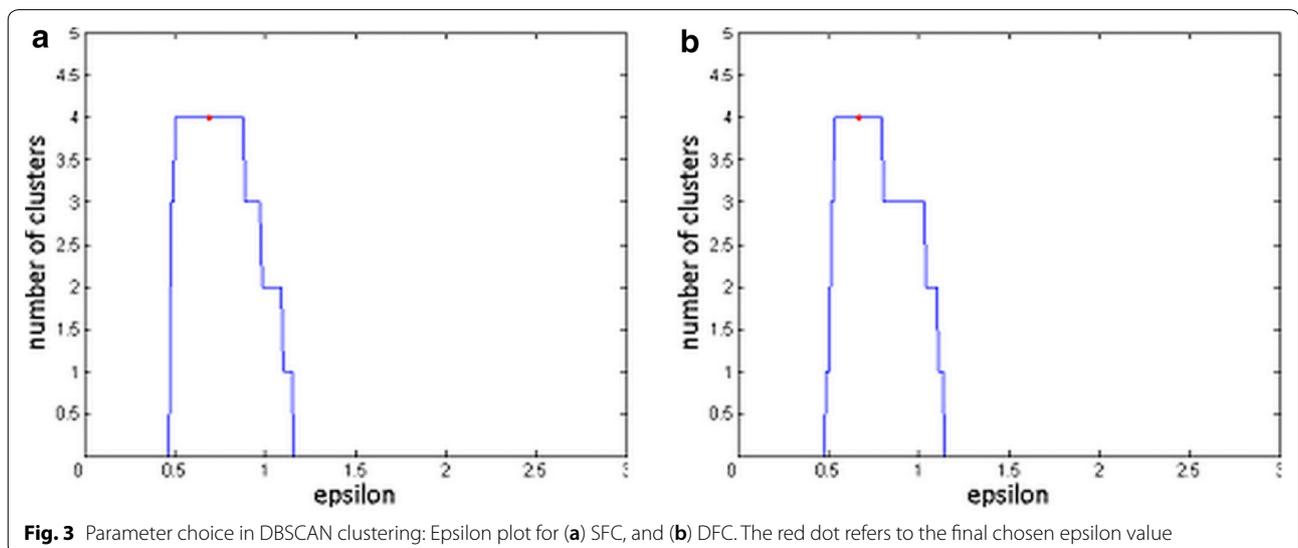
We first present the findings from DBSCAN and OPTICS clustering, before reporting the results from the assessment of robustness using our R-CLAN and separation index techniques. Figure 3a, b shows the epsilon plots



generated using DBSCAN with SFC and DFC features, respectively. With both plots, our method of choosing  $\epsilon$  resulted in DBSCAN detecting a total of four clusters, which was the expected number, given that we had four diagnostic groups.

Clustering performances are presented in Table 2. With DBSCAN clustering, the average group-wise cluster purity with SFC and DFC features were 75% and 87.88%, respectively; while OPTICS clustered with 93.18% cluster purity using SFC and 95.46% using DFC features. From this, it is clear that, (i) OPTICS performed better than DBSCAN, (ii) DFC features resulted in higher performance than SFC features, (iii) OPTICS clustering using DFC features resulted in the overall best performance, and (iv) the performance with control and AD groups were higher than that with the intermediate EMCI and LMCI groups.

The execution times for a single instance of DBSCAN and OPTICS algorithms on our computer (Intel® Xeon® quad-core processor, 3.5 GHz) were 0.0206 s and 0.0204 s, respectively, which is impressive. These numbers are interesting, because the developers of OPTICS (Ankerst et al. [38]) reported that OPTICS was consistently 1.6 times slower than DBSCAN. However, their observations may not be applicable to today’s computers since they assessed it in their computer 20 years ago. It should be noted that the time complexity of both algorithms is dependent on  $\epsilon$  and the underlying data structure, and the general runtime complexity is the same for both:  $O(n*\log(n))$ . Hence, to determine any consistent differences in execution time, the two algorithms must be repeatedly run on different datasets.



**Table 2 Success rate of clustering for each group and each feature, for both DBSCAN and OPTICS**

	Success rate of clustering				Row-wise average %
	DBSCAN		OPTICS		
	SFC %	DFC %	SFC %	DFC %	
Control	80	97.14	97.14	100	<i>93.57</i>
EMCI	73.53	79.41	100	91.18	<i>86.03</i>
LMCI	64.71	82.35	82.35	91.18	<i>80.15</i>
AD	82.76	93.10	93.10	100	<i>92.24</i>
Mean	<i>75</i>	<i>87.88</i>	<i>93.18</i>	<i>95.46</i>	

Row-wise averages (last column) and column-wise averages (last row), shown in italics, provide summary statistics

**Table 3 SNR values obtained as a measure of robustness**

	SFC	DFC
DBSCAN	55	46
OPTICS	34	21

Lower value indicates better performance

**3.1 Assessing robustness using additive noise**

As noted earlier, higher cluster purity does not necessarily imply its robustness to noise or outliers. We devised a novel technique to assess the robustness of clustering (R-CLAN). A lower SNR value identified by R-CLAN indicates that the clustering method is more robust, because it would have delivered the same clustering performance even at higher noise levels. We applied this technique to clustering of both SFC and DFC features using both DBSCAN and OPTICS (see Table 3). Clearly, OPTICS clustering was more robust to noise than DBSCAN since lower SNR values are indicative of higher robustness. We obtained higher robustness with both SFC and DFC features using OPTICS, indicating that OPTICS could deliver the same level of clustering performance even at higher noise levels. OPTICS could withstand noise levels of 125–315 times (or 21–25 dB) more than what DBSCAN could. It was also observed that DFC features were more robust to noise compared to SFC features, and the best performance was delivered by OPTICS using DFC features, with an SNR of 21 dB. DFC features could withstand about 20 times (or 13 dB) more noise than SFC features using OPTICS. With these observations, we concluded that OPTICS is a more robust clustering technique than DBSCAN, and that DFC features could result in more robust clustering performance than SFC features. These findings also corroborate with clustering results presented earlier, wherein OPTICS and DFC resulted in better cluster purities than DBSCAN and SFC. Our findings attribute both superior performance and higher robustness to OPTICS clustering using DFC features.

**Table 4 Separation index as a measure of OPTICS robustness**

	Group-wise value of separation index				Mean
	Control	EMCI	LMCI	AD	
SFC	3.2914	3.6381	3.1946	3.2487	3.3432
DFC	4.4227	3.8152	3.4321	4.1152	3.9463

Higher value indicates better performance

**3.2 Separation index to evaluate robustness of reachability plot in OPTICS clustering**

The reachability plot determines the quality of OPTICS clustering, and hence we devised a measure called the separation index to evaluate the robustness of the reachability plot. A higher value indicates superior robustness. Upon evaluating the measure with both SFC and DFC features (see Table 4), we found that DFC features provided a higher value of separation index, which in turn indicates that DFC, compared to SFC features, is a more robust measure for performing clustering using OPTICS. This result corroborates with the result obtained from R-CLAN regarding robustness using additive noise, wherein the DFC features resulted in higher robustness compared to the SFC features. These findings further reiterate that OPTICS is a better clustering algorithm than DBSCAN for unsupervised clustering of fMRI connectivity features obtained from subjects with cognitive impairment, and that DFC features would result in better clustering performance than SFC features.

**4 Discussion**

In this work, density-based clustering was performed on static and dynamic functional connectivity features obtained from fMRI data of subjects with cognitive impairment. The data consisted of subjects with early and late mild cognitive impairment, Alzheimer’s disease and matched aged healthy controls. DBSCAN and OPTICS clustering techniques were applied to SFC as well as DFC features obtained from fMRI data. Since the input data came from four diagnostic categories, we expected to identify four clusters using DBSCAN and OPTICS, although the algorithms were not biased with this information a priori. Upon clustering, we found that both DBSCAN and OPTICS resulted in four clusters, obtained in a blind unsupervised manner. This shows that these techniques were able to detect inherent disease clusters from neuroimaging data without any supervision.

We found that higher cluster purities were obtained with OPTICS clustering compared to DBSCAN, and that DFC features always resulted in superior clustering performance compared to SFC features. Furthermore, the same trend was observed with the measures

of robustness and separation index. OPTICS clustering combined with DFC features resulted in highest performance as well as most robustness to noise. Previous work by Jia et al. [46] reported a similar superior performance of DFC over SFC, wherein they determined that DFC features explained significantly more variance in human behavior than SFC. Another report by Jin et al. [47] found that DFC features have better ability in predicting psychiatric disorders (such as post-traumatic stress disorder) compared to SFC features. Our findings corroborate with these previous reports.

Referring to Table 2, the average clustering performance was found to be higher in control (93.57%) and AD groups (92.24%), as compared to the intermediate EMCI (86.03%) and LMCI (80.15%) groups. This indicates a rather expected trend wherein the control (completely healthy) and AD (completely ill) groups are on the extreme, resulting in less outliers, as compared to EMCI and LMCI groups that form the mid-band of the spectrum, resulting in more outliers and lower performance.

A novel technique using additive noise (called R-CLAN) was devised by us to assess the robustness of DBSCAN and OPTICS clustering. Table 3 shows that OPTICS clustering was more robust to noise than DBSCAN for both SFC and DFC feature sets, demonstrating that OPTICS would be able to withstand higher noise levels than DBSCAN. In addition, clustering with DFC features resulted in superior clustering performance as well as superior robustness to noise than clustering with SFC features. Static connectivity is popular in neuroimaging research and is extensively used, while dynamic connectivity is a newer technique that has gained traction more recently [42]. Both SFC and DFC provide characteristically distinct information. While the former provides the strength of connectivity over the entire fMRI scan, the latter provides the temporal variability of connectivity (i.e. change in connectivity over time). Recent studies have demonstrated the unique and superior properties of DFC over SFC [46, 55]. A comparison of SFC and DFC in their ability to identify hidden structures in the data is one of the novel contributions of our work, wherein DFC was found to be better and more robust than SFC. This is important, because the research community still largely prefers to not look beyond static connectivity. We hope our findings encourage researchers to incorporate dynamic connectivity analysis among their research strategies.

Traditionally, supervised classifiers are termed ‘robust’ when the testing error is close to the training error—that is, the classifier is able to perform well even on different datasets on which it is previously not trained. Our definition of robustness works in a similar manner: by adding noise to the underlying data, our form of robustness

is a measure of how much the quality of data, evaluated by SNR, must change before the clustering changes the subject classes. The ‘different’ data here is not a completely new set of data, but the same data changed by noise. Besides the input parameters, feature selection and other algorithmic characteristics, an unsupervised learning algorithm would be sensitive to outliers in the data as well as underlying data quality. Since OPTICS and DBSCAN are already designed to detect outliers [37, 38], our robustness measures are a novel way of evaluating their sensitivity to data quality. According to data mining literature, this is the difference between robustness to class noise (identifying outliers [56]), and to attribute noise (errors in the feature values themselves [57]). Attribute noise tends to be more difficult to detect and eliminate than class noise [58]; hence, it is important that clustering is able to work well in the presence of high attribute noise, as it is common in fMRI data even after pre-processing and noise reduction [39]. To the best of our knowledge, little research has been done in determining the level of attribute noise in fMRI data that either clustering or a supervised classifier can withstand before losing accuracy. Using our robustness measures, results indicate that OPTICS and DFC features are the best combination of clustering method and input features, respectively, to use in the presence of attribute noise, as compared to DBSCAN and SFC features.

During our evaluation of robustness, we modeled the noise present in SFC and DFC features using additive white Gaussian noise. It is notable that modeling noise in fMRI with a white Gaussian distribution is well documented [40, 41], supported by the fact that non-biological noise comes from various independent sources [40], which tend to resemble white noise. In addition, Chen and Tyler [40] showed that the power spectrum of non-biological noise approached the flat line characteristic of a white noise spectrum, and hence could be well-approximated by white Gaussian noise; however, biological noise was non-Gaussian. Other works indicate that the Gaussian noise assumption is appropriate for data with high SNR values, but noise approaches a Rician distribution at lower SNR values [39, 59, 60]. Thus, there may be a weakness in our current model of noise since it does not consider non-white sources of noise that can corrupt features. When developing a noise reduction method on fMRI data obtained from Alzheimer’s disease patients, Garg et al. [41] used additive white Gaussian noise and Rician noise separately to model different levels of noise in artificial and real-world data [41]. Likewise, future work could compare additive Rician noise and white Gaussian noise within the R-CLAN framework. The response to clustering

with various levels of different types of attribute noise could be investigated as well.

The idea of measuring noise robustness of clustering techniques is not alien to the literature. For example, in the original work that presented OPTICS, a steepness value was used to determine at which point clusters begin and end [38]. While it worked well in theory, its success was found to depend highly on the steepness value, an input parameter with no standard way of determining its appropriate value. In another method [61], similar to the peaks we used in finding the separation index, all significant local maxima in the plot were found, and significance was used to distinguish deep valleys representing well-formed clusters from shallow regions that were simply noise. The separation index is similar to these aforementioned methods, in that it is a measure of how well defined each cluster is with respect to the noise points surrounding it. However, it is a much simpler method and requires no human inputs. It could also be used in conjunction with the previous methods and other cluster extracting methods to determine how well-separated a specific cluster is from the points that form its boundaries on the reachability plot. In addition, separation index could be used as a measure of relative density of each cluster, since higher separation indices usually correspond to deeper valleys, which in turn indicates high-density regions within the dataset.

It is interesting to note that DBSCAN and OPTICS were found to be just as competitive in classifying Alzheimer's disease patients as traditional supervised learning classifiers [62]. These two density-based clustering algorithms automatically detect outliers as part of their function, leaving us able to test for robustness to data quality—something that is difficult to do with supervised classifiers since outlier noise is more harmful to their performance [58]. In addition, unsupervised algorithms such as DBSCAN and OPTICS can be used in applications beyond just classification. They can also be used to determine hidden disease states or sub-states that are often undetected in traditional diagnostic classification, future prediction of diagnostic status of new subjects based on current clusters [63] and hypothesis generation and testing, to name a few. For example, with cerebrospinal fluid data, structural MRI and FDG-PET scans as features, an earlier study used hierarchical clustering on healthy controls to identify subgroups within these subjects that could later be susceptible to Alzheimer's disease [64]. However, the number of clusters had to be chosen through visual assessment prior to clustering. Our results indicate that similar experiments could be performed using density-based clustering methods that require few input parameters, and no requirement to provide the number of clusters.

We used static and dynamic functional connectivity in this work. The brain networks obtained from them can also be used in a graph-theoretic framework and future studies could attempt clustering of network properties to obtain newer insights.

Supervised algorithms such as support vector machine (SVM) are popular in brain imaging. Backed by results of this study, we encourage researchers to consider density-based clustering methods in subject grouping and classification. Future work in this area could involve performing these analyses on various neurological and psychiatric disease conditions such as epilepsy, attention deficit hyperactivity disorder (ADHD), schizophrenia, autism spectrum disorder (ASD), post-traumatic stress disorder (PTSD), etc. to determine if disease classification findings are consistently good across various clinical populations.

Our main contributions are as follows: (i) we demonstrated that unsupervised learning, which is not prone to overfitting, could be used to determine inherent disease clusters that provide performances comparable to supervised learning. (ii) We demonstrated that density-based clustering methods, which require minimal input parameters, perform satisfactorily, and that the OPTICS technique is the suitable choice for fMRI connectivity data. (iii) We proposed two novel robustness assessment techniques and demonstrated that OPTICS clustering is a superior and noise-robust technique for fMRI connectivity data. (iv) For the first time in the literature, we assessed and compared the ability of static and dynamic connectivity features in identifying inherent disease clusters and found dynamic connectivity features to be superior and more robust. This is a significant finding given that the research community is still largely inclined towards the continued use of static connectivity alone.

## 5 Conclusion

Unsupervised learning algorithms present some key theoretical advantages compared to supervised learning algorithms (e.g. a priori diagnostic labeling is not required, and they possess the potential to detect unknown data structures). Our work presented and contrasted two relatively new methods to perform unsupervised clustering on fMRI data in a relatively large clinical dataset. We briefly described and assessed the performance of two density-based clustering algorithms: DBSCAN and OPTICS. These algorithms were used to cluster three stages of cognitive impairment (EMCI, LMCI and AD) and matched healthy controls using static and dynamic functional connectivity features. DBSCAN was found to be relatively more sensitive to noise and less precise, whereas OPTICS accurately identified all four groups and was more robust to noise as measured from

our proposed R-CLAN and separation index robustness measures. OPTICS clustering using DFC features was found to be more dependable than DBSCAN and SFC features. With the superiority of DFC and OPTICS, we encourage researchers to incorporate dynamic connectivity analysis among their research strategies and hope that we have motivated the community sufficiently to consider employing OPTICS clustering for subject classification and identifying hidden disease clusters.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s40708-020-00120-2>.

**Additional file 1.** Supplemental information.

## Abbreviations

AD: Alzheimer's disease; ADNI: Alzheimer's disease neuroimaging initiative; DBSCAN: Density-based spatial clustering of applications with noise; DFC: Dynamic functional connectivity; EMCI: Early mild cognitive impairment; FC: Functional connectivity; fMRI: Functional magnetic resonance imaging; LMCI: Late mild cognitive impairment; MCI: Mild cognitive impairment; OPTICS: Ordering points to identify the clustering structure; R-CLAN: Recursive-clustering using additive-noise; rs-fMRI: Resting-state functional magnetic resonance imaging; SFC: Static functional connectivity; SNR: Signal-to-noise ratio; SVM: Support vector machine.

## Acknowledgements

Data used in this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). Investigators within ADNI contributed to design and implementation of ADNI and provided data but did not participate in analysis or writing of this report. Complete listing of ADNI investigators: [https://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf). Data collection and sharing for this work was funded by 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.; 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 NIH ([www.fnih.org](http://www.fnih.org)). The grantee is the Northern California Institute for Research and Education, and the study is coordinated by Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. The authors thank Katherine Buchanan, Auburn University for helping in language editing.

## Authors' contributions

DR, NDD and GD were involved in the conception of this work. DR, TO and GD analyzed and interpreted the data. All authors contributed to manuscript writing. All authors read and approved the final manuscript.

## Funding

The authors acknowledge support from the Auburn University MRI Research Center. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Availability of data and materials

The datasets used in this work were taken from the publicly available ADNI database, which is available for download from their website (<https://www.loni.ucla.edu/ADNI>).

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup> Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA, USA. <sup>2</sup> Department of Radiology, Harvard Medical School, Boston, MA, USA. <sup>3</sup> Division of Health Sciences and Technology, Harvard University and Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>4</sup> Division of Engineering Science, Faculty of Applied Science & Engineering, University of Toronto, Toronto, ON, Canada. <sup>5</sup> Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India. <sup>6</sup> AU MRI Research Center, Department of Electrical and Computer Engineering, Auburn University, 560 Devall Dr, Suite 266D, Auburn, AL 36849, USA. <sup>7</sup> Department of Psychological Sciences, Auburn University, Auburn, AL, USA. <sup>8</sup> Alabama Advanced Imaging Consortium, University of Alabama Birmingham, Alabama, USA. <sup>9</sup> Center for Health Ecology and Equity Research, Auburn University, Auburn, AL, USA. <sup>10</sup> Center for Neuroscience, Auburn University, Auburn, AL, USA. <sup>11</sup> School of Psychology, Capital Normal University, Beijing, China. <sup>12</sup> Key Laboratory for Learning and Cognition, Capital Normal University, Beijing, China. <sup>13</sup> Department of Psychiatry, National Institute of Mental Health and Neurosciences, Bangalore, India.

Received: 16 October 2020 Accepted: 29 October 2020

Published online: 26 November 2020

## References

- Wang Y, Chattaraman V, Kim HJ, Deshpande G (2015) Predicting purchase decisions based on spatio-temporal functional MRI features using machine learning. *IEEE Transact Automom Mental Devel* 7(3):248–255
- Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X, Just M, Newman S (2004) Learning to Decode Cognitive States from Brain Images. *Mach Learn* 57(1–2):145–175
- Clark IA, Niehaus KE, Duff EP, Di Simplicio MC, Clifford GD, Smith SM, Mackay CE, Woolrich MW, Holmes EA (2014) First steps in using machine learning on fMRI data to predict intrusive memories of traumatic film footage. *Behav Res Ther* 62:37–46
- Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cognit Sci* 10(9):424–430
- Demirci O, Clark VP, Magnotta VA, Andreasen NC, Lauriello J, Kiehl KA, Pearson GD, Calhoun VD (2008) A review of challenges in the use of fMRI for disease classification / characterization and a projection pursuit application from Multi-site fMRI schizophrenia study. *Brain Imag Behav* 2(3):147–226
- Davatzikos C, Ruparel K, Fan Y, Shen DG, Acharyya M, Loughhead JW, Gur RC, Langleben DD (2005) Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage* 28(3):663–668
- Deshpande G, Libero L, Sreenivasan KR, Deshpande H, Kana RK (2013) Identification of neural connectivity signatures of autism using machine learning. *Front Hum Neurosci* 7:670
- Deshpande G, Wang P, Rangaprakash D, Wilamowski B (2015) Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data. *IEEE Transact Cybernet* 45(12):2668–2679
- Libero LE, DeRamus TP, Lahti AC, Deshpande G, Kana RK (2015) Multi-modal neuroimaging based classification of Autism Spectrum Disorder using anatomical, neurochemical and white matter correlates. *Cortex* 66:46–59

10. Ghebrea S, Smeulders A, Pieter A (2008) Predicting brain states from fMRI data: incremental functional principal component regression," Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, pp. 537–544.
11. Varoquaux G, Thirion B (2014) How machine learning is shaping cognitive neuroimaging. *Giga Sci* 3:28
12. Venkataraman A, Whitford TJ, Westin C-F, Golland P, Kubicki M (2012) Whole brain resting state functional connectivity abnormalities in schizophrenia. *Schizophr Res* 139(1–3):7–12
13. Michel V, Gramfort A, Varoquaux G, Eger E, Keribin C, Thirion B (2012) A supervised clustering approach for fMRI-based inference of brain states. *Pattern Recogn* 45(6):2041–2049
14. Wei L, HuaFu C, Qin Y, Xu L (2008) Analysis of fMRI data using improved self-organizing mapping and spatio-temporal metric hierarchical clustering. *IEEE Trans Med Imag* 27(10):1472–1483
15. Wang Y, Li T-Q (2013) Analysis of Whole-Brain Resting-State fMRI Data Using Hierarchical Clustering Approach. *PLoS ONE* 8(10):e76315
16. Heller R, Stanley D, Yekutieli D, Rubin N, Benjamini Y (2006) Cluster-based analysis of FMRI data. *Neuroimage* 33(2):599–608
17. Liu D, Lu W, Zhong N (2010), Clustering of fMRI Data Using Affinity Propagation. International Conference on Brain Informatics, pp. 399–406.
18. S. Rani, G. Sikka (2012) Recent techniques of clustering of time series data: a survey. *Int J Comput Appl.* 52(15).
19. Katwal SB, Gore JC, Marois R, Rogers BP (2013) Unsupervised spatiotemporal analysis of FMRI data using graph-based visualizations of self-organizing maps. *IEEE Transact Bio-Med Engin* 60(9):2472–2483
20. Warren Liao T (2005) Clustering of time series data—a survey. *Pattern Recognit.* 38(11):1857–1874
21. Venkataraman A, Van Dijk KRA, Buckner RL, and Golland P (2009) Exploring functional connectivity in fmri via clustering," In Proc. ICASSP: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 441–444.
22. Yul Byun H, Lu JJ, Mayberg HS, Günay C (2014) Classification of resting state fMRI datasets using dynamic network clusters, AAI Workshop: Modern Artificial Intelligence for Health Analytics.
23. Plant C, Teipel SJ, Oswald A, Böhm C, Meindl T, Mourao-Miranda J, Bokde AW, Hampel H, Ewers M (2010) Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *Neuroimage* 50(1):162–174
24. Honorio J (2014) Classification on brain functional magnetic resonance imaging: dimensionality, sample size, subject variability and noise. *Front Med Imag.* 153–165.
25. Jiang T, Navab N, Plum J, Vieregger M, Janoos F, Machiraju R, Sammet S, Knopp MV, Mórocz IÁ (2010) Unsupervised learning of brain states from fMRI data, medical image computing and computer-assisted intervention—MICCAI 2010 Lecture Notes in Computer Science. Springer Berlin Heidelberg, Heidelberg. pp. 201–208
26. Sato JR, Rondina JM, Mourão-Miranda J (2012) Measuring abnormal brains: building normative rules in neuroimaging using one-class support vector machines. *Front Neurosci* 6:178
27. Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 45(1):S199–S209
28. Asano T, Bhattacharya B, Keil M, Yao F (1988) Clustering algorithms based on minimum and maximum spanning trees," Proceedings of the fourth annual symposium on Computational geometry, Illinois: Urbana-Champaign, 252–257.
29. Kettnering JR (2006) The Practice of Cluster Analysis. *J Classif* 23(1):3–30
30. Raczynski L, Wozniak K, Rubel T, Zaremba K (2010) Application of density based clustering to microarray data analysis. *Int J Electron Telecommunicat* 56(3):281
31. Maqbool O, Babri HA (2007) Hierarchical Clustering for Software Architecture Recovery. *IEEE Trans Software Eng* 33(11):759–780
32. Antani S, Kasturi R, Jain R (2002) A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. *Pattern Recogn* 35(4):945–965
33. Sakai T, Tamura K (2015) Real-time analysis application for identifying bursty local areas related to emergency topics. *Springer Plus* 4:162
34. Ashtawry HM, Mahapatra NR (2015) A comparative assessment of predictive accuracies of conventional and machine learning scoring functions for protein-ligand binding affinity prediction. *IEEE/ACM Trans Comput Biol Bioinf* 12(2):335–347
35. Xu R, Wunsch D (2009) *Partitioning Clustering, Clustering*, IEEE Press Series on Computational Intelligence. Wiley, Hoboken, pp 70–71
36. Xu R, Wunsch D (2005) Survey of clustering algorithms. *IEEE Transact Neural Netw.* 16(3):645–678
37. Ester M, Kriegel H-P, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, AAAI Press. 226–231.
38. Ankerst M, Breunig MM, Kriegel H-P, Sander J (1999) OPTICS: ordering points to identify the clustering structure. *SIGMOD Rec* 28(2):49–60
39. Lindquist MA (2008) The statistical analysis of fMRI data. *Stat Sci* 23(4):439–464
40. Onozuka M, Yen CT, Chen CC, Tyler C (2008) Spectral analysis of fMRI signal and noise, novel trends in brain science. Springer Japan, Tokyo, pp 63–76
41. Garg G, Prasad G, Coyle D (2013) Gaussian Mixture Model-based noise reduction in resting state fMRI data. *J Neurosci Methods* 215(1):71–77
42. Hutchison RM, Womelsdorf T, Allen EA, Bandettini PA, Calhoun VD, Corbetta M, Della Penna S, Duyn JH, Glover GH, Gonzalez-Castillo J, Handwerker DA, Keilholz S, Kiviniemi V, Leopold DA, de Pasquale F, Sporns O, Walter M, Chang C (2013) Dynamic functional connectivity: promise, issues, and interpretations. *Neuro Image* 80:360–378
43. Lee MH, Smyser CD, Shimony JS (2013) Resting-State fMRI: A review of methods and clinical applications. *Am J Neuroradiol* 34(10):1866–1872
44. Fox MD, Greicius M (2010) Clinical applications of resting state functional connectivity. *Front Syst Neurosci* 4:19
45. Fox MD, Raichle ME (2007) Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci* 8(9):700–711
46. Jia H, Hu X, Deshpande G (2014) Behavioral relevance of the dynamics of the functional brain connectome. *Brain Connect* 4(9):741–759
47. Jin C, Jia H, Lanka P, Rangaprakash D, Li L, Liu T, Hu X, Deshpande G (2017) Dynamic brain connectivity is a better predictor of PTSD than static connectivity. *Hum Brain Mapp* 38(9):4479–4496
48. Rangaprakash D, Dretsch MN, Venkataraman A, Katz JS, Denney TS, Deshpande G (2018) Identifying disease foci from static and dynamic effective connectivity networks: illustration in soldiers with trauma. *Hum Brain Mapp* 39(1):264–287
49. Liang P, Li Z, Deshpande G, Wang Z, Hu X, Li K (2014) Altered causal connectivity of resting state brain networks in amnesic MCI. *PLoS ONE* 9(3):e88476
50. Friston KJ, Ashburner JT, Kiebel S, Nichols TE, Penny WD (2007) *Statistical parametric mapping: the analysis of functional brain images*. Elsevier Academic Press, Amsterdam (**ISBN 9780123725608**)
51. Yan C, Zang Y (2010) DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. *Front Syst Neurosci* 4:13
52. Craddock RC, James GA, Holtzheimer PE, Hu XP, Mayberg HS (2012) A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum Brain Mapp* 33:1914–1928
53. Tench CR, Tanasescu R, Auer DP, Cottam WJ, Constantinescu CS (2014) Coordinate based meta-analysis of functional neuroimaging data using activation likelihood estimation; full width half max and group comparisons. *PLoS ONE* 9(9):e106735
54. Huang T (2009) Reckon the parameter of DBSCAN for multi-density data sets with constraints," International Conference on Artificial Intelligence and Computational Intelligence (AICI), pp.375–379.
55. Rangaprakash D, Deshpande G, Daniel TA, Goodman AM, Robinson JL, Salibi N, Katz JS, Denney TS, Dretsch MN (2017) Compromised hippocampus-striatum pathway as a potential imaging biomarker of mild-traumatic brain injury and posttraumatic stress disorder. *Hum Brain Mapp* 38(6):2843–2864
56. Brodley CE, Friedl MA (1999) Identifying mislabeled training data. *J Artif Intell Res* 11:131–167
57. Hulse JD, Khoshgoftaar TM, Huang H (2007) The pairwise attribute noise detection algorithm. *Knowl Inf Syst* 11(2):171–190
58. Zhu X, Wu X (2004) Class noise vs. attribute noise: a quantitative study of their impacts. *Artif Intell Rev* 22(3):177–210
59. Wink AM, Roerdink JBTM (2006) BOLD noise assumptions in fMRI. *Int J Biomed Imag* 2006:12014

60. Noh J, Solo V (2011) Rician distributed fMRI: asymptotic power analysis and cramer-rao lower bounds. *IEEE Trans Signal Process* 59(3):1322–1328
61. J. Sander, X. Qin, Z. Lu, N. Niu, and A. Kovarsky (2003). Automatic extraction of clusters from hierarchical clustering representations. *Proceedings of the 7th Pacific-Asia conference on Advances in knowledge discovery and data mining*, Seoul, Korea, pp. 75–87.
62. Jiang X, Zhang X, Zhu D (2014) Intrinsic functional component analysis via sparse representation on Alzheimer's disease neuroimaging initiative database. *Brain Connect* 4(8):575–586
63. Halkidi M, Batistakis Y, Vazirgiannis M (2001) On clustering validation techniques. *J Intell Informat Syst* 17(2–3):107–145
64. Nettiksimmons J, Harvey D, Brewer J, Carmichael O, DeCarli C, Jack CR Jr, Petersen R, Shaw LM, Trojanowski JQ, Weiner MW, Beckett L (2010) Subtypes based on cerebrospinal fluid and magnetic resonance imaging markers in normal elderly predict cognitive decline. *Neurobiol Aging*. 31(8):1419–1428

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---