Supplementary Material

Investigating the correspondence of clinical diagnostic grouping with underlying neurobiological and phenotypic clusters using unsupervised machine learning

Xinyu Zhao[1,2], D Rangaprakash[1,3], Bowen Yuan[1], Thomas S. Denney Jr.[1,4,5], Jeffrey S. Katz[1,4,5], Michael N. Dretsch[6,7] and Gopikrishna Deshpande[1,4,5,8*]

[1]AU MRI Research Center, Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA

[2]Quora, Inc., Mountain View, CA, USA

[3]Department of Psychiatry and Biobehavioral Sciences, University of California Los Angeles, Los Angeles, CA, USA

[4]Department of Psychology, Auburn University, Auburn, AL, USA

[5]Alabama Advanced Imaging Consortium, Auburn University and University of Alabama Birmingham, AL, USA

[6]Human Dimension Division, HQ TRADOC, Fort Eustis, VA, USA

[7]U.S. Army Aeromedical Research Laboratory, Fort Rucker, AL, USA

[8]Center for Health Ecology and Equity Research, Auburn University, AL, USA

***Correspondence to:**
Gopikrishna Deshpande, Ph.D.
560 Devall Dr., Suite 266D,
Auburn University, Auburn, AL 36849, USA
Tel.: +1 334 844-7653; Fax: +1 334 844-0214
Email: gopi@auburn.edu

Disclosures: The authors report no competing interests.

1 Data Acquisition

1.1 ADHD

Subjects were scanned on different MRI scanners (KKI: Philips 3T, Peking: SIEMENS MAGNETOM Trio 3T, NYU: SIEMENS MAGNETOM Allegra 3T) using standard T2* weighted echo-planar imaging sequence with the following parameters at resting state: For KKI, TR=2500ms, TE=30ms, FA=75˚, slice thickness=3mm, number of slices=47. For Peking, TR=2000ms, TE=30ms, FA=90˚, slice thickness=3mm, number of slices=33. For NYU, TR=2000ms, TE=15ms, FA=90˚, slice thickness=4mm, number of slices=33.

1.2 AD

Data were acquired using a Philips 3T MRI scanner. A total of 140 volumes (TR=3000ms, TE=30ms, FA=80˚, and slice thickness=3.3mm, number of slices=48) were obtained. For each subject, the first 7 volumes were discarded for signal equilibrium and to allow the participant's adaptation to the circumstances.

1.3 ASD

For Caltech, data were acquired using a 3T MAGNETOM Trio scanner with TR=2000ms, TE=30ms, FA=75˚, slice thickness=3.5mm, number of slices=34. For CMU, data were acquired using a 3T MAGNETOM Verio scanner with TR=2000ms, TE=30ms, FA=73˚, slice thickness=3.0mm, number of slices=34. For NYU, data were acquired using a 3T MAGNETOM allegra scanner with TR=2000ms, TE=15ms, FA=90˚, slice thickness=4.0mm, number of slices=33. For Pitt, data were acquired using a 3T MAGNETOM allegra scanner with TR=1500ms, TE=25ms, FA=70˚, slice thickness=4.0mm, number of slices=29. For SDSU, data were acquired using a GE 3T MR750 scanner. A total of 180 volumes (TR=2000ms, TE=30ms, FA=90˚, slice thickness=3.4mm) were obtained. For Trinity, data were acquired using a Philips 3T Achieva MRI scanner with TR=2000ms, TE=28ms, FA=90˚, slice thickness=3.5mm, number of slices=38. For UCLA, data were

acquired using a 3T MAGNETOM Trio scanner with TR=3000ms, TE=28ms, FA=90˚, slice thickness=4.0mm, number of slices=34.

1.4 PTSD and PCS

Data were acquired using a 3T MAGNETOM Verio scanner (Siemens Healthcare, Erlangen, Germany) at Auburn University using T2* weighted multiband echo-planar imaging (EPI) sequence in resting state with TR=600ms, TE=30ms, FA=55˚, multiband factor=2, voxel size=3×3×5 mm^3, 1000 volumes and two sessions (cerebellum was excluded). During scanning, participants were asked to keep their eyes open and stare at a screen with white and black cross displayed background using an Avotec projection system without thinking of anything specifically.

For all sites, the data acquisition protocol was approved by the local Institutional Review Boards (IRBs). The experiments were performed in compliance with internationally accepted ethical standards and all subjects provided informed consent.

2 Preprocessing

A standard preprocessing procedure of fMRI data was applied on different datasets, including slice timing correction, motion correction, realignment, normalization to MNI space, and regressing out nuisance variance (six head motion parameters, white matter signal, cerebrospinal fluid signal and global mean signal regression). For the ADHD dataset, the preprocessing was performed using the Athena pipeline using AFNI (Cox, 1996) and FSL (Smith et al., 2004). For other datasets, we used the DPARSF toolbox (Gan and Feng, 2010) and SPM8 package (http://www.fil.ion.ucl.ac.uk/spm).

3 Clustering Methods and Parameters Optimization

Let $\boldsymbol{Z} = \{\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_j, \cdots, \boldsymbol{Z}_N\}$ represent a set of N objects with object $\boldsymbol{Z}_j = \left(Z_{j1}, Z_{j2}, \cdots, Z_{jp}\right) \in \mathbb{R}^p$, where p equals to the number of input features. Assuming the entire dataset is partitioned into k clusters. Each cluster contains a set of objects within the range $[1, N]$, and each object $\boldsymbol{Z}_j$ belongs to

exactly one cluster (which is referred to as hard clustering). Clustering is performed in the p-dimensional feature space, where each object is a point in the p-dimensional space.

3.1 Hierarchical Clustering (HC; Agglomerative)

The main idea of hierarchical clustering (Liao et al., 2008, Liao et al., 2008, Cheng et al., 2006, Dasgupta and Long, 2005) is to build a binary tree structure of the objects that iteratively merges similar groups of objects (in terms of cluster distance) together. A general procedure of hierarchical clustering is described below:

1) Assign each object $\boldsymbol{Z}_j$ to a cluster of its own.
2) Calculate the distance between any two clusters and merge the closest pair of clusters.
3) Repeat step 2-3 until all objects $\boldsymbol{Z}_j$ are ultimately merged into one big cluster.

The result of hierarchical clustering is usually depicted as a tree-like structure, which is referred to the dendrogram (**Figure S1**). The root of the dendrogram represents the entire dataset with all objects, each leaf represents one object, and the height of the dendrogram represents the distance between each pair of clusters. Different data partitions can be obtained by cutting the dendrogram at different heights.

Note that the cluster distance can be measured in a variety of ways, referred to as linkage methods. In general, there are three linkage methods that are commonly used in hierarchical clustering, the single linkage, the complete linkage, and the average linkage. The single linkage (Andreopoulos et al., 2008) calculates the shortest distance between two clusters. It can handle non-elliptical shape of clusters, but can be affected by noise and outliers. The complete linkage (Andreopoulos et al., 2008) calculates the longest distance, which is less sensitive to noise and outliers but tends to break large clusters. The average linkage (Andreopoulos et al., 2008) calculates the mean distance, which is a

compromise between single linkage and complete linkage. In this study, the average linkage was employed to measure the distance between different clusters.

3.2 Ordering Points to Identify the Clustering Structure (OPTICS)

OPTICS (Ankerst et al., 1999) is one of the most popular density-based clustering methods (Kriegel et al., 2011). Similar to hierarchical clustering, it is based on grouping objects within certain distance thresholds (ε). However, in OPTICS each cluster must contain at least a minimum number of objects ($MinPts$), which is not the case in hierarchical clustering. OPTICS can discover clusters with arbitrary shapes. It can also detect independent "noise" objects that do not belong to any of the clusters.

For each object $\boldsymbol{Z}_j$ in the feature space, there are two critical values: core-distance and reachability-distance. Let $N_\varepsilon(\boldsymbol{Z}_j)$ represent the number of nearby neighborhood objects within ε distance (called $\varepsilon - neighborhood$), and $MinPts - distance(\boldsymbol{Z}_j)$ represent the distance from $\boldsymbol{Z}_j$ to its $MinPts$' neighbor. An object $\boldsymbol{Z}_j$ is considered as a core-object if at least $MinPts$ objects are found with its $\varepsilon - neighborhood$. The core distance of $\boldsymbol{Z}_j$ is defined as:

$$core - dist_{\varepsilon,MinPts}(\boldsymbol{Z}_j) = \begin{cases} Undefined & if\ N_\varepsilon(\boldsymbol{Z}_j) < MinPts \\ MinPts - distance(\boldsymbol{Z}_j) & otherwise \end{cases} \quad (1)$$

Which is the smallest distance between $\boldsymbol{Z}_j$ and an object in its $\varepsilon - neighborhood$ such that $\boldsymbol{Z}_j$ could become a core-object. The reachability-distance of $\boldsymbol{Z}_i$ with respect to $\boldsymbol{Z}_j$is then defined as:

$$reachability - dist_{\varepsilon,MinPts}(\boldsymbol{Z}_i, \boldsymbol{Z}_j)$$

$$= \begin{cases} Undefined & if\ N_{\varepsilon}(\boldsymbol{Z}_j) < MinPts \\ max(core - distance(\boldsymbol{Z}_j), dist(\boldsymbol{Z}_i, \boldsymbol{Z}_j)) & otherwise \end{cases} \quad (2)$$

Where $dist\ (\boldsymbol{Z}_i, \boldsymbol{Z}_j)$ is the Euclidean distance between $\boldsymbol{Z}_i$ and $\boldsymbol{Z}_j$.

The complete procedure of OPTICS is described below:

1) Choose one object $\boldsymbol{Z}_j$ arbitrarily.
2) Retrieve the $\varepsilon - neighborhood$ of $\boldsymbol{Z}_j$, determine the core-distance, and set the reachability-distance to undefined.
3) If $\boldsymbol{Z}_j$ is not a core-object, go to step 5. Otherwise, go to step 4.
4) For each object $\boldsymbol{Z}_i$ in the $\varepsilon - neighborhood$ of $\boldsymbol{Z}_j$, update its reachability-distance to $\boldsymbol{Z}_j$ and insert $\boldsymbol{Z}_i$ into an OrderSeeds list if it has not been processed yet.
5) If the input dataset is fully consumed and the OrderSeeds list is empty, go to step 6. Otherwise, move on to the next object in the OrderSeeds list (or the input list, if the OrderSeeds list is empty) and go to step 2.
6) Output core-distance, reachability-distance of each object, and processed order of objects.

The output of OPTICS is a reachability plot, which depicts the hierarchical structure of different clusters. In the reachability plot, objects are plotted in the processed order together with their respective reachability-distance. Objects belonging to the same cluster have a low reachability-distance to their nearby neighbors. Therefore, the clusters show up as valleys in the reachability plot (see **Figure S2**). The final data partition can be obtained by using a threshold on the reachability plot.

3.3 Density Peak Clustering (DPC)

DPC (Rodriguez and Laio, 2014) is a relatively new method based on the idea that cluster centers can be characterized by two criteria: (i) they have a higher density than their neighbors, and (ii) they have a relatively large distance from other points with higher densities. Like OPTICS, this method can detect clusters with arbitrary shape and spot outliers. In addition, DPC outperforms commonly used clustering methods, e.g., k-means and hierarchical clustering, when the dataset contains complicated characteristics such as narrow bridges between clusters, uneven-sized clusters, clusters with high overlap, etc.

For each object $\boldsymbol{Z}_j$, two quantities are computed: local density $\rho(\boldsymbol{Z}_j)$ and minimum distance with higher density $\delta(\boldsymbol{Z}_j)$. $\rho(\boldsymbol{Z}_j)$ is defined as:

$$\rho(\boldsymbol{Z}_j) = \sum_i \chi(dist(\boldsymbol{Z}_i, \boldsymbol{Z}_j) - d_c) \tag{3}$$

Where d_c is a cutoff distance, and $\chi(x)$ can be computed by:

$$\chi(x) = \begin{cases} 1 & if\ x < 0 \\ 0 & otherwise \end{cases} \tag{4}$$

From Eqn. **(3)** and **(4)**, it can be seen that $\rho(\boldsymbol{Z}_j)$ equals to the number of objects within d_c with respect to object $\boldsymbol{Z}_j$. $\delta(\boldsymbol{Z}_j)$ is defined as:

$$\delta(\boldsymbol{Z}_j) = \min_{i:\rho(\boldsymbol{Z}_i)>\rho(\boldsymbol{Z}_j)} dist(\boldsymbol{Z}_i, \boldsymbol{Z}_j) \tag{5}$$

For the object with highest density, $\delta(\boldsymbol{Z}_j)$ is conventionally set to,

$$\delta(\boldsymbol{Z}_j) = \max_i dist(\boldsymbol{Z}_i, \boldsymbol{Z}_j) \tag{6}$$

Note that if $\boldsymbol{Z}_j$ is local or global maxima in the density, $\delta(\boldsymbol{Z}_j)$ will be much larger than its typical nearest neighbor. Thus, objects with larger ρ and δ are considered as cluster centers. Objects with smaller ρ and larger δ are considered as outliers. Other objects are assigned to the same cluster as their nearest neighbor of higher density (see **Figure S3**).

3.4 Input Parameter Optimization

In each clustering method, there are several user-specified input parameters, which can affect clustering results significantly. For hierarchical method, the cutting height of the dendrogram needs to be specified. For OPTICS, ε can simply be set to the maximum possible value, and Ankerst and colleagues (Ankerst et al., 1999) showed that $MinPts$ values between 10 and 20 is generally a reasonable choice. However, the threshold for the reachability plot, which is used to extract clusters, still needs to be optimally determined. For DPC, d_c can be chosen based on the rule that the average number of neighbors is often around 1-2% of the total number of objects in the data set (Rodriguez and Laio, 2014). A threshold for ρ and δ needs to be defined to distinguish cluster centers and outliers. To find the optimal value of these parameters, the Calinski-Harabasz (CH) index (Calinski and Harabasz, 1974, Y. Liu et al., 2013) was applied in this work. Let C_k represent the center of cluster k, where $1 \leq k \leq K$, and C represent the center of entire data set. The CH index is then defined as:

$$CH = \frac{B/(K-1)}{W/(N-K)} \tag{7}$$

Where the between-cluster variation B is computed by,

$$B = \sum_{k=1}^{K} \|C_k - C\|^2 \tag{8}$$

And the within-cluster variation W is computed by:

$$W = \sum_{k=1}^{K} \sum_{\boldsymbol{Z}_j \in cluster-k} \left\|\boldsymbol{Z}_j - C_k\right\|^2 \tag{9}$$

Based on the definition of clustering, we would want to minimize W and maximize B. Thus, the optimal parameters are determined by maximizing the CH index. The optimal number of clusters can be identified, simultaneously. For example, for hierarchical clustering, we started with a relatively high cutting height for the dendrogram. In each iteration, the cutting height was reduced by a small amount and the CH index was computed and recorded based on the current data partition. The iteration continued until the cutting height was smaller than a specified baseline (e.g., the average height of the dendrogram). The optimal height was determined as the one with the largest CH index. The same iterative procedure was applied to OPTICS to determine optimal threshold of reachability plot, and to DPC to determine the optimal threshold of ρ and δ. This iterative method is referred to as "grid search" (Bergstra and Bengio, 2012), which has been commonly used for hyperparameter optimization in machine learning.

4 Feature Selection and Cluster Identification

Assuming d is the initial number of features, an exhaustive search of 2^d possible subsets is computationally intractable. Thus, three alternative methods are proposed in our framework, i.e.,

Sequential Feature Ranking (SFR), Sequential Forward Searching of Greedy Algorithm (SFS), and Genetic Algorithm (GA), to find the optimal subset of features.

4.1 Evaluation of Feature Selection and Clustering

To evaluate the result for both feature selection and clustering, a "similarity criteria" was employed that compares the similarities of diagnostic clusters (i.e. similarity of objects grouping) obtained using the following pairs of features: (i) clinical diagnostic measures and connectivity, and (ii) phenotypic measures and connectivity. And the similarity between two different clustering results are quantified using Torres' method (Torres et al., 2008). Let $C = \{C_1, C_2, \cdots, C_m\}$ and $D = \{D_1, D_2, \cdots, D_n\}$ represent objects grouping obtained from two clustering results. The similarity matrix for *C* and *D* is an $m \times n$ matrix defined as:

$$S_{C,D} = \begin{bmatrix} S_{11} & \cdots & S_{1j} & \cdots & S_{1n} \\ & & \vdots & & \\ S_{i1} & \cdots & S_{ij} & \cdots & S_{in} \\ & & \vdots & & \\ S_{m1} & \cdots & S_{mj} & \cdots & S_{mn} \end{bmatrix} \tag{10}$$

Where $S_{ij} = size_{inter}/size_{union}$, which is the Jaccard's Similarity Coefficient with $size_{inter}$ being the size of intersection and $size_{union}$ being the size of the union of cluster sets C_i and D_j. The similarity of clustering C and D is then defined as:

$$Sim(C, D) = \frac{\sum_{i \le m, j \le n} S_{ij}}{max(m, n)} \tag{11}$$

From Eqn. (10) and (11), it can be seen that $0 \le Sim\ (C, D)\ \le 1$, and $Sim\ (C, D) = 1$ when two clustering results are identical (i.e. objects are classified into the exact same groups in two clustering results).

Using the similarity criteria, the optimal feature selection along with clustering is then determined by the one resulting in the largest similarity between objects grouping obtained using following pairs of features (i) clinical diagnostic measures and connectivity, and (ii) phenotypic measures and connectivity.

4.2 Sequential Feature Ranking (SFR)

In this study, we proposed a new feature selection method by ranking all features based on statistical significance in descending order. The process started with a subset of features with only one feature (with highest statistical significance) in it. In each iteration, a feature was added to the subset from the sorted list, and the corresponding value of similarity criterion was computed. The iteration stopped when all the features were added to the subset. SFR is derived based on an intuitive thinking that statistically significant features can lead to better clustering results. However, as the statistical significance of individual feature does not necessarily guarantee cluster separation when they are combined, SFR does not guarantee a global optimum in feature selection.

4.3 Sequential Forward Searching of Greedy Algorithm (SFS)

SFS is one of the most commonly used heuristic methods in finding an optimal solution (Dy and Brodley, 2004) in different problems. Specifically, when used for feature selection, the method started with no feature in the “optimal subset” and sequentially added one feature at a time. The feature added to the optimal subset was the one that provided the largest similarity criterion value when used in combination with the features chosen previously. The iterations continued until adding more features did not increase the similarity criterion value. Although SFS converges faster than other heuristic methods, it does not guarantee a global optimum.

4.4 Genetic Algorithm (GA)

GA is also a heuristic search method inspired by stochastic evolution theory that is routinely used in generating solutions to optimization and searching problems (Yang and Honavar, 1997, Shahamat and Pouyan, 2015, Hassanzadeh and Mobayen, 2011, Hassanzadeh et al., 2008, Iraj Hassanzadeh et al., 2008, Hassanzadeh and Mobayen, 2008). GA maintains a set of candidate solutions (M_0) called "population". Each "population" represents a specific selection of a subset of features. GA is an iterative process of selecting "survival solutions". The survival solutions would be those resulting in larger values of the similarity criteria, which were generated from the crossover and mutation operations. Crossover was applied to produce new solutions by randomly combining two current solutions, and a mutation was applied on newly generated solutions with a small probability (Yang and Honavar, 1997, Shahamat and Pouyan, 2015). The iteration continued until the maximum similarity among candidate solutions did not increase any further.

In this study, an array of N_0 bits was used to represent the selected subset of features, with each bit in the array indicating the status of one specific feature, 1 indicates 'selected' and 0 indicates 'discarded'. The complete procedure of GA is described below (**Figure S4**):

1) Initialization: 400 candidate solutions were generated by randomly setting 1 or 0 for each bit.
2) Crossover: two candidate solutions *A* and *B* were randomly selected from the current population. A value *v* between 1 and N_0 was randomly selected. Then a new solution was formed by combining the feature bits 1 to *v* from *A* and feature bits *v* + 1 to N_0 from *B*.
3) Mutation: for each new generated new solution, a mutation was applied by reversing bits in the vector with a probability of 0.1. The mutation probability was determined empirically.
4) Evaluation: clustering method was applied on each candidate solution (i.e., a subset of features). Then, for each clustering result, the cluster similarity was computed between itself and reference cluster (i.e., clinical cluster or phenotypic cluster).

5) Selection: 70% of candidate solutions (i.e., 280) with higher similarity were selected and 30% candidate solutions (i.e., 120) from the rest of the solutions were randomly selected to increase the diversity of the solution. This ratio was determined by a trial-and-error method.
6) If the largest similarity did not converge, we iterated back to step-2. Otherwise, the clustering results with the largest similarity and the corresponding selected subset of features were saved as outputs.

5 References

Andreopoulos, B., An, A., Wang, X., Schroeder, M., 2008. A roadmap of clustering algorithms: finding a match for a biomedical application. Brief. Bioinform. 10, 297–314. doi:10.1093/bib/bbn058

Ankerst, M., Breunig, M.M., Kriegel, H.-P., Sander, J., 1999. Optics: Ordering points to identify the clustering structure. ACM Sigmod Rec. 49–60. doi:10.1145/304182.304187

Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. J. Mach. Learn. Res. 13, 281–305.

Calinski, T., Harabasz, J., 1974. A Dendrite Method for Cluster Analysis. Commun. Stat. - Simul. Comput. 3, 1–27. doi:10.1080/03610917408548446

Chao-Gan, Y., Yu-Feng, Z., 2010. DPARSF: A MATLAB Toolbox for "Pipeline" Data Analysis of Resting-State fMRI. Front. Syst. Neurosci. 4, 13. doi:10.3389/fnsys.2010.00013

Cheng, D., Kannan, R., Vempala, S., Wang, G., 2006. A divide-and-merge methodology for clustering. ACM Trans. Database Syst. 31, 1499–1525. doi:10.1145/1189769.1189779

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173. doi:10.1006/cbmr.1996.0014

Dasgupta, S., Long, P.M., 2005. Performance guarantees for hierarchical clustering. J. Comput. Syst. Sci. 70, 555–569. doi:10.1016/j.jcss.2004.10.006

Dy, J.G., Brodley, C.E., 2004. Feature Selection for Unsupervised Learning. J. Mach. Learn. Res. 5, 845–889.

Hassanzadeh, I., Mobayen, S., 2011. Controller design for rotary inverted pendulum system using evolutionary algorithms. Math. Probl. Eng. 2011. doi:10.1155/2011/572424

Hassanzadeh, I., Mobayen, S., 2008. PSO-based controller design for rotary inverted pendulum system. J. Appl. Sci. 8, 2907–2912. doi:10.3923/jas.2008.2907.2912

Hassanzadeh, I., Mobayen, S., Harifi, A., 2008. Input-output feedback linearization cascade controller using genetic algorithm for rotary inverted pendulum system. Am. J. Appl. Sci. 5, 1322–1328. doi:10.3844/ajassp.2008.1322.1328

Hassanzadeh, I., Mobayen, S., Sedaghat, G., 2008. Design and implementation of a controller for magnetic levitation system using genetic algorithms. J. Appl. Sci. 8, 4644–4649. doi:10.3923/jas.2008.4644.4649

Kriegel, H.P., Kröger, P., Sander, J., Zimek, A., 2011. Density-based clustering. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 1, 231–240. doi:10.1002/widm.30

Liao, W., Chen, H., Yang, Q., Lei, X., 2008. Analysis of fMRI data using improved self-organizing mapping and spatio-temporal metric hierarchical clustering. IEEE Trans. Med. Imaging 27, 1472–1483. doi:10.1109/TMI.2008.923987

Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J., Wu, S., 2013. Understanding and enhancement of internal clustering validation measures. IEEE Trans. Cybern. 43, 982–994. doi:10.1109/TSMCB.2012.2220543

Rodriguez, A., Laio, A., 2014. Machine learning. Clustering by fast search and find of density peaks. Science 344, 1492–6. doi:10.1126/science.1242072

Shahamat, H., Pouyan, A.A., 2015. Feature selection using genetic algorithm for classification of schizophrenia using fMRI data. J. Artif. Intell. Data Min. 3, 30–37.

doi:10.5829/idosi.JAIDM.2015.03.01.04

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, 208–219. doi:10.1016/j.neuroimage.2004.07.051

Torres, G., Basnet, R., Sung, A., 2008. A similarity measure for clustering and its applications. Proc. World Acad. Sci. Eng. Technol. 31, 490–496.

Yang, J., Honavar, V., 1997. Feature Subset Selection Using A Genetic Algorithm. Pattern Recognit. 13, 380. doi:10.1.1.55.5748

FIGURES

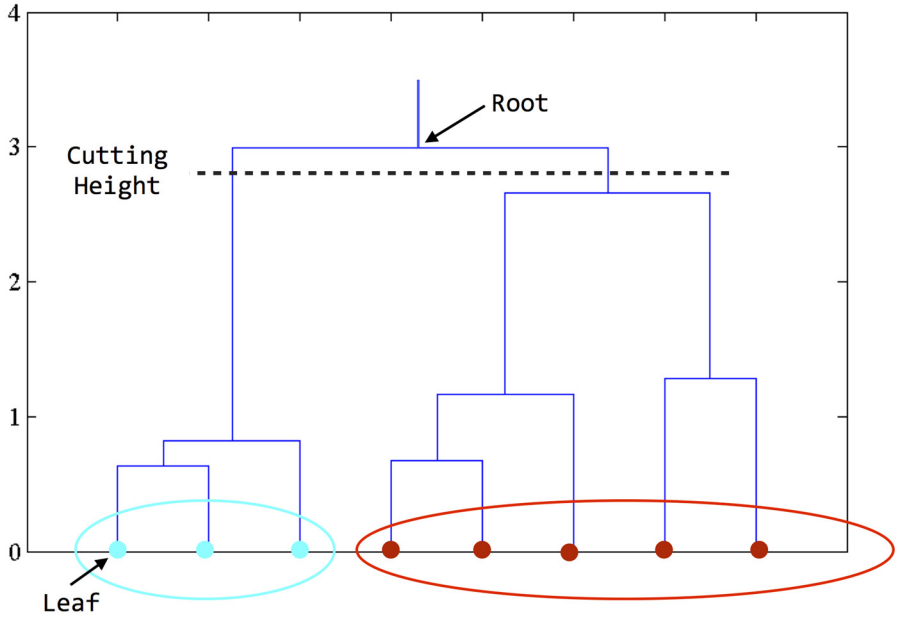

Figure S1 *Dendrogram derived from hierarchical clustering. The final clustering result is obtained by cutting the tree at defined level. The two clusters are shown by different colors (red and cyan).*

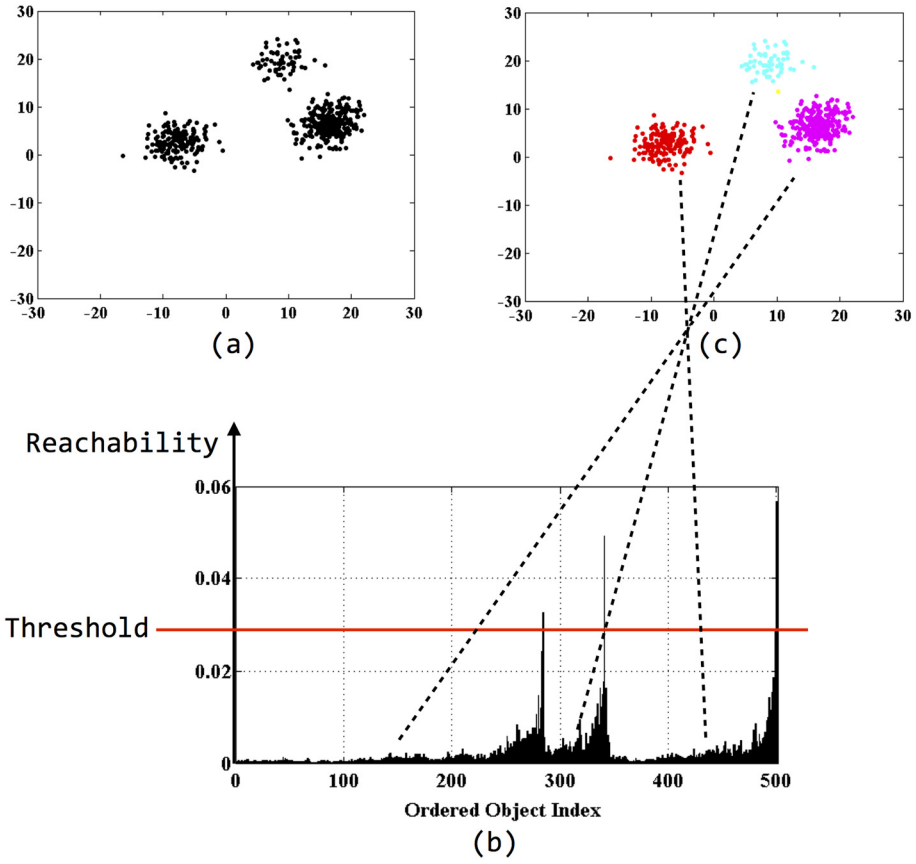

Figure S2 *(a) Simulated data (unclustered), (b) reachability plot obtained from OPTICS, and (c) clustering result. Each cluster corresponds to one valley in the reachability plot.*

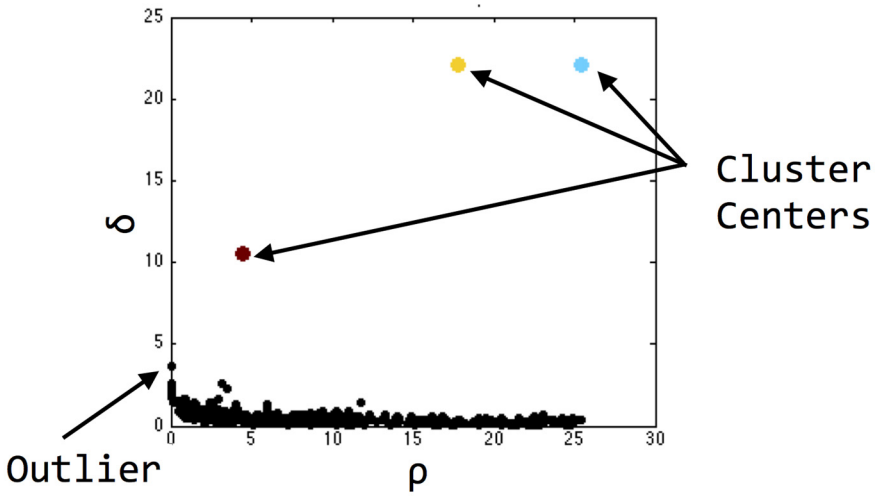

Figure S3 *Illustration of the DPC method: Plot of δ as a function of ρ for each object. Objects with larger ρ and δ are cluster centers and objects with smaller ρ, and larger δ are outliers.*

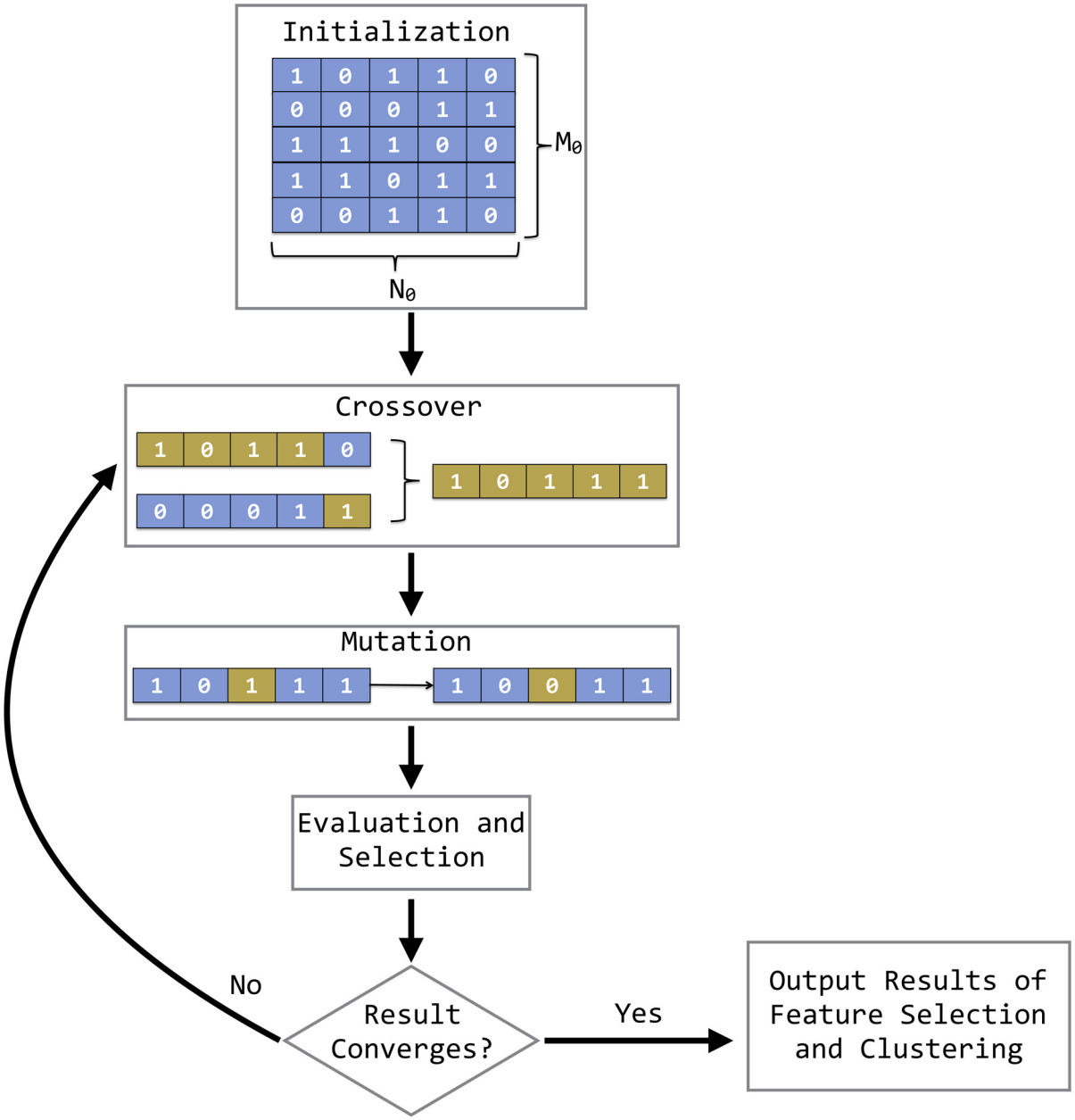

Figure S4 *Flowchart of GA for feature selection and clustering. In the M_0-by-N_0 matrix, each row represents a candidate solution, describing a subset of selected features. Each number in the row represents whether a feature is selected (1) or discarded (0).*