# Evaluating 35 Methods to Generate Structural Connectomes Using Pairwise Classification

Dmitry Petrov, Alexander Ivanov, Joshua Faskowitz, Boris Gutman, Daniel Moyer, Julio Villalon, Neda Jahanshad, Paul Thompson

github.com/USC-IGC/connectome-evaluation-miccai2017

## Summary

In this work, we study how variations in pre-processing steps of structural brain networks affect reliability and ability to distinguish subjects. We compare 35 structural connectome-building pipelines for which we vary diffusion reconstruction models, tractography algorithms and parcellations. Next, we classify structural connectome pairs as either belonging to the same individual or not and compare our results to Intraclass Correlation Coefficient (ICC). **Our main result: probabilistic tracking using either a Constrained Spherical Deconvolution local model [2] or the Constant Solid Angle method [1] gives the best combination of pairwise classification and mean ICC**.

## Pipelines and features
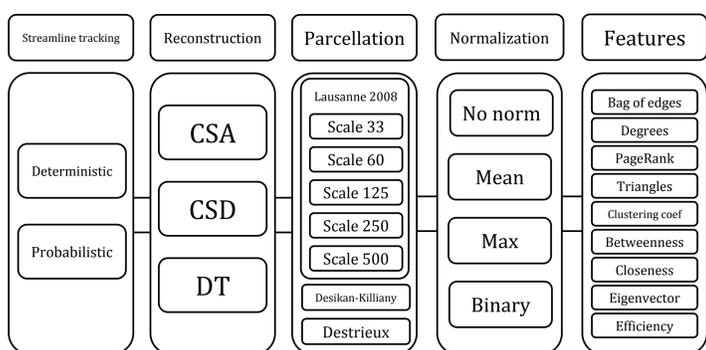


## Methods

**We propose to use pairwise classification accuracy (PACC) as a supplement to the usual ICC used in test-retest datasets.** As an additional validation of pipelines and features, we also evaluate accuracy of a sex classification. We denote a set of connectomes derived from particular pipeline as $\{C_j^i\}$, where $j$ is an index of a subject and $i$ is an index of an image.

**Pairwise classification.** For each connectome, each normalization and each set of features described in 'examine pipelines' section we make all possible pairs of connectome features $- (f(C_{j_1}^{i_1}), f(C_{j_2}^{i_2}))$. For each pair, we assign a binary target variable $- 1$ if connectomes were from the same subject ($j_1 = j_2$), $0 -$ if they were from different subjects ($j_1 \neq j_2$). Finally, for each pair we build a vector of three features, describing their difference $\|f(C_1) - f(C_2)\|$ according to $l_1$, $l_2$ and $l_\infty$ norms. Due to imbalance of classes in generated pairs, we use all samples with label 1 and equally sized random subsample of 0.

**Validation.** We use linear classifiers for pairwise and sex classification problems. We measure model performance and accuracy in a two-step validation procedure. First, for each dataset, we perform hyperparameter grid search based on a 5-fold cross-validation with a fixed random state for reproducibility. Then we evaluate the best parameters on 50 train/test splits with fixed different random states (test size was set to 20% of data). We characterize each connectome building pipeline and feature by mean pairwise/sex classification accuracy on these 50 test splits.

## Data

We used three datasets from the Consortium for Reliability and Reproducibility [3]. Summary is in the table.

| Dataset | N | Scans per subject | Age, years | Number of females | Retest period, days | DWI directions |
|---------|-----|------|-----------------|-----|-------|----|
| BNU 1   | 49  | 2    | $23.0 \pm 2.3$  | 23  | 33-55 | 30 |
| HNU 1   | 30  | 10   | $24.4 \pm 2.4$  | 15  | 3-40  | 30 |
| IPCAS 1 | 26  | 2    | $20.7 \pm 1.7$  | 19  | 5-29  | 60 |

## Results

Figure 1 shows scatter plots of PACC vs ICC depending on the reconstruction model, tractography, normalization and parcellation. **The combination of CSA/CSD reconstruction model and probabilistic tractography performs best**. Weighted degrees, number of triangles, clustering coefficient and PageRank all have scatter patterns closely mimicking that for bag of edges; likewise, a pattern similar to closeness centrality holds for betweenness centrality, eigenvector centrality and local efficiency.

Figure 2 shows the accuracy of the sex classification task in four groups: high/low PACC and high/low ICC. **Unlike their combination, neither low ICC nor low PACC alone is sufficient to predict poor sex classification accuracy**.
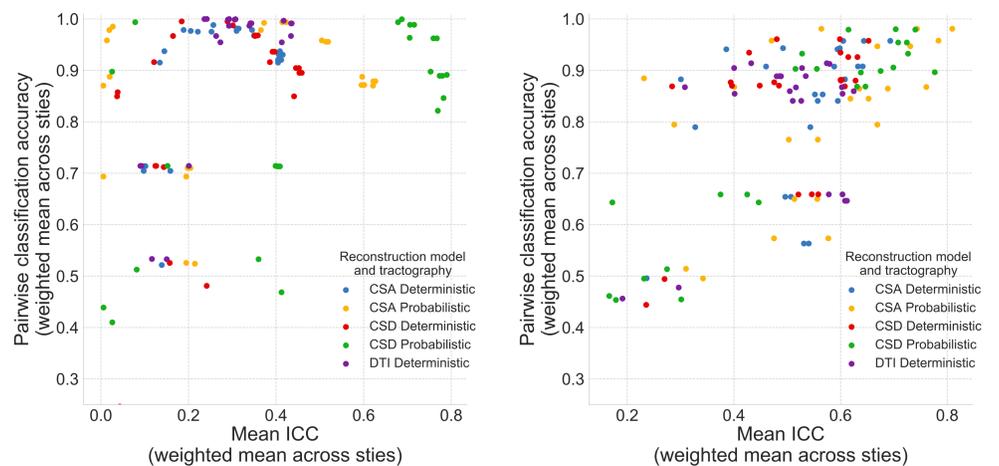


Figure 1: Scatter plots for mean feature ICC and PACC for **bag of edges (left)** and **closeness centrality (right)** depending on reconstruction model, tractography, connectome normalization and parcellation. Each point represents a weighted mean of ICC/PACC across three datasets. ICC was weighted by the number of subjects and pairwise classification by the number of scans per subject.
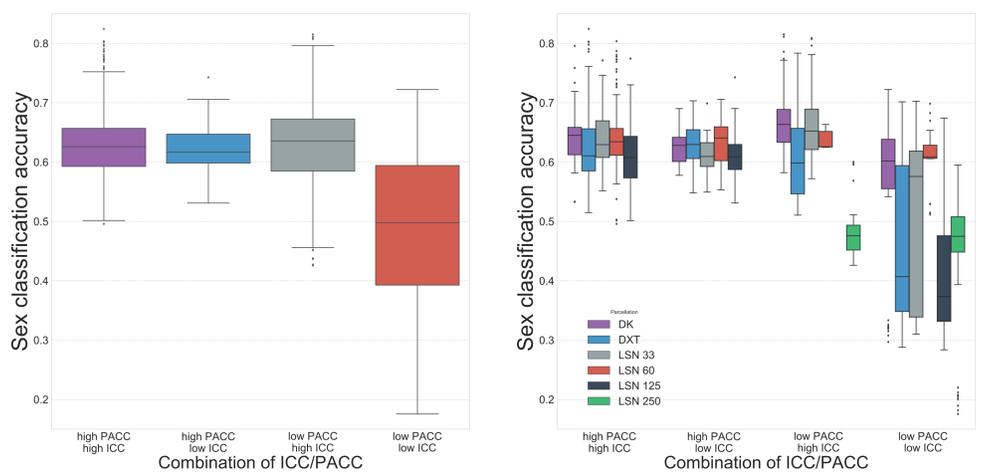


Figure 2: Distribution of sex classification accuracy depending reconstruction model in terms of high/low PACC and ICC (left), specified by parcellation (right). Each sex classification accuracy value is a weighted mean across datasets weighted by the number of subjects. PACC threshold was set at 0.9, ICC threshold was 0.6.

**Our results suggest that the method may be useful in identifying overall trends in connectome usefulness beyond simply feature-wise reliability measures, particularly with respect to DWI model choice and tractography approaches**. As well, our results appear to confirm the intuition that having both low ICC and low pairwise classification accuracy generally leads to poor performance in unrelated classification tasks. It is also notable that PACC or ICC alone are not sufficient to identify reliably features poorly suited for our test classification task. Though the results are promising, they must be viewed with some skepticism given the limited nature of our validation.

## References

[1] Iman Aganj, Christophe Lenglet, Guillermo Sapiro, Essa Yacoub, Kamil Ugurbil, and Noam Harel. Reconstruction of the orientation distribution function in single-and multiple-shell q-ball imaging within constant solid angle. *Magnetic Resonance in Medicine*, 64(2):554–566, 2010.

[2] J-Donald Tournier, Fernando Calamante, and Alan Connelly. Robust determination of the fibre orientation distribution in diffusion mri: non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, 35(4):1459–1472, 2007.

[3] Xi-Nian Zuo, Jeffrey S Anderson, Pierre Bellec, Rasmus M Birn, Bharat B Biswal, Janusch Blautzik, John CS Breitner, Randy L Buckner, Vince D Calhoun, F Xavier Castellanos, et al. An open science resource for establishing reliability and reproducibility in functional connectomics. *Scientific data*, 1, 2014.