Review of: "A Simple Preprocessing Method Enhances Machine Learning Application to EEG Data for Differential Diagnosis of Autism"

Arnab Bhattacharjee[1]

1 Bangladesh University of Engineering and Technology

Potential competing interests: No potential competing interests to declare.

The authors have applied minimum spanning tree (MST) clustering to the Manhattan Distance matrix computed from the multi-channel EEG electrode time series data. The links for each electrode have been used as features in KNN to distinguish two data classes.

1. It would be nice for the article to have mathematical details of the MST algorithm.
2. The classification results can be reported for more trials. Authors can randomly take samples from both classes multiple times and run the classification algorithm for each trial. The average result can be reported.
3. The first 19 features are electrode numbers, which, I believe, can be omitted. Authors can arrange the link values for each sample maintaining the same electrode number order and can use these 19 values for classification.
4. The article requires details of the PST algorithm. Is PST run on the 38 features constructed for classification? If not, it would be nice to see clustering algorithm performance (t-SNE/UMAP) on the features.
5. The authors need to discuss the choice of selecting the Manhattan distance over others. Classification results over MST performed with different distance metrics can be a nice experiment.
6. What is the distance metric used in KNN? What's the K value?
7. References in Table 2 are mismatched.
8. According to the training-testing protocol, ANN has been used. The article should report the ANN performance and also the details of the architecture.