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

Aleš Smrdel[1]

1 University of Ljubljana

In this article, the authors introduce an innovative approach for analyzing EEG signals to differentiate autistic children from those with different diagnoses. Their method involves computing the Manhattan distance between signals from different EEG electrodes and constructing a minimum spanning tree. Feature vectors derived from this tree are then employed for classification, yielding promising results.

However, I recommend the following modifications to enhance the clarity and completeness of the article:

1. The authors provided general subject data but omitted essential statistical information such as means and standard deviations. Including these details would improve the comprehensiveness of the presented data.
2. While the authors mention calculating the Manhattan distance, they do not specify the exact methodology used.
3. The authors state that they used the 10 minutes eyes-closed part of the signal, but it is unclear if the entire signal was used for the calculation of the distance and if any kind of preprocessing was involved (was it filtered, averaged, …) since the signals can be quite noisy and some noise-cancellation technique should be used. These details should be clarified for a more robust understanding.
4. The article mentions the use of the Manhattan distance as the measure of signal difference. It would be valuable if the authors explained the selection of this distance measure and also explored the performance of their technique when employing alternative distance metrics.
5. The authors describe creating a minimum spanning tree from the obtained values but do not elaborate on the process. Given the critical role of this step in their technique, it is imperative that they provide a detailed explanation of how the tree was constructed.
6. From the abstract and the body, it is unclear if the authors used more than one classifier (in the abstract: "The best machine learning system (KNN algorithm) ...") or several. The authors should present results also of different classifiers. Also, it is unclear how many nearest neighbours they used to obtain these results.