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

Manoj Vishwanath[1]

1 University of California, Irvine

Potential competing interests: No potential competing interests to declare.

The approach outlined in the paper is compelling due to its innovative use of EEG data analysis techniques, particularly in deriving topological features using minimum spanning trees. By applying machine learning, the method aims to distinguish between Autism Spectrum Disorder (ASD) and other Neuro-Psychiatric Disorders (NPD) with promising accuracy.
However, there are some concerns:
More detailed demographics of the subjects should be provided, including age ranges, gender distribution, and any relevant clinical characteristics.
There is a lack of comparison in terms of accuracy and processing time between the proposed approach and existing methods, which could provide valuable insights into its effectiveness and efficiency.
The abbreviation "ASD" should be defined at the beginning of the text to ensure clarity for readers unfamiliar with the term.
Numerous typographical errors are present throughout the paper, which may detract from its overall professionalism and readability.
Figure 1 could benefit from a clearer visual presentation, such as a heatmap, to enhance the understanding of the data and facilitate interpretation.
The paper should discuss the interpretability of the graphs and the relationship between connections and the research questions being addressed, rather than treating the machine learning process as a black box.
Clarification is needed regarding how the two-column data in Figure 4 is converted into one row, specifying whether it is done row-wise or column-wise to ensure reproducibility and clarity.
It would be helpful to use the term "2-fold cross-validation" instead of providing lengthy explanations for the training and testing phases, elucidating any differences in methodology or implementation.
The potential impact of varying EEG data lengths on the results should be discussed, as it may affect the accuracy and generalizability of the predictive model.