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

Kamil Bader el Dine[1]

1 Lebanese University

Potential competing interests: No potential competing interests to declare.

- Graphical Abstract: The graphical abstract serves as a concise overview of the work presented. However, it is suggested to enhance clarity by incorporating a brief summary of the data source and collection methodology prior to delving into technical details. This adjustment would provide readers with a clearer understanding of the study before delving into its intricacies.
- Machine Learning Usage: While the paper employs machine learning techniques, there is a notable absence of discussion regarding this methodology within the Methods section. It is recommended to provide a comprehensive explanation of the machine learning approaches utilized to enhance transparency and reproducibility.
- Literature Review: The paper lacks a thorough exploration of previous studies that have applied machine learning in similar contexts. Including a section discussing relevant prior research would provide valuable context and demonstrate the novelty or advancements presented in this work.
- Choice of KNN: The rationale behind selecting the KNN algorithm is not adequately clarified, nor is its comparative advantage over other machine learning methods elucidated. To enhance the comprehensibility of the methodology, it is advisable to provide a justification for the selection of KNN and discuss its suitability for the specific task at hand.
- Data Subdivision: The division of data into sub-samples A and B is mentioned without elucidating the methodology employed for this division. It is imperative to clarify whether this division was conducted manually or through a specific technique such as k-fold cross-validation for transparency and reproducibility.
- Training Phase: The training phase is referenced without detailing the composition of the training and testing datasets. Clarity regarding the ratio of training to testing data (e.g., 50/50 or 70/30) and the rationale behind this decision is necessary to ensure the rigor of the experimental setup.
- Result Presentation: In Table 1, the author presents the results with corresponding accuracy values, indicating an overall accuracy of approximately 93%. However, the origin of the remaining 7% error rate is not specified. It is recommended to elucidate whether these errors stem from true positives or true negatives. Additionally, providing a confusion matrix and discussing the distribution and significance of true positives, true negatives, false positives, and false negatives would offer deeper insights into the model's performance.