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

Caglar Uyulan[1]

1 İzmir Katip celebi University

Potential competing interests: No potential competing interests to declare.

The authors employed innovative machine learning techniques to analyze EEG data, aiming to distinguish children with ASD from those with other neuropsychiatric disorders.

The methodological approach is robust, involving the use of a Minimum Spanning Tree (MST) for data pre-processing and artificial neural networks for classification.

Here are specific technical and methodological criticisms, along with recommendations for improvement:

Technical and Methodological Gaps

1. While innovative, the description lacks detailed justification for the choice of MST over other potential preprocessing methods. The technical explanation of how MST improves the classification process is somewhat brief and could benefit from a deeper dive into the algorithm's mechanics and its specific advantages in handling EEG data.
2. The paper mentions the use of artificial neural networks (ANNs) for classification but does not provide sufficient detail on the architecture, parameters, or training process of these models. For a technical audience, understanding the specifics of the model design, including the number of layers, type of layers (e.g., convolutional, recurrent), activation functions, and optimization algorithms, is crucial for assessing the robustness and generalizability of the study's findings.
3. The study employs a training/testing cross-validation procedure, which is a standard approach. However, the methodology section could be enhanced by including more information on the validation metrics used (beyond sensitivity, specificity, and global accuracy), such as the area under the receiver operating characteristic (ROC) curve, precision-recall curves, or confusion matrices. Additionally, discussing the choice of cross-validation technique (e.g., k-fold, leave-one-out, or nested CV) would provide insight into how the authors mitigate potential overfitting and ensure the model's generalizability.
4. While the study mentions the use of a dataset comprising 50 ASD children and 50 children with other neuropsychiatric disorders, there is limited discussion on the dataset's diversity, representativeness, and any preprocessing steps taken to ensure data quality and consistency. Details on how artifacts were handled or removed, the criteria for selecting these specific subjects, and any data augmentation techniques would be valuable for understanding the dataset's robustness.

Recommendations for Improvement

1. Provide a more comprehensive rationale for selecting MST as the preprocessing method, including comparisons with other techniques and a detailed explanation of its benefits specifically for EEG data in autism diagnosis.
2. Enhance the description of the ANN models used, including their architecture, training details, and rationale for these choices. This should also include any hyperparameter tuning processes and how the models were validated internally before testing.
3. Include more detailed information on the validation metrics and cross-validation techniques used. Discuss the rationale behind these choices and how they contribute to the robustness and reliability of the study's conclusions.
4. Offer a more detailed description of the dataset, including demographic information, data collection protocols, preprocessing steps, and artifact handling. This will help readers assess the generalizability of the study's findings.
5. Incorporate a more thorough statistical analysis to support the findings. This could include confidence intervals for accuracy metrics, significance testing for comparing model performances, and analysis of model errors to identify potential biases or areas for improvement.

By addressing these technical and methodological gaps, the paper could significantly strengthen its contribution to the field, providing clearer insights into the efficacy and reliability of using MST preprocessing and ANN models for the differential diagnosis of autism through EEG data.