**“Objective evaluation metrics for automatic classification of EEG events”
Ziyabari, Saeedeh; Shah, Vinit; Golmohammadi, Meysam; Obeid, Iyad; Picone, Joseph
Article reference: JNE-102155**

**Reviewer No. 1, Item No. 1**:  **The exact equation/algorithms to compute each metric are not provided, and their hyperparameters not mentioned (when they are key in the results they provide).**

We appreciate the careful review and constructive suggestions of reviewer. We agree with the concern that the equations to compute each metrics should be provided. On this recommendation, we have included equations for all metrics except epoch-based sampling and any-overlap metrics. All previously published work about epoch-based sampling and any-overlap metrices did not include equations. Providing equations without accessing to the software that deals with all special cases is meaningless. Therefore, we implement the algorithms in python and put it online for public access (*https://www.isip.piconepress.com/projects/tuh\_eeg/downloads/nedc\_eval\_eeg*). The software includes the reference implementation of the algorithms, so the users should not deal with optimizing the hyperparameters. As we explained in our paper, optimization of the hyperparameters of the metrices for seizure detection has not produced a significantly different result than we presented in the paper.

**Reviewer No. 1, Item No. 2: The machine learning algorithms evaluated are just described in one line without further details (the corresponding reference is not published yet, and the provided link to the pdf of it is restricted and thus inaccessible).**

We thank the reviewer for this valuable suggestion and agree that it would be interesting to add information about the machine learning algorithms in this paper. We can provide it in a resubmitted version of this paper if that opportunity arises. However, comprehensive details about the machine learning algorithms are presented in other paper which is under the review in the 35th International Conference on Machine Learning (ICML 2018). It is publicly available here: https://www.isip.piconepress.com/publications/unpublished/conferences/2018/icml

**Reviewer No. 1 Item No. 3: The comparisons reported are not statistically analyzed, meaning we do not know whether the different metrics or classifiers are actually any significantly different from each other. Yet the paper discusses on the superiority of some metrics or classifiers, leaving statistical tests for future work. This is not appropriate here.**

Unfortunately, this point was not part of our original manuscript. We can provide a comprehensive information about whether the different metrics or classifiers are actually any significantly different from each other with ours in a resubmitted version of this paper if that opportunity arises.

**Reviewer No. 1, Item No. 4: The paper concludes on the supposed value of the newly proposed metrics where there is no objective criteria to support that, and it is not clear whether such metrics reflect user/community needs (it is not evaluated), nor when or how we should use which metrics and what for**

Unfortunately, this point was not clear in our original manuscript. We would like to apologize for the misunderstanding and have now revised the paper to explain this better. As we explained in the paper, popular metrics such as sensitivity and specificity do not completely characterize the problem and neglect the importance that FA rate plays in achieving clinically acceptable solutions. In our paper, we provide an accurate assessments metric, TAES, that overcome those deficiencies. Additionally, a single popular metric in the speech recognition community, TWV, is proposed to address concerns about tradeoffs for different types of errors. This measure is useful when it is preferred to compare two systems based on a single number. Also, the researchers can find the performance of a system based on the proposed metrices beside the existing ones in our publicly available software.

We would like to thank the reviewer for evaluating our manuscript. We have tried to address all the reviewer’s concerns in a proper way and believe that our paper has improved considerably. We would be happy to make further corrections if necessary and look forward to hearing from you soon.