

Evaluating Effects of Resting-State Electroencephalography Data Pre-Processing on a Machine Learning Task for Parkinson's Disease

Robin VLIEGER ^{a,1}, Elena DASKALAKI ^a, Deborah APTHORP ^b,
Christian J. LUECK ^a and Hanna SUOMINEN ^{a,c}

^a*Australian National University, Canberra, Australian Capital Territory, Australia*

^b*University of New England, Armidale, New South Wales, Australia*

^c*University of Turku, Turku, Southwest Finland, Finland*

Abstract. Resting-state electroencephalography pre-processing methods in machine learning studies into Parkinson's disease classification vary widely. Here three separate data sets were pre-processed to four different stages to investigate the effects on evaluation metrics, using power features from six regions-of-interest, Random Forest Classifiers for feature selection, and Support Vector Machines for classification. This showed muscle artefact inflated evaluation metrics, and alpha and theta band features produced the best results when fully pre-processing data.

Keywords. Parkinson's disease, machine learning, diagnosis, pre-processing, electroencephalography

1. Introduction

Resting-state electroencephalography (RSEEG) pre-processing methods in machine learning studies into Parkinson's disease (PD) classification vary widely [1]. Research shows that non-neural signals, especially muscle artefact, can inflate evaluation metrics [2], which is undesirable classifying PD based on neural activity. This study evaluated classification performance results from different pre-processing pipelines.

2. Methods

Data sets of The Canberra Hospital, Australia, the University of New Mexico (UNM), USA [3], and the University of Turku (UTU), Finland [4], were used. Pre-processing was performed as described here [5], and power features were extracted from six regions-of-interest (ROIs) after re-referencing, after filtering, after full pre-processing, and after pre-processing while retaining muscle artefact. Random Forest Classifiers selected features and Support Vector Machines classified participants. The ACT Health Human

¹ Corresponding Author: Robin Vlieger, School of Computing, Australian National University, 145 Science Road, Acton ACT 2601, Australian Capital Territory, Australia; email: robin.vlieger@anu.edu.au.

Research Ethics Committee (HREC) (ETH.4.16.060) and the Australian National University (ANU) HREC (protocol No. 2020/612) approved this study.

3. Results

UNM and UTU accuracies of data retaining muscle artefact were significantly higher than those for fully pre-processed data ($p < 0.0001$ for both using t-tests). ROIs of selected features differed between these two pre-processing stages in all three data sets. For the fully pre-processed data, selected features came from the alpha and theta bands.

4. Discussion

Our results show that muscle artefact can inflate evaluation metrics. Theta and alpha features were selected most often, but research into frequency band splitting and grouping seems reasonable. Data sets used were rather small and selected ROIs varied per data set, so using larger data sets to obtain more uniform results is encouraged.

5. Conclusions

Artefact removal is essential if the intention is to classify based on neural activity, after which theta and alpha features contribute most to classification accuracy. Further research is needed into which specific features are necessary for accurate classification.

Acknowledgment

This study was funded by Our Health in Our Hands, an initiative of the Australian National University, which aims to transform health care by developing new personalised health technologies and solutions. We gratefully acknowledge the funding from the ANU School of Computing for the first author's PhD studies.

References

- [1] Maitín AM, García-Tejedor AJ, Muñoz JPR. Machine learning approaches for detecting parkinson's disease from EEG analysis: a systematic review. *Appl Sci*. 2020 Dec;10(23):8662, doi: 10.3390/app10238662.
- [2] Weyhenmeyer J, Hernandez ME, Lainscsek C, Sejnowski TJ, Poizner H. Muscle artifacts in single trial EEG data distinguish patients with Parkinson's disease from healthy individuals. *Annu Int Conf IEEE Eng Med Biol Soc*. 2014 Aug;2014:3292-5, doi: 10.1109/EMBC.2014.6944326.
- [3] Steve Carr. UNM researchers design patient repository for more effective EEG diagnoses. Available at <http://news.unm.edu/news/unm-researchers-design-patient-repository-for-more-effective-eeeg-diagnoses>.
- [4] Railo H, Suuronen I, Kaasinen V, Murtojärvi M, Pahikkala T, Airola A. Resting state EEG as a biomarker of Parkinson's disease: Influence of measurement conditions. *bioRxiv*. 2020 May:2020-05, doi: 10.1101/2020.05.08.084343.
- [5] Vlieger R, Daskalaki E, Apthorp D, Lueck CJ, Suominen H. Evaluating effects of resting-state electroencephalography data pre-processing on a machine learning task for parkinson's disease. *medRxiv*. 2023 Mar:2023-03, doi: 10.1101/2023.03.06.23286826.