Machine Learning-Based Sleep Scoring for the Bernese Sleep Registry

Marco Pesce

Supervisor: Prof. Dr. Athina Tzovara
Institutions: University of Bern, ARTORG Center for Biomedical Engineering Research
University Hospital Bern (Inselspital), Department of Neurology
Examiners: Prof. Dr. Athina Tzovara, PD Dr. Wimmer Wilhelm

Introduction

The main test that constitutes the gold-standard sleep assessment and diagnostic tool is the so-called polysomnography (PSG). The analysis of an all-night PSG requires hours of tedious work if done manually by trained operators, which is why research has made great strides to automate this manual procedure: many different techniques and approaches have been proposed, achieving excellent results in terms of accuracy. However, automated scoring is not yet an everyday reality in sleep centers. The aim of this Master's thesis project is to assist in the curation and creation of a database containing clinical and PSG-derived data recorded over the last 20 years at the Schlaf-Wach-Epilepsie Zentrum (SWEZ) of the InselSpital in Bern, and then to use dataset from this database to test and validate two state of the
art sleep scoring algorithms.

Materials and Methods

The database assembled provides access to large collections of de-identified PSG signals and clinical data collected in research cohorts and clinical studies. We targeted two main disorders: patients with Obstructive Sleep Apnea (OSA), and patients with Narcolepsy Type 1 (NT1). The available data have been divided into training, testing and validation sets. Two different sleep scoring algorithms were then applied and compared. The first (YASA) was a tree-based gradient boosting classifier, which is powerful in scoring data of diverse populations and matching inter-scorer agreement.The second (Stanford-stage) was based on neural networks that have been developed specifically for patients with narcolepsy.

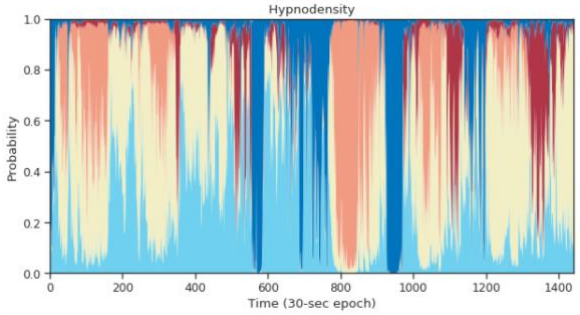

Fig. 1 The predicted hypnodensity for a randomly selected OSA test patient can be generated by the Sleep Stage algorithms studied

Results

Both algorithms achieved results consistent with published predictions. For OSA patients, for whom a larger sample was provided, the best performance was obtained by the YASA algorithm, which achieved a median F1-Score of 79.45% considering all 5 sleep stages, whereas the median F1-Score of the Stanford-stage was 70.01%. Also for the NT1 cohort, the YASA algorithm performed better than the Stanford-stage model (74.03% and 67.23%, respectively, for the best test folds).

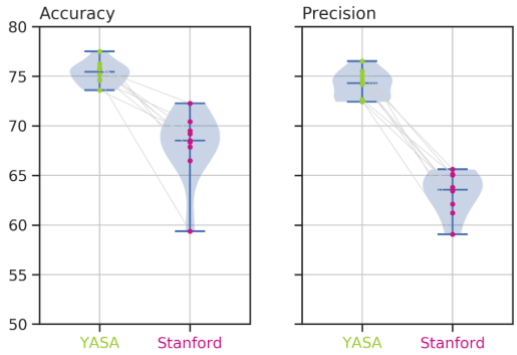

Fig. 2 YASA vs Stanford-stage trained classifier model performances for the OSA subset.

Discussion

Our results demonstrated the capability of the new Bernese Sleep Registry as a digital resource for scientific and clinical use, applied to the domain of sleep stage scoring. Finally, we conclude that machine learning has a strong potential to be used in the future for automating sleep scoring in clinical data. Importantly, automated scoring algorithms introduce a new concept of visually illustrating the probability densities of sleep stages, called hypnodensity graphs. This type of representation can provide further insights into sleep architecture, and provides, at least for one of the two algorithms, higher temporal resolution than traditional sleep staging.

References

Matthew P. Walker Raphael Vallat. "A universal, open-source, high-performance tool for automated sleep staging". In: (2021). doi: https://doi.org/10.1101/2021.05.28.446165.

Acknowledgements

I would like to thank Prof. Dr. Athina Tzovara, Dr. Julia Van der Meer and Florence M. Allen for the great educational experience they provided.