Abstract

IN the ever-evolving landscape of biomedical signal analysis, the pursuit of accuracy can be hindered by interpretation biases. These biases can arise at different stages of the research process, from study design to publication.

The heart of this thesis lies in Explainable Artificial Intelligence (xAI), striving to illuminate the intricacies of “black-box” decision-making processes that underlie most contemporary deep learning algorithms. Our objective is to mitigate the potential for biased conclusions stemming from these non-transparent models. Therefore, we propose a novel xAI approach, termed *human-centered explainable AI*, which leverages intra and inter-subject similarities to extract pivotal features forming the basis for classification or regression tasks. Inspired by human decision-making processes based on comparisons, this technique is applied to sleep data. It yields a novel severity measure for sleep apnea events and uncovers electroencephalographic (EEG) biomarkers associated with severe sleep apnea events.

To promote the prioritization of clinician-interpretable features by AI models, we investigate the incorporation of “white-box” analyses that provide human-friendly representations of the recorded signals. However, these analyses can also

introduce biases and, although easy to understand, may inadvertently limit data exploration. This limitation could cause us to overlook important factors or effects beyond the scope of the analysis. As a response, we propose techniques to address vulnerabilities in experimental protocol design, sub-optimal recordings, and misrepresentations of target information. Focusing on EEG signal processing, this thesis introduces standardized frameworks covering the evaluation of confounder factors' effects, EEG preprocessing, and the validation of brain source reconstruction.

In summary, our overarching goal is to counteract interpretation biases in biomedical signal analysis, thereby fostering transparency, precision, and ethical integrity. Our future endeavors are aimed at extending the human-centered xAI approach to encompass multimodal data and diverse medical applications. This journey holds the promise of not only technological advancement but also a profound shift towards reliable medical diagnostics and research.