

# An Interpretable Deep Learning Model for EEG Signals

Amirhessam Tahmassebi<sup>1</sup>, Anke Meyer-Baese<sup>1</sup>, and Amir Gandomi<sup>2</sup>

<sup>1</sup>Florida State University

<sup>2</sup>University of Technology Sydney

June 15, 2020

## Abstract

Cutting-edge methods in artificial intelligence (AI) have the ability to significantly improve outcomes. However, the struggle to interpret these black box models presents a serious problem to the industry. When selecting a model, the decision to sacrifice accuracy for interpretability must be made. In this paper, we consider a case study on eye state detection using electroencephalogram (EEG) signals to investigate how a deep neural network (DNN) model makes a prediction, and how that prediction can be interpreted.

## Hosted file

AI\_Letter\_Wiley\_EEG\_DNN\_SHAP.pdf available at <https://authorea.com/users/318151/articles/459730-an-interpretable-deep-learning-model-for-eeeg-signals>