



Contribution ID: 54

Type: **not specified**

## Machine learning and interpretability methods to investigate Alzheimer's disease

*Wednesday, 15 May 2024 11:45 (25 minutes)*

The use of machine learning methods in clinical settings is increasing. One reason for this is the availability of more complex models that promise more accurate predictive performance, especially for the study of heterogeneous diseases with multimodal data, such as Alzheimer's disease. However, as machine learning models become more complex, their interpretability decreases. The reduced interpretability of so-called black-box models can have a negative impact on clinicians' and patients' trust in the decisions of a system. For this reason, several methods to overcome this problem have been developed. These methods are summarised under the terms of interpretable machine learning and explainable artificial intelligence.

The presented research investigates how methods from the domain of interpretable machine learning and explainable artificial intelligence can be used for the early detection of Alzheimer's disease. To this end, a systematic comparison of different machine learning and explanation methods is presented. The comparison includes black-box deep learning convolutional neural networks, classical machine learning models as well as models that are interpretable by design. For all models, several reasonable explanation methods such as SHAP and GradCAM were applied. Common problems such as model calibration and feature correlations, which often occur when working with explainability methods, were addressed during the investigations. It is validated whether the activated brain regions and learned connections in different models are biologically plausible. This validation can increase the trust in the decision of machine learning models. In addition, it is investigated whether the models have learned new, biologically plausible connections that can help in the development of new biomarkers.

### Type of presentation

Invited Talk

**Primary author:** BLOCH, Louise (Department of Computer Science, University of Applied Sciences and Arts Dortmund)

**Presenter:** BLOCH, Louise (Department of Computer Science, University of Applied Sciences and Arts Dortmund)

**Session Classification:** Invited session

**Track Classification:** Spring Meeting