AI meets AD: Coining the future of early and accurate prediction of Alzheimer's Disease progression

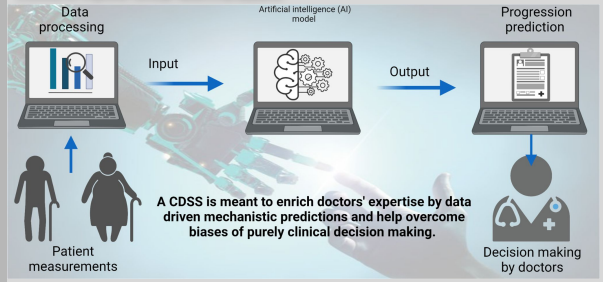
Students: Andrade Leon, R.; Bruton, O.; Jafri, S. / Supervisor: Prof. Dr. Andrea Hildebrandt

Background

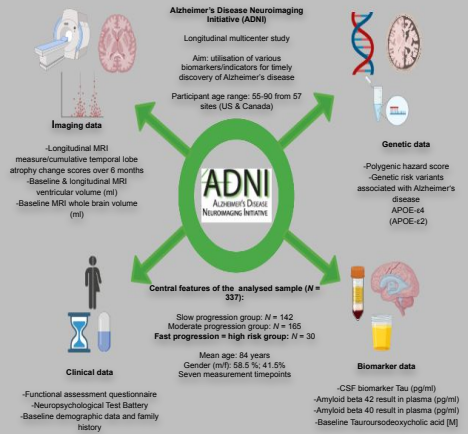
Alzheimer's Disease (AD) is a **progressive** neurodegenerative disease known to severely impact cognitive functions. By the year 2050, 35% in developed countries and 21% of all individuals worldwide are predicted to be older than 60 years (United Nations, 2005).

The **progression strength of AD differs across individuals** and its prediction is challenging even for experienced clinicians. However, mechanistic (statistical) predictions were shown to outperform clinicians' predictions which are prone to biases, such as ignoring base rates, assigning non-optimal weights to features, and failure to properly assess covariation. Machine learning algorithms - when trained with sufficient data - might be more accurate in predicting the progression of AD.

Aim: Contrast different machine learning algorithms with respect to their prediction accuracy of AD progression. These algorithms might be implemented into **Clinical Decision Support Systems** (CDSS) to improve clinical recommendations.

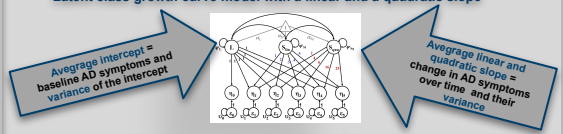


Study design and input features

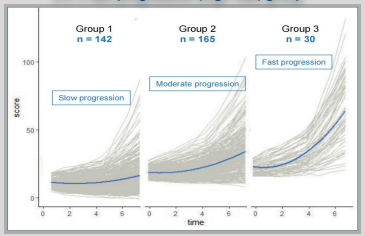


Identifying AD progression classes

Latent class growth curve model with a linear and a quadratic slope



- Latent **classes** were allowed to **vary** in **intercept**, **linear** and **quadratic slope**
- The number of classes was identified in a stepwise manner, by comparing the log-likelihood of models with an additional class
- Adding a fourth class to the model did not substantially improve the model fit
- A three class solution revealed a slow progression, a moderate progression, and a **fast progression (high risk) group**



High risk = fast progression individuals turned out to be much less frequent in this sample → **a class imbalance problem** to be addressed when training machine learning models

Machine Learning methods applied

Naive Bayes classifier builds upon the Bayes' Theorem. It assumes the features to be mutually independent and outputs the highest probability for a given class.

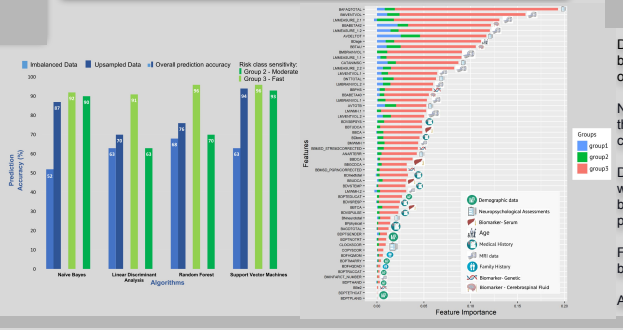
Linear Discriminant Analysis maximizes the component axes for class separation in order to find the axes for best class separability.

Random Forest classifier is an ensemble algorithm, which creates a set of decision trees (each a bit different from the other) from a randomly selected subset of the training dataset, which then aggregates the votes from different decision trees to decide the final class of the test observation.

Support Vector Machines are based on the idea of finding a hyperplane that best divides a dataset into two or more classes depending on the input features.

All analyses were performed by means of the **mlr** package in the R Software for Statistical Computing.

Results



Conclusions and future directions

Developing machine learning algorithms is a challenge in case of imbalanced data between the AD progression classes. Resampling techniques can be used to overcome this problem.

Naïve Bayes and SVM evince higher accuracy in correctly classifying individuals in the original dataset. Predictive sensitivity for both the high and the moderate risk class appears to be best for Naïve Bayes and SVM.

Due to their disproportionate relevance, imaging variables (e.g., baseline MRI whole brain volume), neuropsychological tests and certain biomarkers (e.g., baseline Amyloid-ß 42) should be focussed on when predicting risk for *fast* progression of AD symptoms over time.

Further data is needed to test the robustness of the achieved predictions and to better overcome the class imbalance problem in the future.

Algorithmic predictions need experts' acceptance prior to designing a CDSS.

raxide.eugenia.andrade.leon@uni-oldenburg.de ; oliver.bruton@uni-oldenburg.de ; sumbul.jafri@uni-oldenburg.de