**PROBAST**

Study:

PD Disease State Assessment in Naturalistic Environments Using Deep Learning

Step 2: Type of prediction study

**Is the study a diagnostic or a prognostic study?**

Diagnostic

**Is the study a development only, development and validation or validation only study?**

Development only

**What is the model of interest?**

Restricted Boltzmann Machines

**What is the outcome of interest?**

PD disease state

Step 3: Assess risk of bias

**Domain 1: Participants**

**Describe the sources of data and criteria for participant selection**

Overall 34 participants were recruited who exhibited mild to severe level Parkinson’s Disease (Hoehn and Yahr stages I-IV (Hoehn and Yahr 1998)), were not significantly cognitively impaired and were taking immediate-release levodopa medication.

**1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?**

Y

**1.2 Were all inclusions and exclusions of participants appropriate?**

Y

**Risk of bias introduced by selection of participants:**

Low

**Rationale of bias rating**

Reasonable eligbility criteria for PD

**Domain 2: Predictors**

**List and describe predictors included in the final model, e.g. definition and timing of assessment**

Accelerometer data from wearable devices

**2.1 Were predictors defined and assessed in a similar way for all participants?**

Y

**2.2 Were predictor assessments made without knowledge of outcome data?**

Y

**2.3 Are all predictors available at the time the model intended to be used?**

Y

**Risk of bias introduced by predictors or their assessment**

Low

**Rationale of bias rating**

Predictors are assessed the same way for everyone, independent of outcome and applicable.

**Domain 3: Outcome**

**Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:**

Each participant filled out a disease state diary, a pre-formatted document where ticks indicate disease state for each hour, to the best of their abilities. The diary included: asleep, off, on, and(troublesome) dyskinesia. A total of approx. 5,500 hours of accelerometer data was collected, for which approx. 4,500hourly labels were provided by the participants ( 80% diary compliance). The labels are inherently unreliable, as symptom characteristics are very unlikely to change exactly on the hour, participants may have trouble classifying their own disease state, and diaries may be filled out retrospectively at the end of the day.

**3.1 Was the outcome determined appropriately?**

PN

**3.2 Was a pre-specified or standard outcome definition used?**

PN

**3.3 Were predictors excluded from the outcome definition?**

Y

**3.4 Was the outcome defined and determined in a similar way for all participants?**

PN

**3.5 Was the outcome determined without knowledge of predictor information?**

Y

**3.6 Was the time interval between predictor assessment and outcome determination appropriate?**

PN

**Risk of bias introduced by the outcome or its determination**

High

**Rationale of bias rating**

Outcome is at risk of various different biases such as recall bias due to the possibility of retrospective answer. Also it is not certain whether hourly labels can capture the disease state as described by the researchers.

**Domain 4: Analysis**

**Describe number of participants, number of candidate predictors, outcome events and events per candidate predictor**

5,500 hours of movement data collected from 34 participants. Around 4500 labels

**Describe how the model was developed, predictor selection and risk group definition**

Restricted Boltzmann Machines

**Describe whether and how the model was validated, either internally (cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants)**

K-fold crossvalidation

**Describe the performance measures of the model, e.g. calibration, discrimination, classification, net benefit, and whether they were adjusted for optimism**

F1, ACC, SEN, SPE

**Describe any participants who were excluded from the analysis**

None

**Describe missing data on predictors and outcomes as well as methods used for missing data**

Not described

**4.1 Were there a reasonable number of participants with the outcome?**

Y

**4.2 Were continuous and categorical predictors handled appropriately?**

Y

**4.3 Were all enrolled participants included in the analysis?**

Y

**4.4 Were participants with missing data handled appropriately?**

U

**4.5 Was selection of predictors based on univariable analysis avoided?**

Y

**4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls)**

**accounted for appropriately?**

Y

**4.7 Were relevant model performance measures evaluated appropriately?**

Y

**4.8 Were model overfitting and optimism in model performance accounted for?**

Y

**4.9 Do predictors and their assigned weights in the final model correspond to the results**

**from multivariable analysis?**

U

**Risk of bias introduced by the analysis**

Low

**Rationale of bias rating**

Proper handling of data and good reporting of performance and proper CV. Good amount of outcomes.

**Overall Risk of bias**

High