Prediction of Alzheimer's disease. The impact of information about functional activity level reported by relatives in an early stage of the disease.

Alexandra Vik[a], Marek Kocinski[a,c], Astri J. Lundervold[b], Ingrid Rye[b,a], Alexander S. Lundervold[a,d]

[a]*Mohn Medical Imaging and Visualization Centre (MMIV), Department of Radiology, Haukeland University Hospital*
[b]*Department of Biological and Medical Psychology, University of Bergen & K. G. Jebsen Centre for Research on Neuropsychiatric Disorders*
[c]*Department of Biomedicine, University of Bergen & Mohn Medical Imaging and Visualization Centre, Department of Radiology, Haukeland University Hospital*
[d]*Department of Computer science, Electrical engineering and Mathematical sciences, Western Norway University of Applied Sciences, Bergen, Norway*

Introduction

Managing everyday life activities (ADL) such as finances, medication, running errands, preparing meals and maintaining interests, is one of the criteria differentiating between mild cognitive impairment (MCI) and Alzheimer's disease (AD). Major disabilities in such activities are hallmarks of AD. However mild cognitive symptoms may still interfere with ADL and subtle changes may be noticed by the patients and relatives long before symptoms can be defined as AD. In the present study we expand on previously presented prediction models by adding ADL as a feature in our RF classification model. Furthermore, feature importance will be estimated to investigate the contribution of ADL in predicting trajectories towards AD in a group of patients with MCI.

Methods

Data set

The present study included longitudinal data from the Alzheimer's disease Neuroimaging initiative (ADNI) (adni.loni.usc.edu):

Inclusion criteria for MCI in ADNI [3]:

Cognitive complaint, objective memory impairment as measured by scoring below education-adjusted cutoff on the Wechsler Memory Scale Revised, Logical Memory II sub-scale (delayed recall); Mini-Mental State Examination (MMSE) score between 24-30; Clinical Dementia Rating of 0.5; and preserved functional abilities.

Our group defenitions:

- **sMCI**: MCI diagnose maintained across all visits – stable MCI
- **cAD**: adults progressing to AD at one time point during the examination period – MCI converting to AD.

	sMCI (360) Train (285)/Test (75)	cAD (320) Train (255)/Test (65)
Demographics		
Sex (F:M)	114:171/32:43	99:156/25:40
Age at inclusion [years]: mean (SD)	73.9 (7.4)/72.7(7.3)	73.9 (7.7)/73.9 (6.9)
Age at inclusion [years]: range	55-91/57.8-87.8	55.2-88.3/55-88.4
Education [years]: mean (SD)	15.8 (2.9)/16.2(2.9)	15.8 (2.9)/16.2(2.9)
Number of Visits Cog/Func: mean (SD)	8.4 (3.7)/8.4(3.3)	9.3 (3.8)/10.2(4.2)
Number of Visits MRI: mean (SD)	5.3 (1.6)/5.4(1.4)	5.8 (1.6)/6.0(1.7)

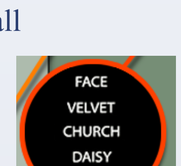
Table 1: Subject Demographics. Subsample from the ADNI cohort.

Neuropsychological features

ADNI database includes several neuropsychological measures, and the following is included in our RF model:

Cognitive tests:

- *Memory function*:
  - Rey Auditory Verbal Learning Test, immediate, **(RAVLT-Im)** (15 word list, sum of 5 trials)
  - RAVLT 30-minute delayed free recall (**RAVLT- Delay**)
  - Recognition between correct and distractor words **(RAVLT-Recog)**



- *Executive functions* (EFs);
  - Trail Making Test, part A **(TMTA)** (150 sec max)
  - TMT, part B **(TMTB)** (300 sec max)
  - Catergory fluency test **(CFT):** animals

Questionnairs:

- Geriatric depression scale **(GDS)** (0-5 normal).
- Functional activities of daily living **(FAQ)** (0 – 30, low-high severity, cutoff point = 9 (dependent in 3 or more activities) [4].

Functional Activities Questionnaire

Administration

Ask informant to rate patient's ability using the following scoring system:
- Dependent = 3
- Requires assistance = 2
- Has difficulty but does by self = 1
- Normal = 0
- Never did [the activity] but could do now = 0
- Never did and would have difficulty now = 1

Writing checks, paying bills, balancing checkbook	
Assembling tax records, business affairs, or papers	
Shopping alone for clothes, household necessities, or groceries	
Playing a game of skill, working on a hobby	
Heating water, making a cup of coffee, turning off stove after use	
Preparing a balanced meal	
Keeping track of current events	
Paying attention to, understanding, discussing TV, book, magazine	
Remembering appointments, family occasions, holidays, medications	
Traveling out of neighborhood, driving, arranging to take buses	
TOTAL SCORE:	

Magnetic resonance imaging (MRI)

Two MRI derived features were included in the model:

i) The hippocampus **(HC)** voulme, regarded as a hallmark region for memory function in neurodegenerative diseases and,

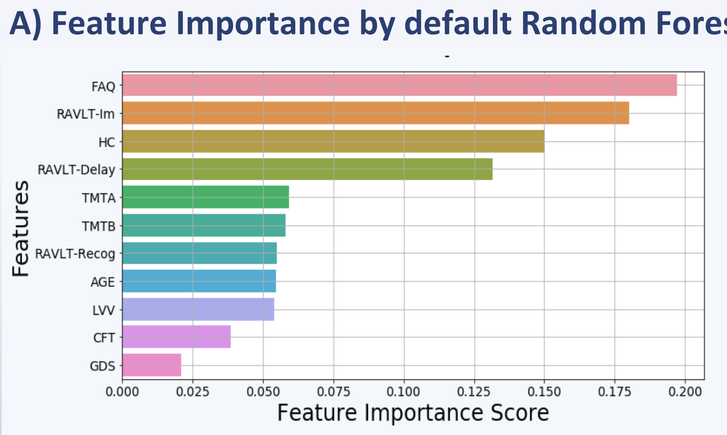
ii) The lateral ventricle volumes **(LVV),** used as a proxy of overall brain tissue loss.

The T1-Weighted MRI images were processed on FreeSurfer longitudinal stream v.7.1.1 [5]. Specifically, an unbiased within-subject template space and image is created using robust, inverse consistent registration [5].

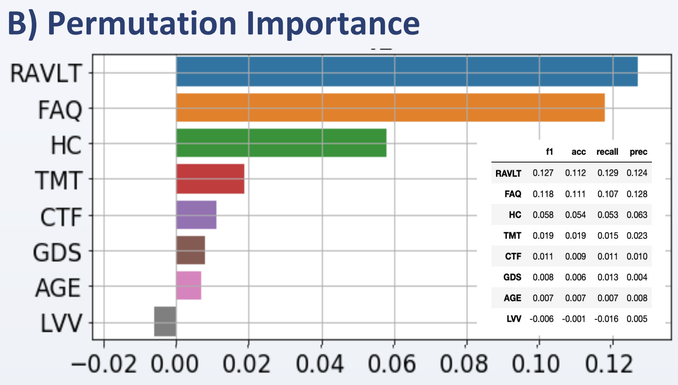
Preliminary Results

When the 10 selected neurocognitive features and age were used as input in a Random Forest (RF) binary classifier (sMCI vs. cAD) we obtained an accuracy-precision- recall- and f1- score of 73.6%/69.2%/72.6%/70.9% using a *k*-fold cross validation procedure with 10-folds. Results from the RF feature importance was also confirmed across three different permutation driven importance estimates, showing that measures of every day functioning, verbal memory and hippocampus volume are highly relevant for future conversion to AD in our prediction model. All results are displayed on the test-set.

A) Feature Importance by default Random Forest



B) Permutation Importance



C) Drop Feature Importance

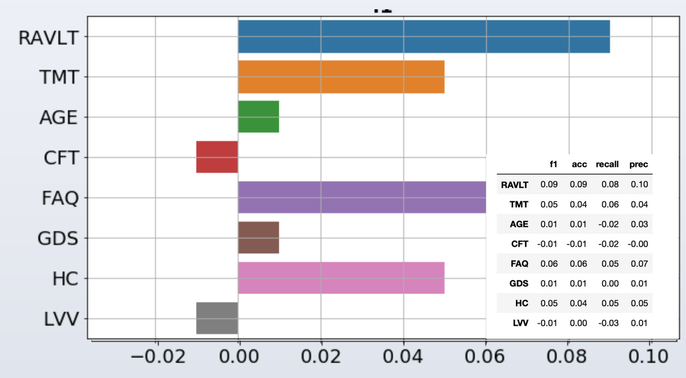


Figure 1: Feature importance: illustrated by the default in RF (**A).** Feature importance was further evaluated across different permutation techniques: **B)** permutation importance (shuffled 2000 times) and **C)** drop column importance (importance score) is illustrated by the F1-score (complete model evaluation table is superimposed). Due to multicollinearity RAVLT subscores (Im, Delay and Recog) were grouped: RAVLT, and also for the two parts, A and B, of the TMT (in B and C).

D) SHAP importance

SHapley Additive exPlanations (SHAP) values [1, 2] is here defined as the average marginal contribution of a feature value across all possible feature coalitions. Thus, a Shapley value for a given feature value can be interpreted as the difference between the actual prediction and the average prediction for the whole dataset.

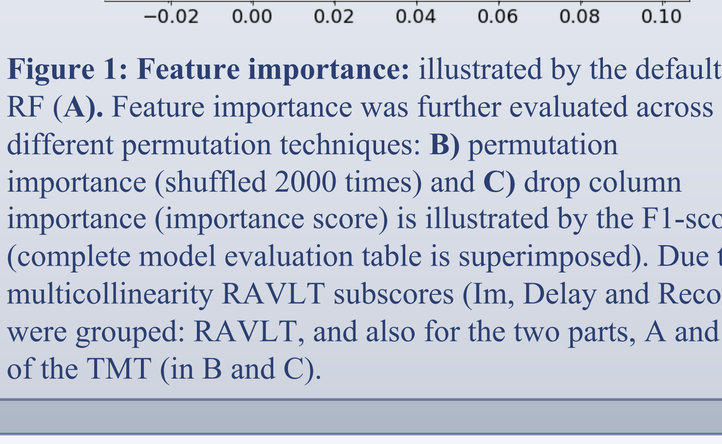


Figure 2: SHAP summary plot: of the features of the RF model. A dot is created for each feature attribution value for the model of each subject. Dots are colored according to feature values. Thus, higher values (represented by red color) for FAQ and lower values for RAVLT –Im, -Delay, and HC (blue) increase the prediction to convert to AD. Symmetrically, low values in FAQ and high values in RAVLT and HC decrease the prediction to convert to AD. When the distribution is clustered around 0 indicates that the feature is less relevant. The more skewed the distribution the more important the feature.

Preliminary Model Exploration

To investigate the specific effect on the prediction model within the range of values of the FAQ we applied the partial dependency plot **(PDP)** and individual composition expectation **(ICE)** plot [2].

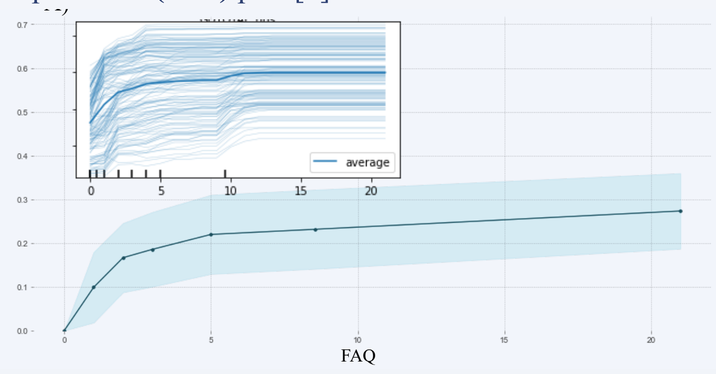


Figure 3: **PDP and ICE**: illustrates the marginal effect of FAQ total score have on our RF model. The X-axis represents the range of FAQ values (0-30) and the Y-axis shows changes in the prediction, positive values represent the contribution of the FAQ to the increase in the odds to convert to AD. The shaded area represents the standard deviation. Interestingly, small increase in the FAQ score (1-2) increase the odds of converting to AD and from the score 5 there is an above 20% increased chance to convert to AD. The same effects can be observed in the superimposed ICE plot.

SHAP auto-cohort feature explanation

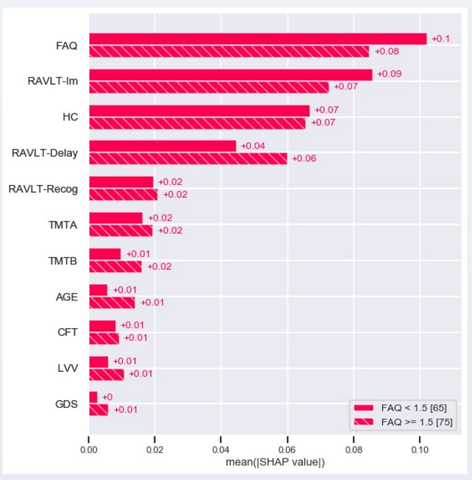


Figure 4: Two cohorts are optimally separated by the SHAP values by applying auto-cohort feature of explanation, utilizing a DecisionTreeRegressor from sklearn. By this, a separation is given between those scoring less than 1.5 and those ≥ 1.5 in the FAQ.

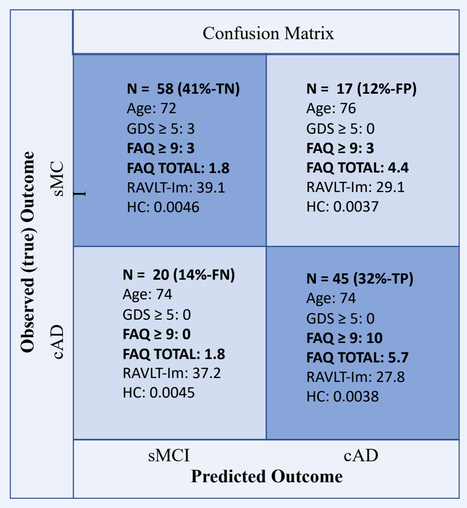
Hence, creating two cohorts with 65 and 75 subjects in each. The bar plot displays the mean SHAP values for each group, for each feature. Lower FAQ scores may contribute to better prediction performance across features. The opposite in the cohort consisting of subjects with higher FAQ scores, the RAVLT-Delay seem to be more influential.

Characteristics of correctly and mis-classified subjects

	False Negative	False Positive	True Negative	True Positive
Demographics				
N	20	17	58	45
Sex (F:M)	8:12	7:10	25:33	17:28
Age [years]: mean (SD)	74 (8.3)	76 (5.5)	72 (7.5)	74 (7.0)
Education [years]: mean (SD)	16.1 (2.5)	16.8 (3.5)	16.0 (2.9)	15.4 (2.8)
Participation length [years]: mean (SD)	6.1 (2.8)	3.4 (2.7)	4.8 (2.7)	5.3 (2.9)
Cognitive function				
RAVLT immediate recall: mean number (SD)	37.2 (6.7)	29.1 (6.2)	39.1(8.7)	27.8(4.8)
RAVLT delayed: mean number (SD)	4.3 (3.2)	1.2 (1.3)	5.7 (3.5)	1.5 (1.9)
RAVLT recognition: mean number (SD)	11.6 (2.6)	9.0 (3.6)	12.1 (2.6)	9.0 (3.7)
TMTA: mean seconds (SD)	41 (29)	43 (8)	37 (12)	45 (25)
TMTB: mean seconds (SD)	115 (82)	131 (56)	93 (36)	140 (80)
CFT animals: mean number (SD)	16.9 (4.7)	16.2 (5.6)	18.7 (4.7)	15.2 (4.1)
MRI measures				
LVV (eTIV normalized): mean (SD)	0.025 (0.010)	0.030(0.011)	0.023 (0.011)	0.030 (0.015)
Hippocampus (eTIV normalized): mean (SD)	0.0045 (0.00)	0.0037(0.00)	0.0046 (0.00)	0.0038 (0.00)
Functional levels				
GDS: mean (SD)	1.5 (1.2)	1.6 (0.9)	1.9 (1.4)	1.4 (1.2)
FAQ Total: mean (SD)	1.8 (2.3)	4.4 (4.8)	1.8 (3.7)	5.7 (4.8)

Table N: Classification in test set. Characteristics of the correctly and miss-classified sMCI and cAD labels returned

Figure 4:
Confusion matrix



The diagonal cells (shaded in blue) are those representing correctly classified subjects (N: number of occurrences). Off-diagonal cells represents various events of misclassification. Each cell is also accompanied with corresponding information about FAQ etc.

Conclusion

Results from the prediction model, here presented in different plots and by presenting characteristics of correctly and mis-classified subjects, support the importance of subtle, behavioural changes in everyday situations noticed by relatives or other informants when predicting prognosis for MCI patients at risk for conversion to AD. In further work unique and interaction effects between measures of ADL, memory function and related brain structures and function will be investigated to obtain a more detailed information about their contributions in a classification model.

References

[1] Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I.: From local explanations to global understanding with explainable ai for trees. Nature machine intelligence **2**, 56–67 (2020). DOI 10.1038/s42256-019-0138-9

[2] Molnar, C.: Interpretable machine learning. Lulu. com (2020)

[3] Petersen, R.C.: Mild cognitive impairment as a diagnostic entity. J Intern Med **256**(3), 183–194 (2004). DOI 10.1111/j.1365-2796.2004.01388.x. URL http://dx.doi.org/10.1111/j.1365-2796.2004.01388.x

[4] Pfeffer, R.I., Kurosaki, T.T., Harrah, C.H., Chance, J.M., Filos, S.: Measurement of functional activities in older adults in the community. Journal of gerontology **37**, 323–329 (1982). DOI 10.1093/geronj/37.3.323

[5] Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B.: Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage **61**(4), 1402–1418 (2012). DOI 10.1016/j.neuroimage.2012.02.084. URL http://dx.doi.org/10.1016/j.neuroimage.2012.02.084

Contact Info: Tel.: +4790959396 Alexandra.Vik@hvl.no/alexandravik3@gmail.com