Supplementary File C
Classification experiments

1 Sample sizes for the automated prediction of progression to AD dementia

MCI participants in the evaluation set were labelled as stable or converter depending on five time windows. The number of MCI patients in the evaluation set who remained stable (sMCI) or converted to dementia (cMCI) within each time window is presented in Table 1. As expected, with longer time windows, the number of stable MCI subjects decreases while the number of MCI who converted to dementia increases. At each iteration of the cross-validation scheme, the RF classifiers were trained with a balanced sets of cases given by a 70%-30% partition of the underrepresented class. This procedure made classifiers within 12 months be trained with balanced sets of 64 cases but tested with highly unbalanced sets of 338 cases, while classifiers within 60 months were trained with 152 cases (76 sMCI - 76 cMCI) and tested with a sample of 79 subjects roughly balanced. Therefore, evaluation of the short term prediction resulted to be more challenging than the long term prediction.

	Total		Training RF		Testing RF	
Time window	sMCI	cMCI	sMCI	cMCI	sMCI	cMCI
12 months	356	46	32	32	324	14
24 months	263	82	57	57	206	25
36 months	206	99	69	69	137	30
48 months	159	114	80	80	79	34
60 months	109	122	76	76	33	46

Table 1: Number of MCI subjects that remained stable (sMCI) and converted to dementia (cMCI) within each time window, along with the number of subjects per class that were used to train and test the Random Forest (RF) classifier at each iteration of the cross-validation scheme.

2 Classification with psychiatric symptoms

To test if including psychiatric symptoms improves the prediction of progression, additional classification experiments we performed following the same cross-validation scheme. The additional classifiers were trained with the domain composite scores, age, sex, years of education, and two neuropsychiatric (NP) assessments: the Geriatric Depression Scale (GDS) and the abbreviated version of the Neuropsychiatric Inventory (NPI-Q). Figure 1 shows the resulting distribution of the Area Under the Curve (AUC) values compared to the classification without psychiatric information.
Given that at each iteration of the cross-validation there are two classifiers, with and without psychiatric information, the mean AUC values across iterations were compared with paired t-tests. As presented in Table 2, the addition of psychiatric information results in small but significant improvements of the AUC.

		12 months	24 months	36 months	48 months	60 months
mean AUC	without NP	0.684	0.753	0.735	0.741	0.755
	with NP	0.688	0.76	0.741	0.753	0.763
mean AUC change		+0.004	+0.007	+0.006	+0.012	+0.008
p-value		≤ 0.00001	≤ 0.00001	≤ 0.00001	≤ 0.00001	≤ 0.00001

Table 2: Number of MCI subjects that remained stable (sMCI) and converted to dementia (cMCI) within each time window, along with the number of subjects per class that were used to train and test the Random Forest (RF) classifier at each iteration of the cross-validation scheme.

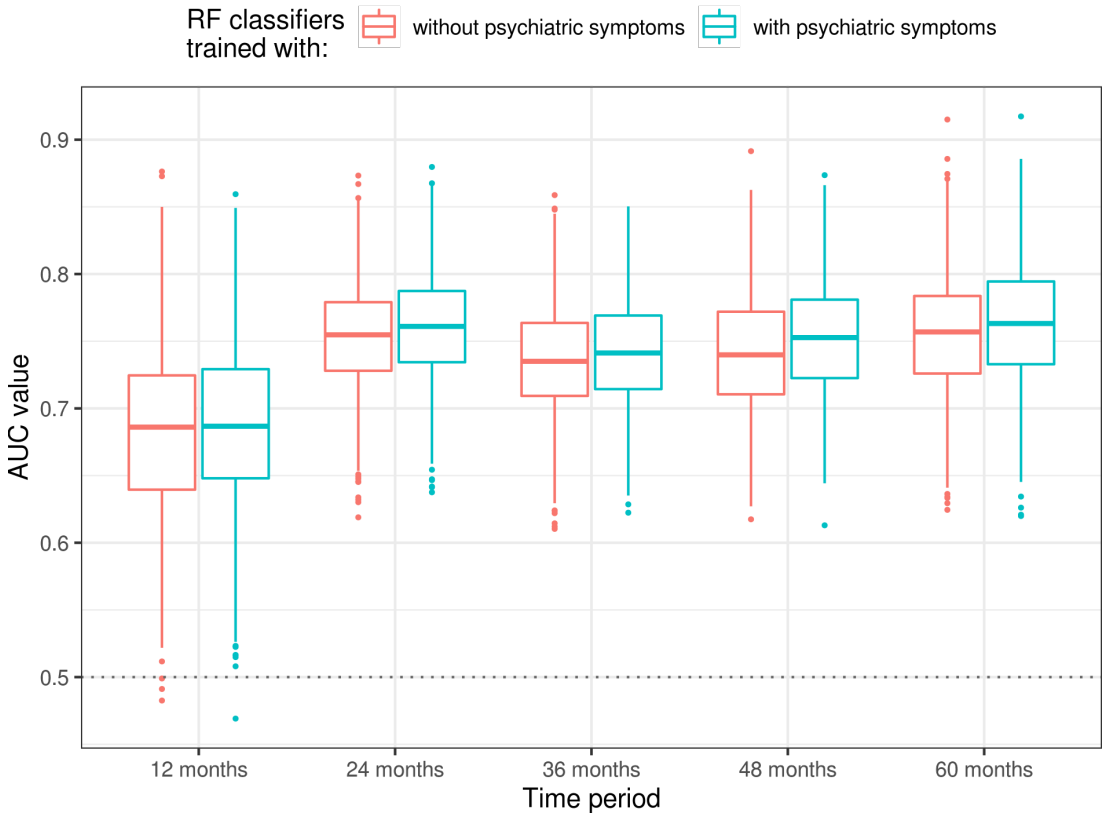

Figure 1: Distribution of AUC values for MCI conversion prediction within 12, 24, 36, 48 and 60 months. Classifiers were trained with domain scores, age, sex, years of education, with and without assessments of psychiatric symptoms.

3 Classification with other composite scores

To compare the prediction of MCI progression to dementia with domain scores against other methods in the literature, nine different sets of features were used to train the same classifiers (Random forest) with the same data at each iteration of the cross validation scheme. The different sets of features related with different composite scores and/or classification results in the literature are:

1. PROPOSED domain-specific composite scores.
2. PROPOSED domain-specific composite scores, with the Clinical Dementia Rating (CDR) - Sum of Boxes, and the Functional Assessment Questionnaire (FAQ).
3. ADAS Tree = 1.05*Q1SCORE + 0.38*Q2SCORE + 0*Q3SCORE + 1.17*Q4SCORE + 0.61*Q5SCORE + 0.13*Q6SCORE + 1.13*Q7SCORE + 0.41*Q8SCORE + 0.54*Q9SCORE + 0.49*Q10SCORE + 0.69*Q11SCORE + 0.39*Q12SCORE + 0.68*Q13SCORE.
   Reported AUC = 0.746 [1].
4. Composite = Q1SCORE + Q4SCORE + Q7SCORE + CDRSB + FAQTOTAL.
   Sensitivity assessed by signal-to-noise ratios [2].
5. Cognitive composite 1 [3]: CC1 = ADAS3 + (75-RAVLT.IMMED) + (30 - MMTOTAL).
6. Cognitive composite 2 [3]: CC2 = ADAS3 + CDMEMORY.
7. Cognitive–functional composite 1 [3]: CFC1 = CC1 + FAQTOTAL.
8. Cognitive–functional composite 2 [3]: CFC2 = CC2 + FAQTOTAL.
   Performance of CC1, CC2, CFC1, and CFC2 was assessed based on sample size requirements for a 2-year clinical trial [3].

9. Selected features [4]: TRABSCOR, Forget.index, RAVLT.IMMED, TOTAL13, TRAASCOR, AVTOT6, LIMMTOTAL, CATANIMSC, AVDEL30MIN, FAQTOTAL, LDELTOTAL, MOCADLREC, AVDELTOT, BNTTOTAL, Q4SCORE, Q8SCORE, MMTOTAL, Q1SCORE, MOCAFLUEN, CDORIENT, CDHOME, AVTOTB.
   Reported AUC values for time widows of 2, 3, and 4 years: 0.821, 0.856, and 0.868, respectively.

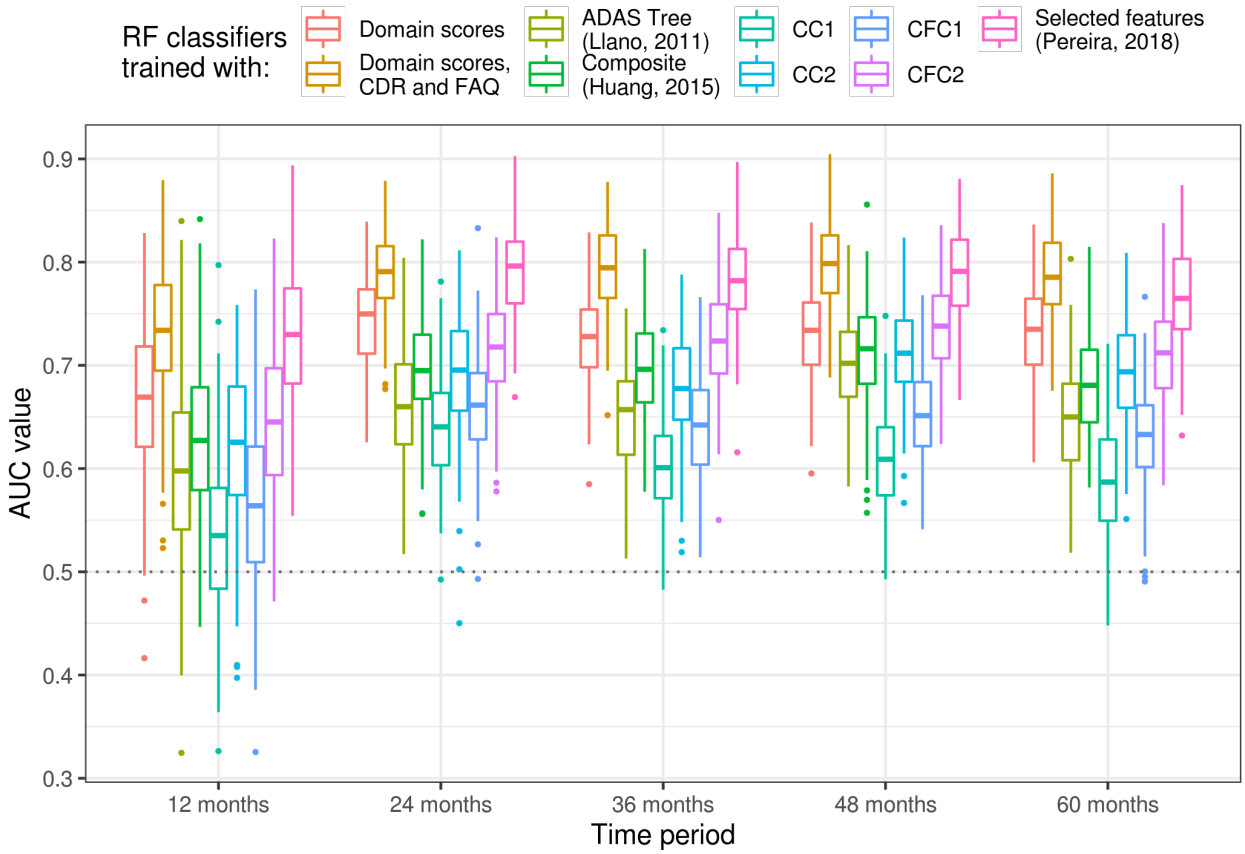

Figure 2: Distribution of AUC values for MCI conversion prediction within 12, 24, 36, 48 and 60 months using different sets of features.

References

[1] D. A. Llano, G. Laforet, and V. Devanarayan, "Derivation of a new ADAS-cog composite using tree-based multivariate analysis," *Alzheimer Disease & Associated Disorders*, vol. 25, no. 1, pp. 73–84, 2011.

[2] Y. Huang, K. Ito, C. B. Billing, and R. J. Anziano, "Development of a straightforward and sensitive scale for mci and early ad clinical trials," *Alzheimer's & Dementia*, vol. 11, no. 4, pp. 404 – 414, 2015.

[3] N. Raghavan, M. N. Samtani, M. Farnum, E. Yang, G. Novak, M. Grundman, V. Narayan, A. DiBernardo, and A. D. N. Initiative, "The ADAS-Cog revisited: Novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials," *Alzheimer's & Dementia*, vol. 9, no. 1S, pp. S21–S31, 2013.

[4] T. Pereira, F. L. Ferreira, S. Cardoso, D. Silva, A. de Mendonça, M. Guerreiro, S. C. Madeira, and for the Alzheimer's Disease Neuroimaging Initiative, "Neuropsychological predictors of conversion from mild cognitive impairment to Alzheimer's disease: a feature selection ensemble combining stability and predictability," *BMC Medical Informatics and Decision Making*, vol. 18, p. 137, Dec 2018.