**Supplementary Materials**

1. **Methods**
   1. **Cognitive Diagnoses**

CN individuals from ADNI and OASIS had no significant impairment in memory or cognitive functions or activities of daily living, and no significant memory concern. In ADNI, a differentiation of CN and SCD was only made starting from ADNI-2, causing us to exclude individuals from ADNI-1 due to uncertain diagnosis. In OASIS, SCD individuals who reported memory concerns or whose spouse or clinician mentioned concerns as assessed in the USD-b9 questionnaire (n=10), were discarded from our sample due to the small sample size. To be considered SCD, either the study participant, an informant, or the clinician (in ADNI)/the study participant (in DELCODE) reported a significant memory concern in the absence of objective impairment of memory of cognitive function. SCD in ADNI were recruited from the general population, whereas SCD in DELCODE were recruited from memory clinics. An MCI diagnosis was provided to individuals with measurable impairment in cognitive function in the absence of dementia or significant impairments of daily living.

* 1. **Pre-processing**

For FDG-PET, preprocessing was performed on average scans of the given time intervals using the Statistical Parametric Mapping 12 toolbox (SPM12; [www.fil.ion.ucl.ac.uk](http://www.fil.ion.ucl.ac.uk)) in MATLAB (r2021b, The MathWorks Inc). All FDG-PET scans were aligned to the anterior commissure/posterior commissure, and subsequently co-registered and normalized to the tissue probability map (TPM) template in standard MNI152 space. Finally, standardized uptake value ratios (SUVRs) were calculated (reference: pons11).

For MRI, first, denoising (spatial-adaptive Non-Local Means), spatial registration, bias correction and skull striping were applied. The scans were then segmented by an adaptive maximum a posteriori approach13 with a partial volume model14. The Geodesic Shooting Algorithm15 was used for the nonlinear transformation.

To extract regional averages from 90 cortical and sub-cortical brain regions, we used the AAL1 atlas. The spatial overlap of AAL1 and template images in standard space, such as TPM, is sub-optimal, therefore, to avoid strong influence of zero-valued voxels, we resized AAL1 to fit the TPM template by only considering those voxels or AAL which corresponded to a gray matter tissue probability of at least 0.3 in TPM. Given the limited spatial resolution of PET images compared to MRI, computation of non-zero means would have falsified our comparison between modalities.

* 1. **Outlier exclusion**

Outlier exclusion was performed in the outer cross-validation loop to ensure data quality in an automated manner. Interquartile ranges (IQRs) were inferred from the CNADNI training sets. Subjects outside 6xIQR were removed from the training and respective test sets of CNADNI and CNOASIS. Importantly, as previous works have shown, MCI and even SCD subjects show an advanced brain age, which likely translates to a reduced signal in age-relevant brain regions5. Thus, outlier exclusion was not applied to the patient samples.

* 1. **Hyperparameters**

For both the support vector, and relevance vector regression models, the following hyperparameters were available for tuning:

**Kernel**: [‘linear’, ‘rbf’, ‘poly’] # kernel type to be used in algorithm  
**Degree**: [2, 3] # degree if kernel is polynomial

For support vector regression, we additionally tuned the regularization parameter C. Unlike support vector regression, relevance vector regression does not involve a margin-based optimization problem that requires C to control the trade-off between fitting the training data and controlling the model complexity.  
**C**: [0.001, 0.01, 0.1, 1, 10, 100, 500] # regularization parameter (strength of regularization is inversely proportional to C)

1. **Bias correction**

Brain age is subject to a frequently reported bias, in which the brain age of older individuals is under- and the brain age of younger individuals is overestimated21, regardless of the data or method under consideration23. Here, bias correction parameters were estimated using a linear model21 in the validation set, and subsequently applied to all test sets. The final brain age was calculated using slope (ɑ) and an intercept (β) as follows:

1. **Measures of cognitive performance**

The ADNI-MEM combines several scores used to evaluate individuals’ memory performance from the Rey Auditory Verbal Learning Test, Alzheimer’s Disease Assessment Scale and Mini Mental State Exam. The ADNI-EF is a summary score of several executive function tasks: Category Fluency, Trails, Digit span backwards, Wechsler Adult Intelligence Scale-R Digit Symbol Substitution, Number Cancellation, and Clock Drawing items. Correlations of BAG with cognitive performance were tested against a Bonferroni-corrected α-level of .025 (0.05/2).

1. **Measures of AD neuropathology**

For AV45-PET, mean SUVR are publicly available from previous analyses26–29. CSF Aβ1-42, and p-Tau181/ Aβ1-42 were acquired via lumbar puncture and analyzed using the Roche Elecsys® immunoassays30. The number of tau PET scans already evaluated for SUVR in the current cohorts was too small to include this biomarker in the current analyses. Correlations of BAG with AD neuropathology were tested against a Bonferroni-corrected α-level of 0.0125 (0.05/2).

1. **Results**
   1. **Hyperparameters**

For reproducibility, Table SM1 outlines hyperparameter configurations of the final models.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table SM1. Final models with optimal hyperparameter configurations.** | | | | | |
|  |  | **Model** | **C** | **Kernel** | **Degrees** |
| **MRI** | Model 1 | RVR | NA | Rbf | NA |
| Model 2 | RVR | NA | Linear | NA |
| Model 3 | SVR | 1 | Linear | NA |
| Model 4 | RVR | NA | Rbf | NA |
| Model 5 | SVR | 1 | Linear | NA |
| **FDG-PET** | Model 1 | RVR | NA | Linear | NA |
| Model 2 | RVR | NA | Linear | NA |
| Model 3 | SVR | 1 | Rbf | NA |
| Model 4 | RVR | NA | Linear | NA |
| Model 5 | RVR | NA | Linear | NA |

* 1. **Bias correction**

Bias correction successfully removed bias in all samples except the MRI-based MCIADNI predictions.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table SM2. Remaining correlation between BAG and age after bias correction in all cohorts** | | | |
|  | Modality | *n* | Bias |
| **CNADNI** | MRI | 175+ | r = .016  95% CI [-.13, .16] |
| FDG-PET | 175+ | r = -.022  95% CI [--.17, .13] |
| **CNOASIS** | MRI | 49+ | r = -.149  95% CI [-.37, .09] |
| FDG-PET | 49+ | r = -.058  [-.29, .18] |
| **SCDADNI** | MRI | 102 | r = -.075  95% CI [-.27, .12] |
| FDG-PET | 102 | r = -.055  95% CI [-.25, .14] |
| **MCIADNI** | MRI | 595 | r = .140\*\*  95% CI [.06, .22] |
| FDG-PET | 595 | r = .030  95% CI [-.05, .11] |
| **SCDDELCODE** | FDG-PET | 88 | r = -.022  95% CI [-.23, .19] |
| **MCIDELCODE** | MRI | 80 | r = -.019  95% CI [-.24, .2] |
| **Notes.** Remaining bias (the correlation coefficient between BAG and age) after bias correction was assessed using Pearson correlation. +trend significant with α = 0.1, \* significant with α = 0.05, \*\* significant with α = 0.01 | | | |

* 1. **Accuracy of estimated brain age using Schaeffer and Tian atlas**

Brain age estimation using the composite Schaeffer and Tian atlas (216 regions of interest) was largely comparable to results obtained from the AAL atlas. We found no significant difference between modalities in the accuracy of brain age estimation in the CN cohorts (ADNI: t(342) = -1.56, p = .12, 95% CI [-0.88, 0.10]; OASIS: t(71) = -1.34, p = 0.18, 95% CI [-1.81, 0.35]). Moreover, accuracy of brain age estimation (MAE and r²) was lower in the external datasets compared to ADNI, thus again proving the cohort effect of brain age estimation frameworks (XX STATISTICS XX). Finally, the FDG-PET BAG of ADNI SCD patients was slightly higher compared to CN, which was not observed when BAG was derived from MRI (XX STATISTICS XX).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table SM2.** Accuracy of predicting chronological age from FDG-PET and MRI scans using the composite atlas. | | | | | | | | |  |  |
|  | CNADNI | | SCDADNI | | CNOASIS | | MCIADNI | | SCDDELCODE | MCIDELCODE |
|  | MRI | FDG-PET | MRI | FDG-PET | MRI | FDG-PET | MRI | FDG-PET | FDG-PET | MRI |
| *n* total | 173+ | 173+ | 102 | 102 | 41+ | 41+ | 595 | 595 | 88 | 80 |
| MAE | 2.94 | 2.55 | 2.44 | 2.45 | 3.31 | 2.58 | 3.92 | 2.43 | 2.57 | 4.07 |
| Range | [-9.7; 10.3] | [-12.9; 13.6] | [-6.9; 7.2] | [-5.5; 10.5] | [-7.4; 11.8] | [-6.7; 5.5] | [-7.9; 14.8] | [-11.2; 11.0] | [-4.9; 7.9] | [-6.9; 14.9] |
| ME | 0.14 | -0.08 | -0.08 | 0.23 | 0.75 | -0.9 | 2.61 | 0.5 | 1.7 | 2.85 |
| R² | 0.62 | 0.7 | 0.7 | 0.7 | -0.07 | 0.38 | 0.51 | 0.8 | 0.64 | 0.24 |
| *Notes.* +After outlier exclusion using CN train set (IQR > 6). | | | | | | | | | | |

* 1. **Feature Importance**

Most important regions were defined as mean(δ) - 2\*SD(δ) < δi < mean(δ) + 2SD(δ), with δ being the average weight coefficient average across all models of a modality with a linear kernel.  
Across the 5 MRI models with linear kernel, most important regions were:  
17Networks\_LH\_ContA\_PFCl\_2 (r = -.191, p = .0002)\*\*,   
*17Networks\_RH\_VisPeri\_StriCal\_1 (r = -.391, p = 2.157e-15)\*\**,   
*17Networks\_RH\_VisPeri\_ExStrSup\_3 (r = -.221, p = 1.338e-05)\*\**,   
17Networks\_LH\_SalVentAttnA\_ParMed\_1 (r = -.191, p = .0002)\*\*,   
17Networks\_LH\_DefaultB\_PFCd\_2v (r = -.099, p = .054),   
17Networks\_LH\_DefaultB\_PFCd\_4 (r = -.093, p = .068),   
*17Networks\_RH\_SalVentAttnA\_Ins\_1 (r = -.159, p = .002)\*\** and   
*17Networks\_RH\_ContB\_IPL\_2 (r = -.177, p = .0005)\*\**   
from the Schaefer atlas, as well as

*NAc.rh (r = -.256, p = 3.13e-07)\*\**,   
*GP.rh (r = .135, p = .008)\**,   
*PUT.rh (r = -.179, p = .0004)\*\*,   
CAU.rh (r = -.062, p = .2241), and   
HIP.rh (r = -.381, p = 1.267e-14) \*\**from the Tian atlas.

For the 4 FDG-PET models with linear kernel, most important regions were:  
*17Networks\_RH\_SalVentAttnB\_PFCmp\_1 (r = -.451, p < 2.2e-16)\*\*,   
17Networks\_RH\_TempPar\_4 (r = -.223, p = 9.416e-06)\*\*,   
17Networks\_RH\_DorsAttnA\_TempOcc\_1 (r = -.244, p = 1.185e-06)\*\*,  
17Networks\_RH\_DefaultA\_PFCm\_1 (r = -.372, p = 3.856e-14)\*\*,   
17Networks\_RH\_DefaultA\_PFCm\_2 (r = -.323, p = 7.254e.11)\*\*, and   
17Networks\_RH\_DefaultC\_Rsp\_1* *(r = -.291, p = 5.192e-09)\*\**  
from the Schaefer atlas and

*NAc.rh (r = -.210, p = 3.096e-05)\*\*,   
GP.rh (r = -.152, p = .003)\*\** and   
*CAU.rh (r = -.410, p < 2.2e-16)\*\**   
from the Tian atlas.

Right hemispheric regions in italics for visibility. Pearson’s r and p values are yielded from correlating mean signal in these regions with age in the whole sample of cognitively normal individuals. \* significant, \*\* significant after Bonferroni correction (αMRI = .00385, αFDG-PET = 0.00555).

* 1. **Prediction of Cognitive Outcome**

In the CN+SCIADNI sample 2, 217 individuals remained stable until year two, while 24 obtained a diagnosis of cognitive impairment (MCI or dementia) within two years. Consequently, a subsample of 24 stables and all 24 decliners constituted the subsample for prediction of cognitive outcome in CN+SCDADNI. We found that, holding all other predictor variables constant, FDG-PET-derived BAG and APOE-ε4 carriership significantly, and years of education marginally predicted cognitive outcome after two years. The odds of a cognitive impairment diagnosis within two years were increased by 31% (95% CI [1.047, 1.739], *p* = .032) for every FDG-PET-derived BAG year. Moreover, the odds of developing cognitive impairment were increased by seven-fold with a positive APOE-ε4 carriership status (95% CI [1.209, 67.541], *p* = .46). Finally, a trend was observed that with every additional year of education, the odds of cognitive impairment were decreased by 23% (95% CI [0.573, 1.001], p = .065). To obtain a cutoff for prognoses of cognitive impairment, we fit a logistic regression model on cognitive outcome by BAG on FDG-PET. The intersection of the curve with 50% probability of receiving such a diagnosis was at 0.32 years FDG-PET BAG. In the current sample, this yielded a sensitivity of 75% and a specificity of 67%. In DELCODE, this cutoff enabled to correctly identify all eight decliners (sensitivity = 100%), while most stable individuals would be falsely predicted to develop cognitive impairment (specificity = 8%).

In sample 2 of MCIADNI patients, 93 patients converted to dementia, while 276 remained stable. Consequently, a subsample of 93 stables and all 93 decliners constituted the subsample for prediction of cognitive outcome. Holding all other predictor variables constant, MRI-derived BAG, a positive amyloid status in CSF, and APOE-ε4 carriership significantly predicted cognitive outcome after two years. With every one-year increase in BAG on MRI, the odds of converting to MCI or dementia were increased by 40% (95% CI [1.208, 1.655]), while a positive amyloid status and APOE-ε4 carriership both increased those odds by threefold (95% CIABETA+ [1.114, 8.748]; 95% CIAPOE+ [1.336, 6.767]), respectively. The intersection of the curve with 50% probability of receiving a diagnosis of dementia within two years was at 1.88 years of MRI BAG. In the current MCIADNI subsample, stratification by this cutoff yielded a sensitivity of 76% and a specificity of 70%. In the MCIDELCODE sample the cutoff had a sensitivity of 69% and a specificity of 71%.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table SM2. Coefficients of predicting cognitive outcome in CN without MRI BAG as predictor | | | | |
|  | **Estimate** | **Std. Error** | **Z value** | **Pr (>|z|)** |
| **FDG-PET BAG [Years]** | 0.25011 | 0.09655 | 2.590 | 0.00959 \*\* |
| **ABETA=NA** | -0.49955 | 0.78849 | -0.634 | 0.52637 |
| **ABETA=+** | -0.23647 | 0.72048 | -0.328 | 0.74275 |
| **APOE-e4=+** | 2.07650 | 0.94512 | 2.197 | 0.02801 \* |
| **Education [Years]** | -0.16003 | 0.11289 | -1.418 | 0.15633 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table SM3. Coefficients of predicting cognitive outcome in CN without FDG-PET BAG as predictor | | | | |
|  | **Estimate** | **Std. Error** | **Z value** | **Pr (>|z|)** |
| **MRI BAG [Years]** | 0.050259 | 0.113422 | 0.443 | 0.6577 |
| **ABETA=NA** | -0.004879 | 0.714663 | -0.007 | 0.9946 |
| **ABETA=+** | 0.188144 | 0.648534 | 0.290 | 0.7717 |
| **APOE-e4=+** | 1.292239 | 0.782596 | 1.651 | 0.0987 |
| **Education [Years]** | -0.109366 | 0.101333 | -1.079 | 0.2805 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table SM4. Coefficients of predicting cognitive outcome in MCI without MRI BAG as predictor | | | | |
|  | **Estimate** | **Std. Error** | **Z value** | **Pr (>|z|)** |
| **FDG-PET BAG [Years]** | 0.188953 | 0.058779 | 3.215 | 0.00131 \*\* |
| **ABETA=NA** | 1.257901 | 0.531529 | 2.367 | 0.01795 \* |
| **ABETA=+** | 1.783196 | 0.457962 | 3.894 | 9.87e-05 \*\*\* |
| **APOE-e4=+** | 1.047244 | 0.333144 | 3.144 | 0.00167 \*\* |
| **Education [Years]** | -0.004455 | 0.059747 | -0.075 | 0.94057 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table SM5. Coefficients of predicting cognitive outcome in MCI without FDG-PET BAG as predictor | | | | |
|  | **Estimate** | **Std. Error** | **Z value** | **Pr (>|z|)** |
| **MRI BAG [Years]** | 0.43237 | 0.07523 | 5.747 | 9.08e-09 \*\*\* |
| **ABETA=NA** | 0.48095 | 0.58592 | 0.821 | 0.41174 |
| **ABETA=+** | 1.42208 | 0.48662 | 2.922 | 0.00347 \*\* |
| **APOE-e4=+** | 1.09720 | 0.36234 | 3.028 | 0.00246 \*\* |
| **Education [Years]** | -0.04905 | 0.06627 | -0.740 | 0.45924 |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1