Grey matter volume and CSF biomarkers predict neuropsychological subtypes of MCI

Jeremy Lefort-Besnard, Mikaël Naveau, Nicolas Delcroix, Leslie M. Decker, Fabien Cignetti

Grey matter volume and CSF biomarkers predict neuropsychological subtypes of MCI

Lefort-Besnard J.[1], Naveau M.[2], Delcroix N.[2], Decker L.M[1,3,*], Cignetti F.[4,*], for the Alzheimer's Disease Neuroimaging Initiative[+]

[1] Normandie Univ, UNICAEN, INSERM UMR-S 1075, COMETE, 14000 Caen, France,
[2] Normandie Univ, UNICAEN, CNRS, CEA, INSERM, UAR 3408, GIP Cyceron, 14000 Caen, France,
[3] Normandie Univ, UNICAEN, CIREVE, Caen, France,
[4] Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France,
[*] These authors contributed equally to this work

[+] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at:http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Corresponding authors:
Dr. Leslie Marion DECKER,
UMR-S 1075, INSERM / Université de Caen Normandie
COMETE – Mobilités : Vieillissement, Pathologie, Santé
Pôle des Formations et de Recherche en Santé
2, Rue des Rochambelles, 14032 Caen Cedex 5
Electronic address: leslie.decker@unicaen.fr
Phone: +33 (0)6 70 40 58 44
Fax: +33 (0)2 31 56 82 19

Dr. Fabien CIGNETTI,
UMR 5525, CNRS / Université Grenoble Alpes
TIMC - Technique de l'Ingénierie Médicale et de la Complexité
Bâtiment Jean Roget
Faculté de Médecine et Pharmacie
Place du Commandant Nal, 38700 La Tronche
Electronic address: fabien.cignetti@univ-grenoble-alpes.fr

ABSTRACT

We demonstrated that mild cognitive impairment (MCI) participants of the ADNI database (N=640) can be discriminated into 3 coherent and neuropsychologically-defined subgroups. Our clustering approach revealed an amnestic MCI, a mixed MCI and a false positive subgroup. Furthermore, we investigated the neurobiological foundation of these automatically extracted MCI subgroups. Classification modelling exposed that specific predictive features can be used to differentiate amnestic and mixed MCI from healthy controls: CSF $A\beta_{1\text{-}42}$ concentration for the former and CSF $A\beta_{1\text{-}42}$ concentration, tau concentration as well as cortical atrophies (especially in the temporal and occipital lobes) for the latter. In contrast, false positive participants exhibited an identical profile to healthy participants in terms of cognitive performance, brain structure and CSF biomarker levels. Our comprehensive data-analytics strategy provide further evidence that multimodal neuropsychological subtyping is both clinically and neurobiologically meaningful.

HIGHLIGHTS

- Our clustering approach revealed an amnestic MCI, a mixed MCI and a false positive subgroup within the ADNI-defined MCI individuals.
- We investigated the neurobiological foundation of these neuropsychologically-defined MCI subtypes.
- Classification modelling exposed that specific predictive features can be used to differentiate amnestic and mixed MCI from healthy controls.
- However, a subgroup of ADNI-defined MCI individuals displayed a profile similar to healthy participants in terms of brain structure, CSF biomarker levels, and cognitive performance.
- These results suggest that neuropsychologically-defined MCI subtypes are neurobiologically grounded.

KEYWORDS

MCI subtypes; neuropsychological profile; Grey matter; CSF biomarker; ADNI; Machine learning

INTRODUCTION

Mild cognitive impairment (MCI) is considered a transitional stage between normal aging and Alzheimer's disease (AD). There is evidence that around 10-15 percent of MCI patients progress to AD each year, compared to 1–2 percent in the healthy older adult population (Alzheimer's Association 2019; Anderson 2019). However, there is considerable heterogeneity among the MCI-diagnosed individuals and not all of them are at risk for developing AD dementia later in life. Some patients develop non-AD dementia or other neuropsychiatric diseases (Slot et al. 2019). Others remain stable with respect to neuropsychological performance (Overton, Pihlsgård, and Elmståhl 2019), or even revert to normal cognitive functioning (Thomas, Edmonds, et al. 2019). There is also a high rate of misdiagnosis using conventional diagnostic criteria based on the DSM-5, with many

‘false-positive’ MCI cases (Edmonds et al. 2019). This heterogeneity of MCI have led the researchers to place great emphasis on subtyping or risk stratification of MCI patients to identify those at increased risk of developing AD and who constitute the optimal target population for therapeutic interventions (Dams-O’Connor et al. 2021; Winblad et al. 2016).

A common subtyping approach is to classify MCI individuals based on their neuropsychological test scores. Early on, MCI were staged into early and late MCI based on their level of impairment on one memory measure, with the latter being more impaired than the former. This “classical criteria” approach can be seen in the North American Alzheimer’s Disease Neuroimaging Initiative (ADNI) and in other samples (eg., Jessen et al. 2014). This approach has proven to be useful for staging MCI severity by demonstrating a higher risk of conversion to AD in individuals with late MCI compared to those with early MCI. However, there are also a number of limits with this approach including the unreliability of using a single neuropsychological test score to form subgroups, resulting in false positive MCI cases (Edmonds et al. 2019; Thomas, Eppig, et al. 2019), as well as the low sensitivity for detecting non-amnestic forms of MCI (Jak et al. 2009). Researchers then developed a “comprehensive criteria” from which multiple subtypes of MCI were identified based on performance on several tests covering a number of cognitive domains (eg., Clark et al. 2013; Jak et al. 2009; Bondi et al. 2014). They consistently revealed an amnestic subtype (impaired memory), a language or dysnomic subtype (impaired language), and a mixed subtype (impaired memory, executive function, attention, verbal fluency, and visuospatial function). It should be noted that, in some studies, the dysexecutive subtype is distinguished from the mixed subtype, with memory being affected only in the latter one; while, in other studies, the mixed subtype is alternately labelled ‘dysexecutive’ or ‘mixed’ depending on the authors, even when referring to a subgroup with substantial impairment in overall cognitive performance, including memory. In the present study, this specific group will be referred to as ‘mixed MCI’. Interestingly, the mixed MCI subtype has been repeatedly reported to have a higher rate of progression to AD dementia than the other subtypes. More recently, this finding has been consolidated by studies that empirically derived the exact same

subtypes (i.e, amnestic and mixed) using cluster analysis performed on neuropsychological test data (Machulda et al. 2019; Junquera et al. 2019; Blanken et al. 2020; Edmonds et al. 2016).

Several studies further characterized the above neuropsychologically-defined MCI subtypes in terms of their underlying AT(N) biomarkers, namely cerebrospinal fluid (CSF) beta amyloid deposition ('A') and pathologic tau ('T'), and neurodegeneration ('N') as assessed from structural MRI. The objective using the AT(N) framework for AD research (Jack et al. 2016) was to better understand the potential etiologic distinctions underlying the MCI subtypes. Overall, patterns of grey matter atrophy among the MCI subtypes were found to correspond to their profiles of cognitive impairment. Amnestic MCI individuals were reported to have smaller hippocampi (He et al. 2009). Medial temporal lobe thinning was found in both the amnestic and dysnomic subtypes (Whitwell et al. 2007; Edmonds et al. 2016). Lateral temporal lobe atrophy was also found in the dysnomic subtype. A widespread pattern of grey matter atrophy spanning parietal, temporal, and frontal regions was reported in the mixed MCI subtype (Dickerson and Wolk 2011; Edmonds et al. 2016). Regarding CSF biomarkers (i.e., p-tau and $A\beta_{1-42}$ level), the mixed subtype showed a greater proportion of individuals with positive CSF AD biomarkers than the dysnomic and amnestic subtypes (Edmonds, Delano-Wood, Clark, et al. 2015). In sum, these results tend to support the idea that MCI subtypes are rather homogeneous in terms of their biological and cerebral injury biomarkers.

A recent study by Kwak and colleagues (2021) addressed the opposite question as to whether heterogeneity in brain atrophy patterns of MCI individuals could allow identification of biologically and clinically meaningful subgroups. They reported one MCI subgroup in which the pattern of brain atrophy resembled that of AD patients (MCI-AD) and another MCI subgroup in which grey matter was similar to that of healthy individuals (MCI-CN). The rate of progression to AD for the MCI-AD subgroup was higher than for the MCI-CN. In terms of biological features, they reported marked differences between MCI-AD and MCI-CN subgroups, including especially more elevated tau and beta-amyloid burden in MCI-AD compared to MCI-CN. On the other hand, they found only a limited

degree of overlap between these two MRI-derived (atrophy-centered) subgroups and those empirically derived from neuropsychological test scores, including the amnestic, dysnomic and mixed ones. Thus, whether or not neuropsychological profiles of patients with MCI correspond to real distinct biological subtypes is still an open question.

In the present study, we pursue the question of the correspondence between MCI subtypes derived from neuropsychological assessment and their underlying patterns of neurodegeneration and CSF biomarker composition. For this purpose, using the ADNI data (640 MCI individuals and 326 healthy controls), we investigated the accuracy with which brain (i.e. grey matter) atrophy on the one hand and CSF beta amyloid and tau levels on the other hand, can predict neuropsychological subtypes of MCI. If predictive models derived from AT/N biomarkers perform well in classifying neuropsychological profiles of MCI, then such findings will provide compelling evidence of concordance between neuropsychological and neurobiological subtypes. More broadly, the study will provide valuable information about the neuropsychological and neurobiological fingerprintings of MCI, and, by extension, about the need (or not) to profile patients on the basis of multi-modal assessments.

METHODS

Participants

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Written informed consent was obtained from all participants or authorized representatives participating in the study. For more information, including criteria for eligibility, see http://www.adni-info.org. To be included in this work, each participant must have a status of mild cognitive impairment (MCI) or cognitively normal (CN). CN participants showed no signs of depression, mild cognitive impairment, or dementia. ADNI criteria for MCI were : i) subjective memory concern as reported by the subject, their study-partner or clinician, ii) abnormal memory function documented by scoring within education-adjusted ranges on delayed free recall of Story A from the WMS-R Logical Memory II subtest, iii) Mini–Mental State Examination (MMSE) score between 24 and 30, iv) global Clinical Dementia Rating (CDR) score of 0.5, with a Memory Box score of at least 0.5, and v) general cognition and functional performance sufficiently preserved so that a diagnosis of AD could not be made. Included participants must also have an exploitable T1 scan (i.e., the image successfully passed the preprocessing steps as well as visual quality assessment), an exploitable level of CSF biomarkers (no missing or not-a-number quantity), as well as an exploitable score on each questionnaire used in our study (no missing or not-a-number scores). A total of 966 participants met these conditions and thus were included in our study. See table 1 for more information.

Neuropsychological assessments

All MCI subjects in ADNI underwent a neuropsychological assessment at baseline (visit at one month from the screening in the ADNI protocol). The ADNI database provided the raw results of this assessment. For our study, we selected a list of neuropsychological tests according to two criteria : i) the test scores must not be missing and be a valid value, and ii) the test scores must have been used in previous studies using clustering (L. Q. Park et al. 2012; Edmonds et al. 2019) in order to allow

comparison of results. Neuropsychological test scores meeting these two criteria were included in our analysis. These tests included three measures of language : Animal Fluency Test, Boston Naming Test, Naming Object and Fingers Task of the Alzheimer's Disease Assessment Scale-Cognitive Subscale / ADAS-Cog (Rosen, Mohs, and Davis 1984), two measures of executive function : Trial Making Test: score A and score B minus A, two measures of visuo-spatial ability : Constructional Praxis Task and Ideational Praxis Task of the ADAS-Cog, and seven measures of memory : Word Recognition Task of the ADAS-Cog, Logical Memory Scale II (Chelune, Bornstein, and Prifitera 1990), and short delayed recall, long delayed recall, recognition, learning and forgetting items of the Rey Auditory Verbal Learning Test -RAVLT- (Rey 1958). Neuropsychological test scores, for which a lower score represents better performance, were multiplied by minus one, so that a higher score represents better performance. All scores were then transformed into z-scores by mean centering and unit-variance scaling.

Image acquisition

Processing: The structural brain image was acquired for all participants (n=966) with an anatomical 3D T1-weighted MPRAGE sequence. The sequence specifications of ADNI 1 session were TR = 3000 ms, TE = 3.6 ms, FoV = 192 × 192 mm2, flip angle = 8°, voxel resolution = 1.3 × 1.3 ×1.3 mm^3, and for ADNI 2 session TR = 2300 ms, TE = 3 ms, FoV = 256 × 256 mm^2, flip angle = 9°, voxel resolution = 1 × 1 × 1 mm^3. The brain tissue was segmented into grey matter, white matter and cerebrospinal fluid. Structural MRI data were preprocessed using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) toolbox implemented in Matlab 2022a (MathWorks, Inc., Natick, MA) to derive voxel-wise grey matter volumes for each subject. For a precise spatial normalization into standard (MNI), the Diffeomorphic Anatomic Registration Through Exponentiated Linear algebra algorithm (DARTEL) (Klein et al. 2009) was performed. Standard settings of SPM12 were used for the preprocessing steps (DARTEL normalization to the ICBM-152 template, affine and non-linear spatial normalization). The images were segmented into grey matter, white matter, and cerebrospinal fluid, and modulated with Jacobian determinants. Finally, the

modulated grey matter images were smoothed with an 8 mm isotropic FWHM Gaussian kernel. Volume extraction: Using the probabilistic Harvard-Oxford Cortical Structural lateralized atlas (RRID:SCR_001476) available from Scikit-learn (Pedregosa et al., 2011, using the argument 'cort-maxprob-thr25-2mm'), quantitative measures of grey matter volume were extracted within the 96 macroscopic brain structures labeled in this atlas in every participant. For the extraction of relevant signal from the structural brain data, the total of 96 regions served as topographic masks to sum the volume information across the voxels belonging to a given region. All region-wise structural grey matter volumes were transformed into z-scores by mean centering and unit-variance scaling. Variance explained by total intracranial volume (TIV) and age were regressed out based on a glm approach (Friston et al. 1994).

Cerebrospinal fluid (CSF) biomarkers collection

The ADNI database provided the raw CSF levels of amyloid ß plaques ($Aß_{1-42}$), total tau (tau) and tau phosphorylated at threonine 181 (ptau). In this work, these 3 biomarkers levels were recorded for each included participant. They were selected according to the AT(N) framework which was proposed to differentially assess the likelihood of progression to AD dementia at the MCI stage. “A” refers to β-amyloid deposition ($Aβ_{1-42}$), “T” refers to pathologic tau, and “N” to neurodegeneration (Jack et al. 2016). More details of the CSF collection and measurements in the ADNI can be found in Shaw and colleagues (2009). All biomarkers were transformed into z-scores by mean centering and unit-variance scaling.

Identifying hidden group structure: hierarchical clustering

We applied a hierarchical clustering algorithm (agglomerative) to automatically partition patient neuropsychological profiles into homogeneous groups using the standardized (z-scored) neuropsychological scores from all MCI participants (n = 640). Hierarchical clustering is a general family of clustering algorithms that build nested clusters by merging or splitting them successively (Kärkkäinen et al. 2020). This hierarchy of clusters is represented as a tree (or dendrogram, see Figure 1). The root of the tree is the unique cluster that gathers all the samples, the leaves being the

clusters with only one sample. Here, agglomerative clustering was performed using a bottom up approach: each observation starts in its own cluster, and clusters are successively merged together. The metric used for the merge strategy was the sum of squared differences within all clusters (Ward's method), which here was minimized. It is a variance-minimizing approach and in this sense is similar to the k-means objective function but tackled with an agglomerative hierarchical approach. In contrast to previous approach, agglomerative clustering is a method identifying one-to-many mappings (Bzdok and Yeo 2017): each patient is a member of exactly one group. We used "NbClust"(Charrad et al. 2014), an established R package that simultaneously applied 30 cluster validity metrics. This approach provided complementary indications on the number of groups most supported by the patient data. That is, several clustering schemes were evaluated while varying the number of clusters, to help determining the most appropriate number of clusters for our dataset. These metrics included for example the Duda index, the C-index, and the Gamma index. Please, see the reference above for the full list of metrics. Among the 30 metrics and according to the majority rule, the best number of clusters was 3. Therefore, three groups of patients with distinct neuropsychological profiles were automatically extracted as it provided a useful fit to our clinical sample.

Risk ratio of developing Alzheimer

Risk ratio (RR) was used to assess the risk of developing Alzheimer disease in each extracted MCI subgroups compared to controls. RR was defined as : RR = Cle / Clu where Cle is the cumulative incidence in the exposed group (i.e., each MCI subgroup), and Clu is the cumulative incidence in the unexposed goup (i.e., the control group).

Machine learning prediction of cluster membership from structural brain measures

The relative importance of grey matter volumes to predict membership in each MCI cluster versus control group was analyzed capitalizing on a pattern-learning algorithm L2-penalized logistic regression (Hastie et al. 2009). Unlike the common logistic regression, the L2-penalized logistic regression variant has an additional constraint used to reduce the chances of overfitting, which can

render the models' prediction of future observation unreliable. The L2-penalized logistic regression estimated the separating hyperplane (i.e., a linear function) yielding out-of-sample accuracies for distinguishing between MCI patients of each cluster and healthy participants. Model-fit and accuracy estimation were carried out as a 5-fold cross-validation procedure. Class imbalance, if present, was handled by changing the class-weight of the scikit-learn logistic regression API. The “balanced” mode uses the class membership to automatically adjust weights inversely proportional to class frequencies. The outcome to be predicted was defined by being healthy (0) or being an MCI patient from one of the three extracted clusters (1). In other words, three models were adjusted using grey matter volumes as input with a first model predicting cluster-derived normal versus controls, a second model predicting amnestic MCI versus controls, and a third one predicting mixed MCI versus controls. This way of engineering transformed a four-class problem into 3 two-class problems. In sum, this quantitative investigation detected if grey matter volume would be predictive of cluster belonging.

Machine learning prediction of cluster membership from CSF biomarkers measures

In order to allow for results comparison, the same algorithm was used in the previous setting and this one. This time, the L2-penalized logistic regression used three CSF biomarkers level ($A\beta_{1\text{-}42}$, t-tau and *p*-tau_{p181}) as feature input to estimate the separating hyperplane for distinguishing between MCI patients of each cluster and healthy participants. Again, we deployed a 5-fold cross-validation procedure and handled class-imbalance if present. The outcome to be predicted were exactly the same as in the previous setting. That is, being healthy (0) or being an MCI patient from one of the three extracted clusters (1). Thus, three models were adjusted using CSF biomarkers level as input with a first model predicting cluster-derived normal versus controls, a second model predicting amnestic MCI versus controls, and a third one predicting mixed MCI versus controls. In sum, this quantitative investigation detected if CSF biomarkers level would be predictive of cluster belonging.

Testing for significance

Three models based on grey matter volume and three other models based on CSF biomarkers level were conducted separately. Statistical significance for weights in each of the 6 final models was assessed based on (family wise error, multiple-comparison corrected) p-values derived through a rigorous non-parametric permutation approach using the model weights as the test statistic (Efron 2012; Nichols and Holmes 2002). Relying on minimal modeling assumptions, a valid null distribution was derived for the achieved weights resulting from the logistic regression fit. In 1000 permutation iterations, the input feature matrix (consecutively brain regions volume and CSF biomarkers level) was held constant, while the class membership (control versus each clusters) underwent participant-wise random shuffling. The empirical distribution generated in this manner reflected the null hypothesis of random association between the input features and class membership across participants. The beta coefficients were recorded in each iteration. The *p* values were obtained given the distance between the original beta values and the mean beta values obtained during the permutation iterations.

Testing for complex relationships among the grey matter volumes

L2-penalized logistic regression (cf. above) selected the most predictive ROI volumes for cluster membership. But this predictive algorithm is constrained to estimate additive effects between ROI volumes. To complement the regression model insights, we combined exploration of more sophisticated ROI-ROI relationships with the evaluation of prediction performance. In this way, we tested the hypothesis of existing non-linear relationships between the ROI volumes and their usefulness for prediction.

The goal here was to assess that a non-linear model would not reach a higher accuracy than the linear models in predicting mixed MCI cluster compared to controls. Note that we reported findings of this analysis only in one setting. That is, input features defined as grey matter volumes, and class defined as whether controls or MCI subjects belonging to the mixed MCI subgroup. given that in the main analysis (using logistic regression), only the accuracy between mixed MCI and controls was

significant. However, all the other results can be found in the additional (Supplementary Figure 1). We thus compared the performance of linear models to the performance of models able to exploit non-linear structure in the ROI volumes for predicting mixed MCI versus control. Model-fit and accuracy estimation were carried out using a 5-fold cross-validation procedure as implemented in the previous analysis. Class imbalance was again handled by changing the class-weight of the scikit-learn model API. Three linear models (ridge regression, logistic regression, and support vector machine) were benchmarked against three models allowing looking for non-linear interactions (k nearest neighbor, random forest and adaptive boosting). Note that we used the default regularization parameter for each algorithm.

Among the linear predictive pattern-learning algorithms, the ridge regression is commonly used as a shrinkage method. This model encourages small absolute weights on each ROI volume which emphasized the most predictive ROI volumes, while linear support vector machines mapped the ROI volumes as points in space so that volumes of separate categories (predictive of a mixed MCI or a control participant) were divided by a maximal gap between the individuals. Given that similar predictive ROIs were expected to emerge from these linear models, we also compared the obtained coefficients in a supplementary analysis.

Regarding the non-linear models, the k nearest neighbor estimator uses the k closest training examples (in our case, the closest participant brain structure) in the feature space, while the output is determined by a majority vote across these most similar training examples. In other words, the guessed class belonging (mixed MCI or control) for a given new participant can thus be derived from the brain structure of the k closest participants in the training set. Further, the random forest algorithm is an ensemble learning method that operates by constructing a multitude of decision trees and outputs its prediction estimate that is the committee decision across all trees. The cluster belonging for a given patient was thus derived based on the most consistently predicted outcome of the built decision trees. As the last non-linear prediction algorithm, the adaptive boosting algorithm starts by fitting a model on the dataset and then fits additional copies of that model on the same

dataset but where the weights of incorrectly judged instances are adjusted such that subsequent fine-tuning of models focusses more on difficult cases.

Code availability

Python was selected as the scientific computing engine. Scikit-learn (Pedregosa et al., 2011) provided efficient, unit-tested implementations of state-of-the-art statistical learning algorithms (http://scikit-learn.org). All analysis scripts of the present study are readily accessible to the reader online (https://github.com/JLefortBesnard/MCI_cluster_prediction).

RESULTS

Identifying hidden group structure: hierarchical clustering

To explore distinct subgroups related to cognitive test assessment patterns among MCI patients, each patient was automatically assigned to one dominant symptom constellation based on a number of cognitive tests. This data-driven exploration exposed 3 distinct symptom clusters (see Figure 1a and 1b) grouping the MCI patients: a mixed MCI subgroup (294 MCI patients) harbored low scores at almost every test (maximum 0.6 points on average), an amnestic MCI subgroup (207 MCI patients) scored low only on test assessing memory (maximum 1 point on average), and a cluster-derived normal subgroup (139 MCI patients) included MCI patients with a scoring profile virtually identical to controls (at least 1 point on average) except for one test, the logical memory scale II. The three subgroups were homogeneous in terms of sex and age (Supplementary Table 1).

Repartition of MCI patients developing Alzheimer

We examined the association between cluster memberships and AD. Compared to controls, only the amnestic MCI and mixed MCI subtypes exhibited a higher propensity to develop Alzheimer with twice as much risk for the mixed MCI subgroup (See Figure 2). The risk ratios for the amnestic and mixed MCI subgroup were respectivelly 4.38 ($p < 0.01$) and 7.52 ($p < 0.01$).

MCI cluster prediction based on grey matter volume

We explored the hypothesis that grey matter volume may predict affiliation to MCI subgroups. A regularized logistic regression was used to automatically identify regions of interest (ROI) with a high discriminant value for distinguishing controls from each MCI subgroup (see Figure 3a and 3c). Our analysis strategy revealed that only the mixed MCI subgroup was distinguishable from controls using grey matter volume. The mean accuracy of the averaged GMV models, incorporating only structural MRI data, was 72.13% with a standard error of 3.66%. There were 8 ROIs ($p<0.05$) that consistently contributed to predicting mixed MCI. These ROIs included the left occipital fusiform gyrus (weight = 0.73), the right (weight = 0.93) and left (weight = 1.11) parahippocampal gyrus anterior, the right cuneal cortex (weight = 0.58), the left middle posterior temporal gyrus (weight = 1.05), the left

occipital pole (weight = -1.10), and the right (weight = -0.84) and left (weight = 0.58) parahippocampal gyrus posterior (see Figure 4a).

MCI cluster prediction based on CSF biomarkers level

We then analyzed the relative importance of the level of Amyloid-β 1 to 42 peptide ($A\beta_{1\text{-}42}$), total tau (Tau), and tau phosphorylated (PTau) for distinguishing controls from each MCI subgroup (see Figure 3a and 3b). Our findings indicated a significant prediction accuracy for discriminating both the mixed MCI (71.60% +/- 4.67%) and amnestic MCI (63.42% +/- 5.19%) subgroup from controls. However, our model did not perform better than chance to distinguish controls from the cluster-derived normal subgroup. Only the weight associated with the level of $A\beta_{1\text{-}42}$ (coefficient = 0.40) was significant in predicting amnestic MCI patients while both the level of $A\beta_{1\text{-}42}$ (coefficient = 0.69) and Tau (coefficient = -0.75) were significant in predicting mixed MCI patients (see Figure 4b).

Testing for complex relationships among ROI volumes to distinguish mixed MCI from controls

Here, we wanted to make sure that a regularized logistic regression was more appropriate to our research setting than a model looking for non-linear effects. To test the hypothesis of existing non-linear interaction between the ROI volumes, we compared the prediction performance of mixed MCI compared to controls of different linear models (ridge regression, logistic regression, and support vector machine) to the prediction performance of different non-linear ones (k nearest neighbor, random forest, and adaptive boosting) (Fig. 5). Furthermore, we directly compared significant ROI weights obtained with the three linear models.

The three linear models — support vector machine, logistic regression, and ridge regression — obtained on average a similar performance with respectively 71.12% (+/- 3.24), 70.96% (+/- 3.64), and 72.42% (+/- 3.28) accuracy. Note that the tiny difference between our reported mean accuracy with logistic regression in the main analysis (72.13%) and the mean accuracy of the logistic regression in the benchmark analysis (70.96%) might be explained by the randomness in the cross validation split procedure (training-testing sets). On the other hand, the models looking for non-linear

interactions—k nearest neighbor, random forest and adaptive boosting—obtained on average 69.84% (+/- 4.34), 70% (+/- 3.24), and 67.74% (+/- 4.53).

As a general observation, the mean accuracy for the linear models (71.5%) was on average higher than the mean accuracy for the non-linear models (69.19%) in our sample. Furthermore, the variance was higher for the non-linear model performances (average standard deviation: 4.21%) than within the linear model performances (average standard deviation: 3.45%). These results suggest that the ROI volumes are predominantly predictive for mixed MCI based on their additive effects. However, it is important to note that our claim might be limited to the size of our sample. Indeed, non-linear models such as adaptive boosting might keep learning and thus may predict better with more data involved in the fitting.

All examined linear models showed virtually identical prediction performance. As a next step, we also wanted to evaluate the similarity of results (i.e, significant ROI weights) obtained between these three benchmarked linear models. Three ROI weights were systematically significant (Table 2). These ROI included the left posterior middle temporal gyrus; the left anterior parahippocampal gyrus, and the left occipital pole. Four ROI weights were reported as significant by the logistic regression and the support vector machine including the right cuneal cortex, the right anterior parahippocampal gyrus, the right posterior parahippocampal gyrus, and the left occipital fusiform gyrus. In other words, the three linear models capitalized on virtually similar ROIs to discriminate mixed MCI from control individuals.

DISCUSSION

Our study uncovered three partitions of discrete neuropsychologically-based MCI profiles. The first extracted MCI profile was similar to controls in terms of grey matter volumes, CSF biomarker levels, neuropsychological tests scores, as well as risk of developing Alzheimer's disease. The two other extracted MCI profiles showed regional grey matter volume reductions and abnormal CSF biomarker levels, allowing their discrimination from healthy individuals, and were also more at risk of developing Alzheimer's disease. These results support the conclusion that MCI subtypes derived from neuropsychological test scores have relatively clear biological – grey matter volume and CSF features – boundaries.

Subtyping of the MCI individuals using neuropsychological test scores

Our clustering method revealed two distinct, clinically meaningful, subgroups of MCI patients: a mixed MCI profile with low performance on memory, language, executive functioning, and visuo-spatial function, and an amnestic MCI profile with memory being the only impaired domain. A third profile also came out, with a neuropsychological profile similar to healthy participants. In general, these latent profiles are consistent with those reported in a number of previous studies that also applied clustering methods on a standardized set of neuropsychological tests measuring multiple domains of cognitive functioning (Eppig et al. 2017; Bondi et al. 2014; Edmonds, Delano-Wood, Clark, et al. 2015; Blanken et al. 2020). However, there are also studies that revealed additional profiles to the above-mentioned core MCI profiles, including dysexecutive, visuo-spatial or dysnomic profiles (Clark et al. 2013; Edmonds, Delano-Wood, Galasko, et al. 2015; Edmonds et al. 2016; Kwak et al. 2021). Factors that can explain such variations in the profiles are the criteria used to define MCI (prior to the clustering analysis) as well as the set of neuropsychological test scores included in the cluster analysis. For instance, in addition to the amnestic, mixed and cluster derived normal profiles, Clark and colleagues (2013) also reported dysexecutive and visuo-spatial subtypes. However, in their study, to be included as MCI was not based on the conventional diagnostic criterion (as in our present study), but instead on a specific criterion that required low performances on at least two

measures within a cognitive domain. In addition, they used items from the Wechsler Intelligence Scale and Wechsler Memory Scale while we used items from the ADAS-cog for assessing visuospatial functioning. Likewise, studies that reported dysnomic MCI subtype assessed language from animal fluency and 30-items Boston Naming Test (Kwak et al. 2021; Edmonds, Delano-Wood, Clark, et al. 2015), while we further included the Naming Object and Fingers Task of the ADAS-cog. An additional factor that may explain discrepancies between studies is the stability of the chosen clusters. We used multiple distance metrics (n=30, through the *nbclust* R package) to assess the most stable number of clusters in our sample while a single metric is usually chosen in other studies. Accordingly, we are confident that the choice of three clusters was the most consistent and optimal solution to get non-overlapping homogeneous groups. It is noteworthy that the higher risk of Alzheimer's dementia observed in the mixed MCI subgroup^p compared to the amnestic MCI subgroup and the normal risk level of the cluster-derived normal subgroup provided clinical validity to this clustering scheme. From a the clinical standpoint, the existence of these MCI subtypes illustrates the problem of diagnosing individuals on the basis of a single test in the memory domain, here the WMS-R Logical Memory Test in the ADNI study. First, it places side by side individuals with memory deficits only and individuals with multi-domain cognitive deficits, who are at different risk of progression to dementia. Second, it leads to false positive MCI diagnoses. Accordingly, and in line with previous recommendations (eg., Thomas, Eppig, et al. 2019; Edmonds, Delano-Wood, Clark, et al. 2015a; Jak et al. 2009), MCI diagnosis should include a multi-domain neuropsychological assessment and avoid the 'one test equals one domain' methodology.

Predictive value of regional grey matter volume to distinguish MCI patients from healthy individuals

We automatically assessed the extent to which each MCI subgroup could be differentiated from healthy participants based on regional grey matter volumes. Significant accuracy (72%) was obtained only for predicting the mixed MCI subgroup compared with the healthy participants. This finding suggests that the amnestic MCI subgroup and the cluster-derived normal subgroup have a brain structure more similar to healthy participants. Whereas the similarity of regional grey matter

volumes in the cluster-derived normal MCI subtype and in the cognitively normal group confirms the conclusion of previous studies drawn from cortical thickness (Edmonds et al. 2016; 2020; Blanken et al. 2019; L. R. Clark et al. 2013), that between the amnestic MCI subgroup and healthy participants may appear surprising.

Edmonds and colleagues (2016; 2020) found cortical differences between these two populations (i.e., amnestic MCI and controls) in the medial and lateral temporal lobe regions bilaterally as well as in some parietal and frontal regions. Machulda and colleagues (2020) also found differences in the medial temporal regions. Sun and colleagues (2019) reported decreased cortical thickness in medial orbitofrontal, parahippocampal and precuneus in amnestic MCI individuals. The discrepancy between these findings and ours is presumably due to difference in the methodology. Indeed, previous research focused on differences in brain structure in an explanatory fashion (i.e., modeling for inference using statistical significance) whereas in our study, we sought to find predictive patterns (i.e., modeling for prediction using cross-validation). In particular, there is evidence that successful prediction is often associated with a significant p-value, but not vice versa (Bzdok, Engemann, and Thirion 2020). Hence, previous brain structure impairments reported in amnestic MCI individuals may have rather poor predictive performance. Accordingly, brain structure should not be regarded as an indicator of main importance to detect amnestic MCI. This proposal is further supported by other studies, albeit with a rather small sample size (respectively 49 and 29 amnestic MCI), that used an explanatory approach and found no differences in brain structure between amnestic MCI individuals and controls (Xue et al. 2021; Yang et al. 2019).

Regarding mixed MCI, a total of 8 ROIs with decreased grey matter volume significantly contributed to the prediction performance. These ROIs included 3 regions from the occipital lobe, namely the left occipital fusiform gyrus, the left occipital pole, and the right cuneal cortex, and 5 regions from the temporal lobe including the right and left, anterior and posterior parahippocampal gyrus, and the left middle posterior temporal gyrus. Note that the weights of 3 of these 8 ROIs (the left middle posterior temporal gyrus, the left anterior parahippocampal gyrus, and the left occipital pole) were

systematically significant across the linear model benchmark analysis, suggesting a more robust predictive value for these 3 ROIs. Hence, atrophy in temporal and occipital regions had predictive value for delineating mixed MCI individuals from healthy participants. While widespread atrophy of temporal regions is a typical finding in mixed MCI (Edmonds et al. 2020; 2016; Machulda et al. 2020; Kwak et al. 2021; Johnson et al. 2010; Junquera Fernández et al. 2020; Ghosh, Libon, and Lippa 2014), occipital regions are usually only marginally affected in these individuals. Indeed, it is generally accepted so far that atrophy of the occipital cortex is characteristic of the later stages of Alzheimer's disease (Braak & Braak, 1991). Furthermore, impaired perfusion of the occipital lobe was proposed as a determining marker of dementia with Lewy bodies but not really of Alzheimer's disease (Hanyu et al. 2006; Prosser, Tossici-Bolt, and Kipps 2017). Hence, a striking and novel result of our study is that grey matter volume in occipital cortex is affected as early as the MCI stage. Interestingly, our findings go well with a recent conclusion that loss of grey matter integrity in the lateral and medial temporal lobes as well as in the occipital lobe is responsible of cognitive decline in vulnerable individuals that suffer the deleterious effects of elevated brain amyloid and poor vascular health (Saboo et al. 2022). Hence, atrophy of the temporal and occipital lobes may be very valuable marker of cognitively vulnerable individuals. On the other hand, the above-mentioned studies on mixed MCI reported significant grey matter loss in parietal and frontal regions, which were not found to be particularly predictively relevant in our study. Note that we emphasized our discussion on ROIs with the highest and most robust weights as automatically optimized by the model (i.e., L2-penalized logistic regression). However, it is important to keep in mind that the model chose to shrink a ROI coefficient because it brings little or no additional information on top of the other ROIs. Therefore, ROIs with small weights may still be related to the outcome. Nevertheless, the 8 ROIs found to have a significant weight in our study carried a substantial impact on distinguishing mixed MCI subjects from controls.

Predictive value of CSF biomarkers to distinguish MCI patients from healthy individuals

CSF biomarkers were useful to significantly differentiate (72% accuracy) between mixed MCI patients and healthy participants. In particular, the weights associated to the concentration of $A\beta_{1-42}$ and concentration of total tau were significant. That is, these two features were repeatedly informative for telling apart both groups. In patients, the concentration of $A\beta_{1-42}$ was lower while the concentration of total tau was higher compared to controls. CSF biomarkers were also effective to significantly distinguish amnestic MCI from healthy individuals (63% accuracy). This time, only the weight associated with the concentration of $A\beta_{1-42}$ was significant, suggesting that the concentration of $A\beta_{1-42}$ was the most contributing feature for the prediction. Overall, these results are in line with several previous studies that have examined biomarker characteristics in empirically derived subtypes of MCI and concluded that MCI patients with amnestic or executive symptoms have amyloid brain pathology and neuronal injury (Bangen et al. 2016; Edmonds et al. 2016; 2021; Eppig et al. 2017; Thomas, Eppig, et al. 2019; Edmonds, Delano-Wood, Clark, et al. 2015). Indeed, low CSF $A\beta_{1-42}$ level and high CSF tau level are strong predictors of the presence of pathologic amyloid plaques and neurofibrillary abnormalities in the brain (Tapiola et al. 2009). An important outcome of our work is that total tau was found to be a significantly informative feature to separate mixed MCI, but not amnestic MCI, from controls. Accordingly, both amnestic and mixed MCI subtypes would exhibit amyloid pathology while only the mixed subtype would have disrupted neuronal activity. This conclusion is also supported by a clinical interpretation of the concentrations of CSF $A\beta_{1-42}$ and total tau observed in our sample. In both subgroups, CSF $A\beta_{1-42}$ concentration was less than the cutoff of 192 pg/ml that is commonly used to identify the presence of amyloid pathology (Shaw et al. 2009). On the other hand, the cutoff of 93 pg/ml, which identifies disruption of neuronal activity (Shaw et al. 2009), was exceeded in the mixed subgroup only. Finally, it is also important to draw attention to the fact that above and beyond the above-mentioned impaired levels of CSF $A\beta_{1-42}$ and total tau in the MCI subgroups, both MCI subgroups were at higher risk to develop AD. This suggests a link between

CSF biomarkers and conversion to AD, as pointed out in earlier studies (Hansson et al. 2006; Mattsson et al. 2009; Insel et al. 2018; Ortega et al. 2019; Park et al. 2019).

MCI subgroups as distinct MCI phenotypes or distinct stages along the course of Alzheimer?

The mixed MCI subgroup was distinguished from healthy individuals through structural brain atrophy as well as CSF $A\beta_{1-42}$ and total tau abnormal levels, while the amnestic MCI subgroup was only separated from healthy controls through CSF $A\beta_{1-42}$ abnormal level. Importantly, mixed MCI individuals were at higher risk of conversion to Alzheimer disease than amnestic MCI individuals. Overall, these findings do concur with the amyloid cascade model of AD progression in which Aβ pathology (as measured by CSF $A\beta_{1-42}$ or amyloid PeT) appears first, followed by tau pathology (measured by CSF tau), then neuronal loss (measured by MRI) and then clinical symptoms (Jack Jr et al. 2010; Jack et al. 2013). This model has received strong support over the years (Balsis et al. 2018; Yasuno et al. 2021; Jack et al. 2010; van Rossum et al. 2012; Broadhouse, Winks, and Summers 2021; Han and Shi 2016; Weiner et al. 2015; X. Yang, Tan, and Qiu 2012; Nettiksimmons et al. 2014), although not all findings align with it and alternative scenarios have emerged where Aβ deposition, tau pathology, neuronal degeneration and cognitive loss aligned in a narrow time sequence (Edmonds, Delano-Wood, Galasko, et al. 2015; Braak et al. 2013). Hence, through the prism of the amyloid cascade model, our MCI subgroups would rather represent distinct stages along the course of AD, the disease progressing from the amnestic stage to the mixed stage. However, the question of whether amnestic and mixed MCI subgroups merely reflect different stages along the course of AD or correspond to distinct MCI phenotypes could only receive a definite answer by examining longitudinal data from the two subgroups. This could hopefully be achieved in future research.

LIMITATIONS

It is important to note that data from the ADNI 1 were acquired from a 1.5 T scanner while data from the ADNI 2 were acquired from a 3T scanner. It is an ongoing debate if scans acquired from different scanners can be merged. Many studies reported highly reproducible correspondence between volumes (Roche et al, 2013; Ho et al, 2010) while other studies suggested different methods to

increase consistency across field strengths (Keihaninejad et al 2010). Here, we have made the decision to similarly preprocess all scans. Additionally, we only used grey matter volume as neuroimaging data, other measure can be of interest such as tractography, white matter, cortical thickness, and so on. Another limitation of our study includes the use of one dataset (i.e., ADNI). ADNI is not a population-based study and there are strict inclusion and exclusion criteria for selection of participants, which can affect generalizability of our findings. Therefore, validating our models and outcomes in other population-based studies and clinical trials' data would be an important next step.

CONCLUSION

In summary, our research revealed 3 latent subgroups underlying MCI participants of the ADNI database: an amnestic MCI, a mixed MCI and a false positive subgroup. Leveraging on machine learning, our findings further suggest that MCI subtypes, extracted from multimodal neuropsychological approach, have proper biological and neurological characteristics. As such, our findings pave the way to fine-grained, biologically and neurologically meaningful MCI diagnosis. Finally, our results suggested that AD progression may start affecting memory and CSF biomarkers, followed by an alteration of brain structure and of the other cognitive functions. As such, multimodal neuropsychological subtyping, in addition to being clinically meaningful, is also biologically and neurologically meaningful. Furthermore, our results suggested that AD progression may start affecting memory and CSF biomarkers, followed by an alteration of brain structure and of the other cognitive functions.

Acknowledgements
The study was funded by the CARSAT (Caisse Régionale d'Assurance Retraite et de la Santé au Travail) of Normandie (France) and from donation to the Caen Normandie Santé foundation (Crédit Agricole Normandie and Normandie-Seine, Harmonie Mutuelle, société SAMMed), and co-funded by the European Union and the Normandie Region as part of the Operationnal FEDER program 2014-2020 Normandie.
Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI). See the supplementary for more information.

Declaration of Competing Interest
The authors have no actual or potential conflicts of interest.

Credit author statement

JLB: Study concept, plan of analysis, data management, image analysis, statistical analysis, literature search and review, write-up of parts of the manuscript, revision of the manuscript; MN: Study concept, plan of analysis, assistance with image and statistical analysis, literature search and review, revision of the manuscript; ND: Study concept, plan of analysis, assistance with image and statistical analysis, revision of the manuscript; FC: Study concept, plan of analysis, assistance with image and statistical analysis, literature search and review, write-up of parts of the manuscript, revision of the manuscript; LD: Study concept, plan of analysis, assistance with statistical analysis, literature search and review, revision of the manuscript.

Bibliography

Alzheimer's Association. 2019. '2019 Alzheimer's Disease Facts and Figures'. *Alzheimer's & Dementia* 15 (3): 321–87. https://doi.org/10.1016/j.jalz.2019.01.010.

Anderson, Nicole D. 2019. 'State of the Science on Mild Cognitive Impairment (MCI)'. *CNS Spectrums* 24 (1): 78–87. https://doi.org/10.1017/S1092852918001347.

Balsis, Steve, Lisa Geraci, Jared Benge, Deborah A Lowe, Tabina K Choudhury, Robert Tirso, Rachelle S Doody, and Alzheimer's Disease Neuroimaging Initiative. 2018. 'Statistical Model of Dynamic Markers of the Alzheimer's Pathological Cascade'. *The Journals of Gerontology: Series B* 73 (6): 964–73. https://doi.org/10.1093/geronb/gbx156.

Bangen, Katherine J., Alexandra L. Clark, Madeline Werhane, Emily C. Edmonds, Daniel A. Nation, Nicole Evangelista, David J. Libon, Mark W. Bondi, and Lisa Delano-Wood. 2016. 'Cortical Amyloid Burden Differences Across Empirically-Derived Mild Cognitive Impairment Subtypes and Interaction with APOE E4 Genotype'. *Journal of Alzheimer's Disease : JAD* 52 (3): 849–61. https://doi.org/10.3233/JAD-150900.

Blanken, Anna E., Shubir Dutt, Yanrong Li, Daniel A. Nation, and Alzheimer's Disease Neuroimaging Initiative. 2019. 'Disentangling Heterogeneity in Alzheimer's Disease: Two Empirically-Derived Subtypes'. *Journal of Alzheimer's Disease* 70 (1): 227–39.

Blanken, Anna E., Jung Yun Jang, Jean K. Ho, Emily C. Edmonds, S. Duke Han, Katherine J. Bangen, and Daniel A. Nation. 2020. 'Distilling Heterogeneity of Mild Cognitive Impairment in the National Alzheimer Coordinating Center Database Using Latent Profile Analysis'. *JAMA Network Open* 3 (3): e200413. https://doi.org/10.1001/jamanetworkopen.2020.0413.

Bondi, Mark W, Emily C Edmonds, Amy J Jak, Lindsay R Clark, Lisa Delano-Wood, Carrie R McDonald, Daniel A Nation, David J Libon, Rhoda Au, and Douglas Galasko. 2014. 'Neuropsychological Criteria for Mild Cognitive Impairment Improves Diagnostic Precision, Biomarker Associations, and Progression Rates'. *Journal of Alzheimer's Disease* 42 (1): 275–89.

Braak, Heiko, Henrik Zetterberg, Kelly Del Tredici, and Kaj Blennow. 2013. 'Intraneuronal Tau Aggregation Precedes Diffuse Plaque Deposition, but Amyloid-β Changes Occur before Increases of Tau in Cerebrospinal Fluid'. *Acta Neuropathologica* 126 (5): 631–41. https://doi.org/10.1007/s00401-013-1139-0.

Broadhouse, Kathryn M., Natalie J. Winks, and Mathew J. Summers. 2021. 'Fronto-Temporal Functional Disconnection Precedes Hippocampal Atrophy in Clinically Confirmed Multi-Domain Amnestic Mild Cognitive Impairment'. *EXCLI Journal* 20 (September): 1458–73. https://doi.org/10.17179/excli2021-4191.

Bzdok, Danilo, Denis Engemann, and Bertrand Thirion. 2020. 'Inference and Prediction Diverge in Biomedicine'. *Patterns* 1 (8): 100119. https://doi.org/10.1016/j.patter.2020.100119.

Bzdok, Danilo, and B. T. Thomas Yeo. 2017. 'Inference in the Age of Big Data: Future Perspectives on Neuroscience'. *NeuroImage* 155 (July): 549–64. https://doi.org/10.1016/j.neuroimage.2017.04.061.

Charrad, Malika, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs. 2014. 'NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set'. *Journal of Statistical Software* 61 (November): 1–36. https://doi.org/10.18637/jss.v061.i06.

Chelune, Gordon J., Robert A. Bornstein, and Aurelio Prifitera. 1990. 'The Wechsler Memory Scale—Revised'. In *Advances in Psychological Assessment: Volume 7*, edited by Paul McReynolds, James C. Rosen, and Gordon J. Chelune, 65–99. Advances in Psychological Assessment. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4613-0555-2_3.

Clark, David, Puneet Kapur, David Geldmacher, John Brockington, Lindy Harrell, and Daniel Marson. 2013. 'Latent Information in Verbal Fluency Lists Predicts Functional Decline in MCI: T2311.' *Annals of Neurology* 74.

Clark, Lindsay R., Lisa Delano-Wood, David J. Libon, Carrie R. McDonald, Daniel A. Nation, Katherine J. Bangen, Amy J. Jak, Rhoda Au, David P. Salmon, and Mark W. Bondi. 2013. 'Are Empirically-

Derived Subtypes of Mild Cognitive Impairment Consistent with Conventional Subtypes?' *Journal of the International Neuropsychological Society* 19 (6): 635–45.

Dams-O'Connor, Kristen, Patrick S.F. Bellgowan, Roderick Corriveau, Mary Jo Pugh, Douglas H. Smith, Julie A. Schneider, Keith Whitaker, and Henrik Zetterberg. 2021. 'Alzheimer's Disease-Related Dementias Summit 2019: National Research Priorities for the Investigation of Traumatic Brain Injury as a Risk Factor for Alzheimer's Disease and Related Dementias'. *Journal of Neurotrauma* 38 (23): 3186–94. https://doi.org/10.1089/neu.2021.0216.

Dickerson, Bradford C, and David A Wolk. 2011. 'Dysexecutive versus Amnesic Phenotypes of Very Mild Alzheimer's Disease Are Associated with Distinct Clinical, Genetic and Cortical Thinning Characteristics'. *Journal of Neurology, Neurosurgery, and Psychiatry* 82 (1): 45–51. https://doi.org/10.1136/jnnp.2009.199505.

Edmonds, Emily C., Lisa Delano-Wood, Lindsay R. Clark, Amy J. Jak, Daniel A. Nation, Carrie R. McDonald, David J. Libon, et al. 2015. 'Susceptibility of the Conventional Criteria for Mild Cognitive Impairment to False-Positive Diagnostic Errors'. *Alzheimer's & Dementia* 11 (4): 415–24. https://doi.org/10.1016/j.jalz.2014.03.005.

Edmonds, Emily C., Lisa Delano-Wood, Douglas R. Galasko, David P. Salmon, Mark W. Bondi, and Alzheimer's Disease Neuroimaging Initiative. 2015. 'Subtle Cognitive Decline and Biomarker Staging in Preclinical Alzheimer's Disease'. *Journal of Alzheimer's Disease: JAD* 47 (1): 231–42. https://doi.org/10.3233/JAD-150128.

Edmonds, Emily C., Joel Eppig, Mark W. Bondi, Kelly M. Leyden, Bailey Goodwin, Lisa Delano-Wood, Carrie R. McDonald, and Alzheimer's Disease Neuroimaging Initiative. 2016. 'Heterogeneous Cortical Atrophy Patterns in MCI Not Captured by Conventional Diagnostic Criteria'. *Neurology* 87 (20): 2108–16.

Edmonds, Emily C, Carrie R McDonald, Anisa Marshall, Kelsey R Thomas, Joel Eppig, Alexandra J Weigand, Lisa Delano-Wood, Douglas R Galasko, David P Salmon, and Mark W Bondi. 2019. 'Early versus Late MCI: Improved MCI Staging Using a Neuropsychological Approach'. *Alzheimer's & Dementia* 15 (5): 699–708.

Edmonds, Emily C., Denis S. Smirnov, Kelsey R. Thomas, Lisa V. Graves, Katherine J. Bangen, Lisa Delano-Wood, Douglas R. Galasko, David P. Salmon, and Mark W. Bondi. 2021. 'Data-Driven vs Consensus Diagnosis of MCI: Enhanced Sensitivity for Detection of Clinical, Biomarker, and Neuropathologic Outcomes'. *Neurology* 97 (13): e1288–99. https://doi.org/10.1212/WNL.0000000000012600.

Edmonds, Emily C., Alexandra J. Weigand, Sean N. Hatton, Anisa J. Marshall, Kelsey R. Thomas, Daniela A. Ayala, Mark W. Bondi, and Carrie R. McDonald. 2020. 'Patterns of Longitudinal Cortical Atrophy over 3 Years in Empirically Derived MCI Subtypes'. *Neurology* 94 (24): e2532–44. https://doi.org/10.1212/WNL.0000000000009462.

Edmonds, Emily C., Alexandra J. Weigand, Kelsey R. Thomas, Joel Eppig, Lisa Delano-Wood, Douglas R. Galasko, David P. Salmon, and Mark W. Bondi. 2018. 'Increasing Inaccuracy of Self-Reported Subjective Cognitive Complaints over 24 Months in Empirically-Derived Subtypes of Mild Cognitive Impairment'. *Journal of the International Neuropsychological Society : JINS* 24 (8): 842–53. https://doi.org/10.1017/S1355617718000486.

Efron, Bradley. 2012. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Vol. 1. Cambridge University Press.

Eppig, Joel S., Emily C. Edmonds, Laura Campbell, Mark Sanderson, Lisa Delano-Wood, and Mark W. Bondi. 2017. 'Statistically-Derived Subtypes and Associations with Cerebrospinal Fluid and Genetic Biomarkers in Mild Cognitive Impairment: A Latent Profile Analysis'. *Journal of the International Neuropsychological Society : JINS* 23 (7): 564–76. https://doi.org/10.1017/S135561771700039X.

Friston, Karl J., Andrew P. Holmes, Keith J. Worsley, J.-P. Poline, Chris D. Frith, and Richard SJ Frackowiak. 1994. 'Statistical Parametric Maps in Functional Imaging: A General Linear Approach'. *Human Brain Mapping* 2 (4): 189–210.

Ghosh, Sayantani, David Libon, and Carol Lippa. 2014. 'Mild Cognitive Impairment: A Brief Review and Suggested Clinical Algorithm'. *American Journal of Alzheimer's Disease & Other Dementias®* 29 (4): 293–302. https://doi.org/10.1177/1533317513517040.

Han, Pengcheng, and Jiong Shi. 2016. 'A Theoretical Analysis of the Synergy of Amyloid and Tau in Alzheimer's Disease'. *Journal of Alzheimer's Disease* 52 (4): 1461–70. https://doi.org/10.3233/JAD-151206.

Hansson, Oskar, Henrik Zetterberg, Peder Buchhave, Elisabet Londos, Kaj Blennow, and Lennart Minthon. 2006. 'Association between CSF Biomarkers and Incipient Alzheimer's Disease in Patients with Mild Cognitive Impairment: A Follow-up Study'. *The Lancet Neurology* 5 (3): 228–34. https://doi.org/10.1016/S1474-4422(06)70355-6.

Hanyu, Haruo, Soichiro Shimizu, Kentaro Hirao, Hidekazu Kanetaka, Hirofumi Sakurai, Toshihiko Iwamoto, Kiyoshi Koizumi, and Kimihiko Abe. 2006. 'Differentiation of Dementia with Lewy Bodies from Alzheimer's Disease Using Mini-Mental State Examination and Brain Perfusion SPECT'. *Journal of the Neurological Sciences* 250 (1): 97–102. https://doi.org/10.1016/j.jns.2006.07.007.

Hastie, Trevor, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Vol. 2. Springer.

He, Jing, Sarah Farias, Oliver Martinez, Bruce Reed, Dan Mungas, and Charles DeCarli. 2009. 'Differences in Brain Volume, Hippocampal Volume, Cerebrovascular Risk Factors, and Apolipoprotein E4 Among Mild Cognitive Impairment Subtypes'. *Archives of Neurology* 66 (11): 1393–99. https://doi.org/10.1001/archneurol.2009.252.

Insel, Philip S., Oskar Hansson, R. Scott Mackin, Michael Weiner, and Niklas Mattsson. 2018. 'Amyloid Pathology in the Progression to Mild Cognitive Impairment'. *Neurobiology of Aging* 64 (April): 76–84. https://doi.org/10.1016/j.neurobiolaging.2017.12.018.

Jack, Clifford R., David A. Bennett, Kaj Blennow, Maria C. Carrillo, Howard H. Feldman, Giovanni B. Frisoni, Harald Hampel, et al. 2016. 'A/T/N: An Unbiased Descriptive Classification Scheme for Alzheimer Disease Biomarkers'. *Neurology* 87 (5): 539–47. https://doi.org/10.1212/WNL.0000000000002923.

Jack, Clifford R., Jr, Heather J. Wiste, Prashanthi Vemuri, Stephen D. Weigand, Matthew L. Senjem, Guang Zeng, Matt A. Bernstein, et al. 2010. 'Brain Beta-Amyloid Measures and Magnetic Resonance Imaging Atrophy Both Predict Time-to-Progression from Mild Cognitive Impairment to Alzheimer's Disease'. *Brain* 133 (11): 3336–48. https://doi.org/10.1093/brain/awq277.

Jack, Clifford R., Heather J. Wiste, Stephen D. Weigand, David S. Knopman, Val Lowe, Prashanthi Vemuri, Michelle M. Mielke, et al. 2013. 'Amyloid-First and Neurodegeneration-First Profiles Characterize Incident Amyloid PET Positivity'. *Neurology* 81 (20): 1732–40. https://doi.org/10.1212/01.wnl.0000435556.21319.e4.

Jack Jr, Clifford R, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, and John Q Trojanowski. 2010. 'Hypothetical Model of Dynamic Biomarkers of the Alzheimer's Pathological Cascade'. *The Lancet Neurology* 9 (1): 119–28.

Jak, Amy J., Mark W. Bondi, Lisa Delano-Wood, Christina Wierenga, Jody Corey-Bloom, David P. Salmon, and Dean C. Delis. 2009. 'Quantification of Five Neuropsychological Approaches to Defining Mild Cognitive Impairment'. *The American Journal of Geriatric Psychiatry : Official Journal of the American Association for Geriatric Psychiatry* 17 (5): 368–75. https://doi.org/10.1097/JGP.0b013e31819431d5.

Jessen, Frank, Steffen Wolfsgruber, Birgitt Wiese, Horst Bickel, Edelgard Mösch, Hanna Kaduszkiewicz, Michael Pentzek, et al. 2014. 'AD Dementia Risk in Late MCI, in Early MCI, and in Subjective Memory Impairment'. *Alzheimer's & Dementia* 10 (1): 76–83. https://doi.org/10.1016/j.jalz.2012.09.017.

Johnson, Julene K., Judy Pa, Adam L. Boxer, Joel H. Kramer, Katie Freeman, and Kristine Yaffe. 2010. 'Baseline Predictors of Clinical Progression among Patients with Dysexecutive Mild Cognitive

Impairment'. *Dementia and Geriatric Cognitive Disorders* 30 (4): 344–51. https://doi.org/10.1159/000318836.

Junquera, Almudena, Estefanía García-Zamora, Mario Alfredo Parra, and S. Fernández-Guinea. 2019. 'Clustering Executive Functions Yields MCI Profiles That Significantly Predict Conversion to AD Dementia'. In *Alzheimer's Association International Conference*.

Junquera Fernández, Almudena, Estefanía García, Mario A. Parra, and Sara Fernández Guinea. 2020. 'Patterns of Brain Atrophy in Dysexecutive Amnestic Mild Cognitive Impairment Raise Confidence about Prodromal AD Dementia'. *Alzheimer's & Dementia* 16 (S6): e046365. https://doi.org/10.1002/alz.046365.

Kärkkäinen, Mikko, Mithilesh Prakash, Marzieh Zare, Jussi Tohka, and for the Alzheimer's Disease Neuroimaging Initiative. 2020. 'Structural Brain Imaging Phenotypes of Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD) Found by Hierarchical Clustering'. Edited by George T. Grossberg. *International Journal of Alzheimer's Disease* 2020 (November): 2142854. https://doi.org/10.1155/2020/2142854.

Klein, Arno, Jesper Andersson, Babak A. Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E. Christensen, et al. 2009. 'Evaluation of 14 Nonlinear Deformation Algorithms Applied to Human Brain MRI Registration'. *NeuroImage* 46 (3): 786–802. https://doi.org/10.1016/j.neuroimage.2008.12.037.

Kwak, Kichang, Kelly S. Giovanello, Andrea Bozoki, Martin Styner, and Eran Dayan. 2021. 'Subtyping of Mild Cognitive Impairment Using a Deep Learning Model Based on Brain Atrophy Patterns'. *Cell Reports Medicine* 2 (12): 100467. https://doi.org/10.1016/j.xcrm.2021.100467.

Machulda, Mary M., Emily S. Lundt, Sabrina M. Albertson, Walter K. Kremers, Michelle M. Mielke, David S. Knopman, Mark W. Bondi, and Ronald C. Petersen. 2019. 'Neuropsychological Subtypes of Incident Mild Cognitive Impairment in the Mayo Clinic Study of Aging'. *Alzheimer's & Dementia* 15 (7): 878–87. https://doi.org/10.1016/j.jalz.2019.03.014.

Machulda, Mary M., Emily S. Lundt, Sabrina M. Albertson, Anthony J. Spychalla, Christopher G. Schwarz, Michelle M. Mielke, Clifford R. Jack Jr., et al. 2020. 'Cortical Atrophy Patterns of Incident MCI Subtypes in the Mayo Clinic Study of Aging'. *Alzheimer's & Dementia* 16 (7): 1013–22. https://doi.org/10.1002/alz.12108.

Mattsson, Niklas, Henrik Zetterberg, Oskar Hansson, Niels Andreasen, Lucilla Parnetti, Michael Jonsson, Sanna-Kaisa Herukka, et al. 2009. 'CSF Biomarkers and Incipient Alzheimer Disease in Patients With Mild Cognitive Impairment'. *JAMA* 302 (4): 385–93. https://doi.org/10.1001/jama.2009.1064.

Nettiksimmons, Jasmine, Charles DeCarli, Susan Landau, and Laurel Beckett. 2014. 'Biological Heterogeneity in ADNI Amnestic Mild Cognitive Impairment'. *Alzheimer's & Dementia* 10 (5): 511-521.e1. https://doi.org/10.1016/j.jalz.2013.09.003.

Nichols, Thomas E., and Andrew P. Holmes. 2002. 'Nonparametric Permutation Tests for Functional Neuroimaging: A Primer with Examples'. *Human Brain Mapping* 15 (1): 1–25.

Ortega, Ricard L., Farida Dakterzada, Alfonso Arias, Ester Blasco, Alba Naudí, Francisco P. Garcia, and Gerard Piñol-Ripoll. 2019. 'Usefulness of CSF Biomarkers in Predicting the Progression of Amnesic and Nonamnesic Mild Cognitive Impairment to Alzheimer's Disease'. *Current Aging Science* 12 (1): 35–42. https://doi.org/10.2174/1874609812666190112095430.

Overton, Marieclaire, Mats Pihlsgård, and Sölve Elmståhl. 2019. 'Diagnostic Stability of Mild Cognitive Impairment, and Predictors of Reversion to Normal Cognitive Functioning'. *Dementia and Geriatric Cognitive Disorders* 48 (5–6): 317–29.

Park, Jung Eun, Kyu Yeong Choi, Byeong C. Kim, Seong-Min Choi, Min-Kyung Song, Jang Jae Lee, Jahae Kim, et al. 2019. 'Cerebrospinal Fluid Biomarkers for the Diagnosis of Prodromal Alzheimer's Disease in Amnestic Mild Cognitive Impairment'. *Dementia and Geriatric Cognitive Disorders Extra* 9 (1): 100–113. https://doi.org/10.1159/000496920.

Park, Lovingly Quitania, Alden L Gross, Donald G McLaren, Judy Pa, Julene K Johnson, Meghan Mitchell, and Jennifer J Manly. 2012. 'Confirmatory Factor Analysis of the ADNI Neuropsychological Battery'. *Brain Imaging and Behavior* 6 (4): 528–39.

Pedregosa, Fabian, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. n.d. 'Scikit-Learn: Machine Learning in Python'. *MACHINE LEARNING IN PYTHON*, 6.

Prosser, Angus MJ, Livia Tossici-Bolt, and Christopher M. Kipps. 2017. 'Occipital Lobe and Posterior Cingulate Perfusion in the Prediction of Dementia with Lewy Body Pathology in a Clinical Sample'. *Nuclear Medicine Communications* 38 (12): 1029–35.

Rey, Andŕe. 1958. *L'examen Clinique En Psychologie. [The Clinical Examination in Psychology.]*. L'examen Clinique En Psychologie. Oxford, England: Presses Universitaries De France.

Rosen, Wilma G., Richard C. Mohs, and Kenneth L. Davis. 1984. 'A New Rating Scale for Alzheimer's Disease'. *The American Journal of Psychiatry* 141 (11): 1356–64. https://doi.org/10.1176/ajp.141.11.1356.

Rossum, Ineke A. van, Pieter Jelle Visser, Dirk L. Knol, Wiesje M. van der Flier, Charlotte E. Teunissen, Frederik Barkhof, Marinus A. Blankenstein, and Philip Scheltens. 2012. 'Injury Markers but Not Amyloid Markers Are Associated with Rapid Progression from Mild Cognitive Impairment to Dementia in Alzheimer's Disease'. *Journal of Alzheimer's Disease: JAD* 29 (2): 319–27. https://doi.org/10.3233/JAD-2011-111694.

Saboo, Krishnakant V., Chang Hu, Yogatheesan Varatharajah, Scott A. Przybelski, Robert I. Reid, Christopher G. Schwarz, Jonathan Graff-Radford, et al. 2022. 'Deep Learning Identifies Brain Structures That Predict Cognition and Explain Heterogeneity in Cognitive Aging'. *NeuroImage* 251 (May): 119020. https://doi.org/10.1016/j.neuroimage.2022.119020.

Shaw, Leslie M., Hugo Vanderstichele, Malgorzata Knapik-Czajka, Christopher M. Clark, Paul S. Aisen, Ronald C. Petersen, Kaj Blennow, et al. 2009. 'Cerebrospinal Fluid Biomarker Signature in Alzheimer's Disease Neuroimaging Initiative Subjects'. *Annals of Neurology* 65 (4): 403–13. https://doi.org/10.1002/ana.21610.

Slot, Rosalinde E.R., Sietske A.M. Sikkes, Johannes Berkhof, Henry Brodaty, Rachel Buckley, Enrica Cavedo, Efthimios Dardiotis, et al. 2019. 'Subjective Cognitive Decline and Rates of Incident Alzheimer's Disease and Non–Alzheimer's Disease Dementia'. *Alzheimer's & Dementia* 15 (3): 465–76. https://doi.org/10.1016/j.jalz.2018.10.003.

Sun, Pan, Wutao Lou, Jianghong Liu, Lin Shi, Kuncheng Li, Defeng Wang, Vincent CT Mok, and Peipeng Liang. 2019. 'Mapping the Patterns of Cortical Thickness in Single-and Multiple-Domain Amnestic Mild Cognitive Impairment Patients: A Pilot Study'. *Aging (Albany NY)* 11 (22): 10000.

Tapiola, Tero, Irina Alafuzoff, Sanna-Kaisa Herukka, Laura Parkkinen, Päivi Hartikainen, Hilkka Soininen, and Tuula Pirttilä. 2009. 'Cerebrospinal Fluid β-Amyloid 42 and Tau Proteins as Biomarkers of Alzheimer-Type Pathologic Changes in the Brain'. *Archives of Neurology* 66 (3): 382–89.

Thomas, Kelsey R., Emily C. Edmonds, Joel S. Eppig, Christina G. Wong, Alexandra J. Weigand, Katherine J. Bangen, Amy J. Jak, Lisa Delano-Wood, Douglas R. Galasko, and David P. Salmon. 2019. 'MCI-to-Normal Reversion Using Neuropsychological Criteria in the Alzheimer's Disease Neuroimaging Initiative'. *Alzheimer's & Dementia* 15 (10): 1322–32.

Thomas, Kelsey R., Joel S. Eppig, Alexandra J. Weigand, Emily C. Edmonds, Christina G. Wong, Amy J. Jak, Lisa Delano-Wood, et al. 2019. 'Artificially Low Mild Cognitive Impairment to Normal Reversion Rate in the Alzheimer's Disease Neuroimaging Initiative'. *Alzheimer's & Dementia* 15 (4): 561–69. https://doi.org/10.1016/j.jalz.2018.10.008.

Weiner, Michael W., Dallas P. Veitch, Paul S. Aisen, Laurel A. Beckett, Nigel J. Cairns, Jesse Cedarbaum, Michael C. Donohue, et al. 2015. 'Impact of the Alzheimer's Disease Neuroimaging Initiative, 2004 to 2014'. *Alzheimer's & Dementia* 11 (7): 865–84. https://doi.org/10.1016/j.jalz.2015.04.005.

Whitwell, Jennifer L., Ronald C. Petersen, Selamawit Negash, Stephen D. Weigand, Kejal Kantarci, Robert J. Ivnik, David S. Knopman, Bradley F. Boeve, Glenn E. Smith, and Clifford R. Jack Jr. 2007. 'Patterns of Atrophy Differ Among Specific Subtypes of Mild Cognitive Impairment'. *Archives of Neurology* 64 (8): 1130–38. https://doi.org/10.1001/archneur.64.8.1130.

Winblad, Bengt, Philippe Amouyel, Sandrine Andrieu, Clive Ballard, Carol Brayne, Henry Brodaty, Angel Cedazo-Minguez, et al. 2016. 'Defeating Alzheimer's Disease and Other Dementias: A Priority for European Science and Society'. *The Lancet Neurology* 15 (5): 455–532. https://doi.org/10.1016/S1474-4422(16)00062-4.

Xue, Chen, Haiting Sun, Yingying Yue, Siyu Wang, Wenzhang Qi, Guanjie Hu, Honglin Ge, et al. 2021. 'Structural and Functional Disruption of Salience Network in Distinguishing Subjective Cognitive Decline and Amnestic Mild Cognitive Impairment'. *ACS Chemical Neuroscience* 12 (8): 1384–94. https://doi.org/10.1021/acschemneuro.1c00051.

Yang, Huanqing, Hua Xu, Qingfeng Li, Yan Jin, Weixiong Jiang, Jinghua Wang, Yina Wu, Wei Li, Cece Yang, and Xia Li. 2019. 'Study of Brain Morphology Change in Alzheimer's Disease and Amnestic Mild Cognitive Impairment Compared with Normal Controls'. *General Psychiatry* 32 (2).

Yang, Xianfeng, Ming Zhen Tan, and Anqi Qiu. 2012. 'CSF and Brain Structural Imaging Markers of the Alzheimer's Pathological Cascade'. *PLOS ONE* 7 (12): e47406. https://doi.org/10.1371/journal.pone.0047406.

Yasuno, Fumihiko, Akinori Nakamura, Takashi Kato, Kaori Iwata, Takashi Sakurai, Yutaka Arahata, Yukihiko Washimi, Hideyuki Hattori, and Kengo Ito. 2021. 'An Evaluation of the Amyloid Cascade Model Using in Vivo Positron Emission Tomographic Imaging'. *Psychogeriatrics* 21 (1): 14–23. https://doi.org/10.1111/psyg.12589.

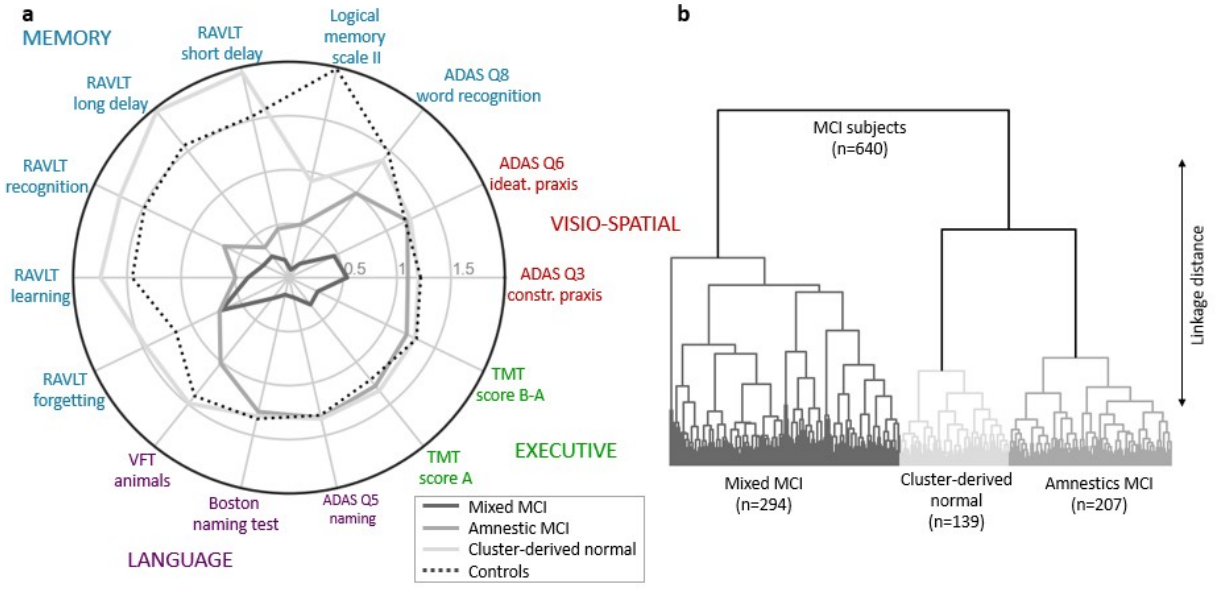

Figure 1: automatic extraction of 3 MCI clusters

Three subgroups were extracted from a cohort of 640 MCI patients.

Polar plot (*a*) shows the Z score for each neuropsychological test included in the clustering procedure. The grey lines represent each extracted cluster (from darker to lighter: *mixed MCI*, *amnestic MCI* and *cluster-derived normal*) while the dotted black line represent the Z score of the controls. A higher score represents a greater performance.

Dendrogram (*b*) shows the best clustering scheme, 3 subgroups according to 30 metrics, extracted from a hierarchical clustering based on a cohort of 640 MCI participants.

In sum, within the MCI diagnosed participants, we could extract three very specific subtypes, the *mixed MCI* subtype scoring low at almost every test, the *amnestic MCI* subtype, scoring low mostly on test assessing memory, and the *cluster-derived normal* subtype, scoring mostly like controls except for the *logical memory scale II*.

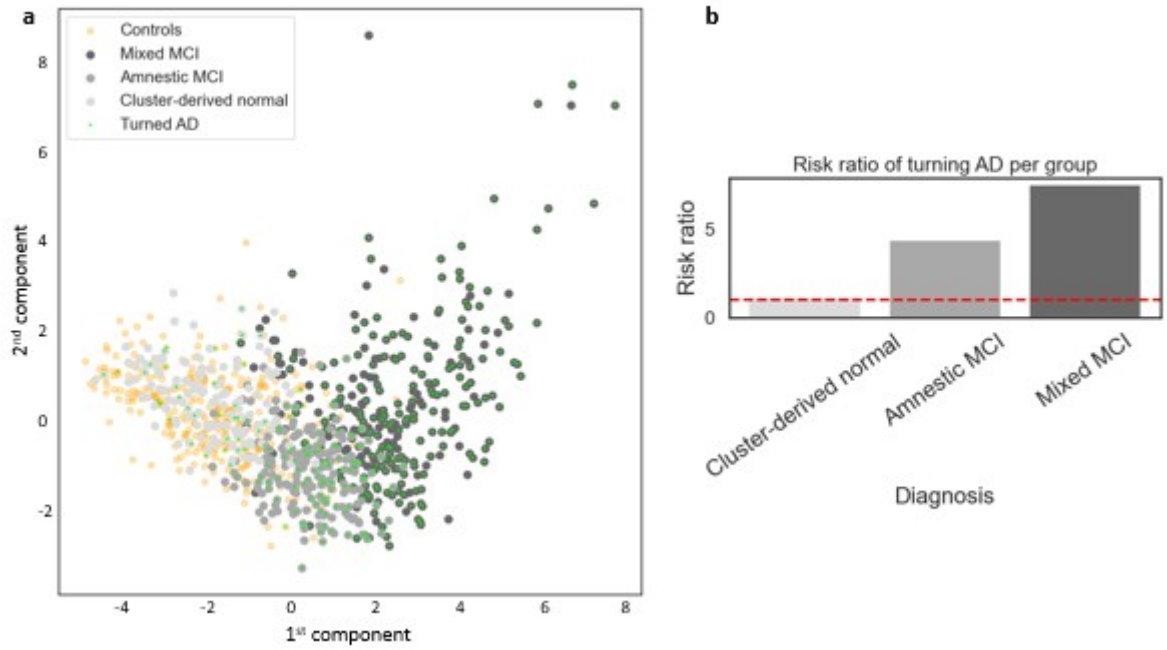

Figure 2: Risk ratio of MCI patients developing Alzheimer

Scatter plot (*a*) displays the participant first and second component of a PCA analysis on the neuropsychological tests included in our clustering analysis. Note that this PCA analysis was computed for the sake of visualization only, however, more details about this analysis can be found in the supplementary.

The grey bar graphs (*b*) display the risk ratio of developing Alzheimer disease in each extracted subgroup compared to controls. The red dotted line represents a risk ratio similar to the control group risk ratio.

These results exhibit a higher propensity to develop Alzheimer for MCI patients in the mixed MCI and amnestic MCI subgroups with twice as much risk for the mixed MCI subgroup.

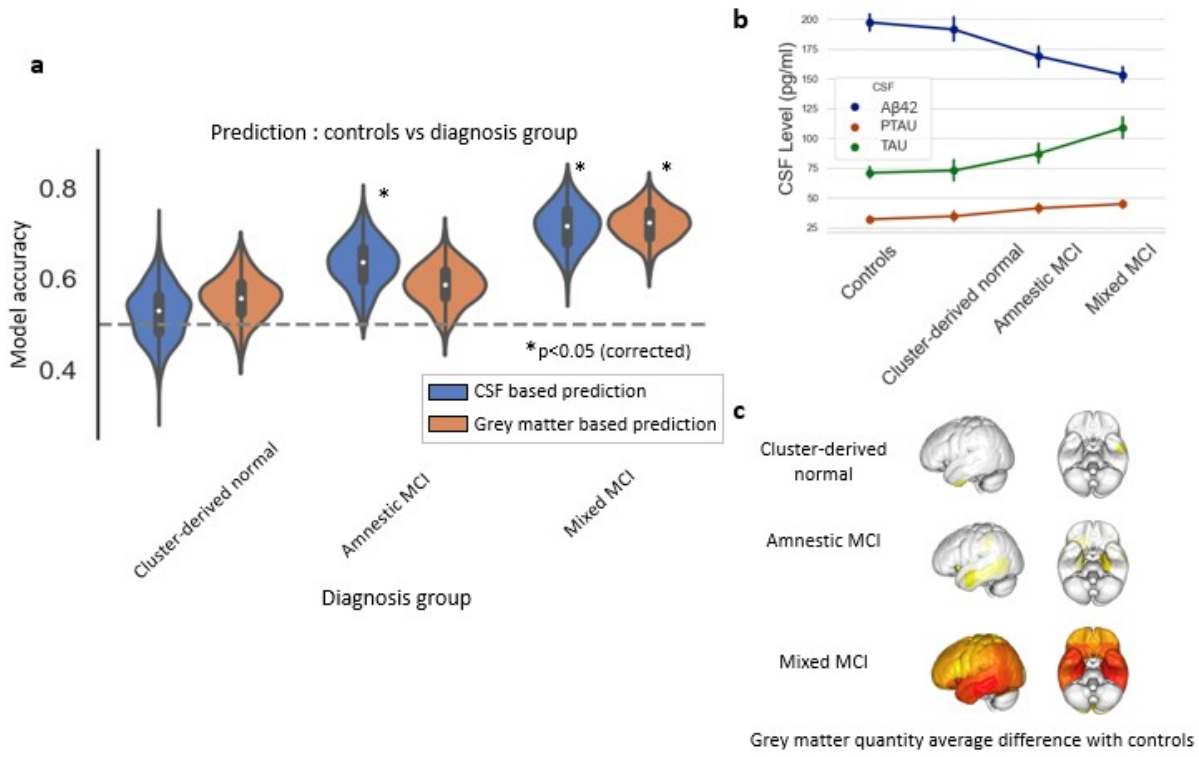

Figure 3: MCI cluster prediction based on grey matter or CSF level

We explored the hypothesis that grey matter volume on the one hand, and level of Amyloid-β 1 to 42 peptide (Aβ42), total tau (tau), and tau phosphorylated (ptau) on the other hand may predict affiliation to MCI subgroups. Note that all subgroups were extracted based on neuropsychological test scores only, thus the CSF biomarkers and grey matter information did not take part into the cluster extraction procedure.

Violin plots (*a*) display the generalization performance (test set) of the prediction using grey matter volume (*blue*) and CSF biomarker level (*orange*) between controls and each MCI subgroup. The width of the violins illustrates the density of the obtained performances. For instance, the shape of the first blue violin plot on the left side (skinny on each end and wide in the middle) indicates that the obtained accuracies are highly concentrated around the median. The height of the violins indicates the variability (i.e., range of the obtained accuracies). Short violins represent a slight while long violins represent a substantial variability. A non-parametric test was applied to assess the (Bonferroni corrected) significance of the accuracy, that is, to evaluate if such an accuracy could be obtained by chance alone. The significant accuracies are represented with a black star.

The line graph (*b*) displays the mean (with the standard deviation) CSF biomarkers level (Aβ42, *tau*, and *ptau*) per MCI subgroups as well as in controls.

The brains (*c*) indicates the grey matter quantity average difference between controls and each MCI subgroups. The redder the area, the higher the atrophy compared to controls.

As a general observation, a better performance was achieved when dissociating mixed MCI from the controls using grey matter volume, as well as when using CSF biomarkers level. However, the models could not segregate cluster-derived normal from controls using these modalities. The amnestic MCI subgroup was distinguishable from the controls based on CSF biomarkers level but not on grey matter volume.

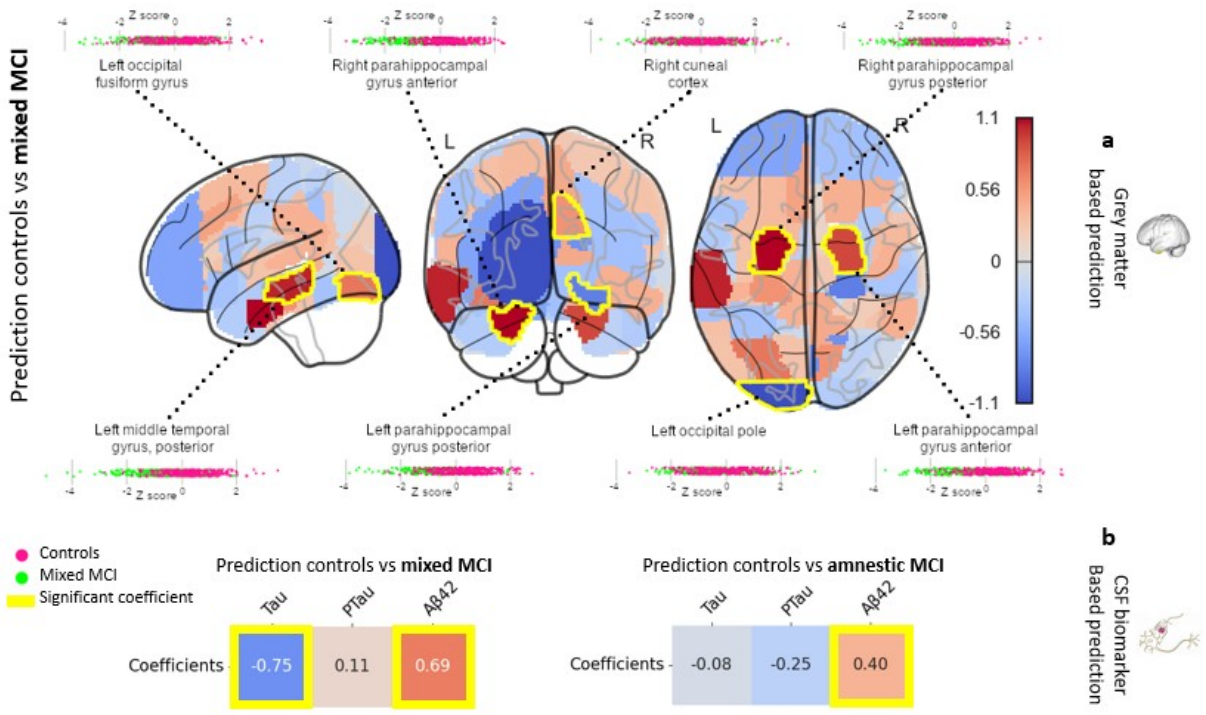

Figure 4: Map of coefficient for the predictions of controls versus each MCI cluster

Prediction of controls versus mixed MCI subgroup membership was assessed using grey matter volume (a) or CSF biomarkers level (b) and using regularized logistic regression models. The colormap on each glass brain (a) depicts the final coefficient value for each region of interest (roi). A non-parametric test was computed to assess significance of the coefficients. That is, to evaluate if a high coefficient was high only by chance or not. Significant rois are outlined in yellow. For each significant roi, boxplots of the distribution of grey matter volume per subjects for controls (*pink*) and mixed MCI (*light green*) are displayed. The heatmap (b) displays the final coefficient value for each CSF biomarker, with significant biomarkers outlined in yellow.

In sum, 8 rois passed the (Bonferroni corrected) threshold and thus had a significant coefficient in predicting mixed MCI versus controls based solely on grey matter volume. These rois are located within the temporal and occipital lobes. Controls had on average more grey matter than mixed MCI subjects in all these significant roi.

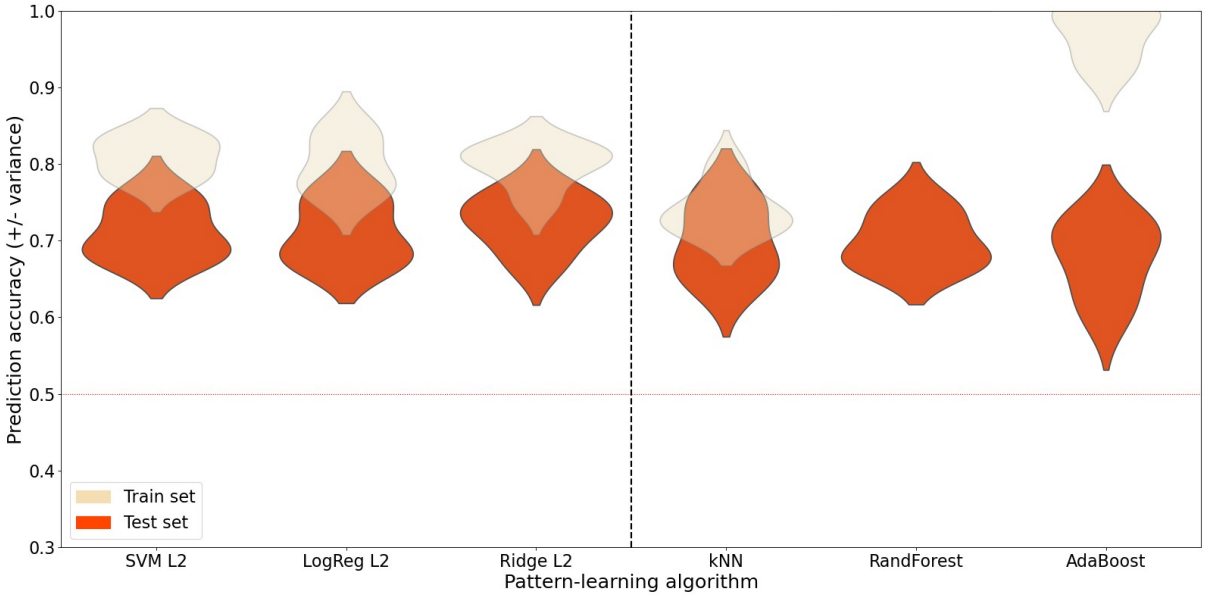

Figure 5: Probing complex relationships among the ROI volumes.

We explored the hypothesis that more complex patterns may explain relationships between different ROI volumes and thus be helpful to distinguish mixed MCI from controls. We thus compared the predicting performance of models looking for additive effects (left side) to the prediction performance of models looking for non-linear effects (right side). The transparent orange violin plots display the in-sample accuracies (train set) while the plain orange plots display the generalization performance (test set). The width of the violins illustrates the density of the obtained performances. For instance, the shape of the first plain orange violin plot on the left side (skinny on each end and wide in the middle) indicates that the obtained accuracies are highly concentrated around the median. The height of the violins indicates the variability (i.e., range of the obtained accuracies). Short violins represent a slight while long violins represent a substantial variability. Linear models including the ridge regression (*Ridge L2*), the logistic regression (*LogReg L2*), and the support vector machine (*SVM L2*) are plotted on the left side of the dashed bar. Non-linear models including the k nearest neighbor (kNN), the random forest (RandForest) and the adaptive boosting (AdaBoost) are plotted on the right side. As a general observation, the plain orange violin plots of the linear models indicate on average a better performance with less variance thus appear to be more adapted to this setting. These results suggest that the ROI volumes are perhaps mostly individually predictive as much as this evidence is supported by dataset

Table 1: information about the included ADNI subjects

Diagnosis	N total	N male	N female	Age mean (SD)	N ADNI 1	N ADNI 2
MCI	640	376	264	73.42(7,66)	374	266
CN	326	162	164	75,15(5.57)	211	115

Table 2: Significant ROI weights for the benchmarked linear models

	LogReg L2	**SVM L2**	**Ridge L2**
Left Middle Temporal Gyrus, posterior division	1,05	0,45	0,30
Right Cuneal Cortex	0,59	0,25	
Left Parahippocampal Gyrus, anterior division	1,11	0,42	0,35
Right Parahippocampal Gyrus, anterior division	0,93	0,39	
Left Parahippocampal Gyrus, posterior division	0,59		
Right Parahippocampal Gyrus, posterior division	-0,84	-0,38	
Left Occipital Fusiform Gyrus	0,73	0,31	
Left Occipital Pole	-1,10	-0,46	-0,29