Alzheimer's & Dementia
Diagnosis, Assessment
& Disease Monitoring

RESEARCH ARTICLE

Quantification of cognitive impairment to characterize heterogeneity of patients at risk of developing Alzheimer's disease dementia

Diana L. Giraldo[1,2] | Jan Sijbers[2] | Eduardo Romero[1] | for the Alzheimer's Disease Neuroimaging Initiative[*]

[1] Computer Imaging and Medical Applications Laboratory - CIM@LAB, Universidad Nacional de Colombia, Bogotá, Colombia

[2] imec-Visionlab, Department of Physics at University of Antwerp, University of Antwerp, Antwerp, Belgium

Correspondence
Eduardo Romero, CIM@LAB, Universidad Nacional de Colombia, Bogotá, Colombia.
E-mail: edromero@unal.edu.co

[*]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

Abstract

Introduction: Neuropsychological test scores are limited and standard outcomes may mask the heterogeneity of cognitive impairment. This article presents the calculation and evaluation of six composite scores that quantify domain-specific impairment.

Methods: Parameters for composite scores calculation were learned by performing confirmatory factor analysis in a sample of participants from the Alzheimer's Disease Neuroimaging Initiative database. The obtained scores were evaluated with a separate sample of mild cognitive impairment (MCI) in two automated tasks: unsupervised partition in different subgroups and prediction of progression to dementia for different time windows.

Results: MCI subgroups with distinctive cognitive profiles and risk of progression emerged from cluster analysis. Composite scores outperform standard neuropsychological tests when automatically predicting progression within time windows up to 5 years.

Conclusions: Domain-specific composite scores are useful to delineate profiles of impairment, stratify the MCI risk, and predict progression to dementia.

KEYWORDS
cluster analysis, cognitive domains, composite scores, mild cognitive impairment, neuropsychological tests

1 | INTRODUCTION

Dementia is a syndrome characterized by deterioration of cognition and behavior to the point that the ability to perform daily activities is impaired. When individuals show some memory or cognitive impairment, but exhibit normal behavior, they are diagnosed with mild cognitive impairment (MCI). Alzheimer's disease (AD) is the most common cause of dementia; therefore, identifying which individuals with MCI are more likely to develop AD dementia is an important research path.[1]

Currently, there is no cure for AD and most clinical trials, designed to pharmaceutically slow down AD progression from MCI to dementia, have so far failed.[2] Aside from the questioned efficacy of the

tested treatments, other possible reasons for failures may come from two sources. First, heterogeneity of recruited participants, including advanced AD and variable MCI manifestations, or participants without any underlying pathology.[2] Second, standard cognitive outcomes, set as endpoints, might be highly variable and not sensitive enough to detect subtle cognitive performance changes.[2,3] This is the case of the widely used Alzheimer's Disease Assessment Scale—Cognitive subscale (ADAS-Cog), that has shown high variability and poor sensitivity, likely by measurement errors, patient heterogeneity, and ceiling effects of its subscores, making some subscores uninformative in patients at early stages.[3–5]

Composite outcomes computed with informative subscores from one or multiple tests have been demonstrated to be more robust and sensible measures to detect cognitive and functional changes in MCI.[3,4] However, single composite scores may mask the heterogeneity of cognitive impairment.

Patients diagnosed with MCI show varying levels of impairment in different cognitive domains beyond memory, including language, visuospatial skills, attention, and executive function.[6–8] This heterogeneity is likely linked to differences in the clinical evolution.[9,10] Therefore, evaluation of domain-specific changes could help to identify individuals at greater risk of progressing to dementia.

Composite scores for measuring specific domain impairment have been proposed for memory[11] and executive function.[12] These scores mitigate the effect of measurement errors for individual items while combining informative subscores from multiple tests. Evaluation of these two previously proposed scores demonstrated they show better performance than individual test scores in detecting domain changes over time and predicting conversion from MCI to dementia.[11,12]

This article presents a data-driven framework combining and weighting subscores from the neuropsychological test battery to calculate a set of domain-specific composite scores that quantify impairment in six domains: memory, language, visuospatial abilities, executive functioning, orientation, and attention. The weighting scheme was obtained by estimating the parameters of a multifactor model with confirmatory factor analysis (CFA). The usefulness of the developed composite scores in MCI was evaluated in two different tasks using machine learning methods. First, the set of composite scores was taken as input for unsupervised cluster analysis, aiming to identify different subgroups of individuals in the MCI sample. Second, we tested the ability of composite scores to predict progression from MCI to dementia within specific time windows, ranging from 1 to 5 years, and compared the performance against standard outcomes.

2 | METHODS

The data-driven methodology presented here is divided into two parts. The first consists of learning the parameters for subscore standardization and domain scores calculation. The second evaluates the composite scores in two automated tasks: clustering of patients diagnosed with MCI and predicting progression to dementia.

RESEARCH IN CONTEXT

1. **Systematic review**: We used PubMed and Google scholar to search in the literature for works investigating cognitive heterogeneity in mild cognitive impairment (MCI) patients. We found these works depend on individual subscore measures that are sensitive to measurement errors and domain-specific composite scores have not been used for this task. In the same direction, we also found that standard outcome measures for clinical trials and patient monitoring mask the cognitive heterogeneity of patients diagnosed with MCI.

2. **Interpretation**: We present a set of six domain-specific composite scores to characterize the cognitive state of MCI patients. Using these measures in cluster analysis, four subgroups of MCI were identified that exhibit different risk of progression to Alzheimer's disease (AD) dementia. The combination of proposed domain scores perform better than standard outcomes in the automated prediction of progression.

3. **Future directions**: Future research includes (1) the study of associations between domain scores and AD-specific biomarkers to improve the understanding of underlying mechanisms causing the cognitive impairment and (2) longitudinal evaluation of scores to delineate the paths of impairment progression.

2.1 | Participant data

Data was provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. ADNI is a public–private partnership with the primary goal of testing whether brain imaging, biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. The dataset herein used comprised 680 patients with MCI and 668 cognitively unimpaired (CU) participants. The demographics and characteristics of these groups are presented in Table 1, corresponding to the first visit with complete information. Subscores from nine different tests were used in the present study, namely: ADAS-Cog,[13] Mini-Mental State Examination (MMSE),[14] Montreal Cognitive Assessment (MoCA),[15] Rey Auditory Verbal Learning Test (AVLT), Logical Memory test immediate and delayed,[16] Clock Drawing test,[17] Category Fluency test,[18] Trail Making A and B,[19] and one of the naming tests depending on its availability: Boston Naming Test[20] or Multilingual Naming Test.[21]

2.2 | Data partition

The ADNI sample was split following the two methodological parts: learning and evaluation. For the learning set, 60% of the CU sample

TABLE 1 Description of sets used in each step of the methodology, including the percentage of carriers of the ε4 allele of the apolipoprotein E (*APOE*) gene, and distributions of total scores for the Mini-Mental State Examination (MMSE), Clinical Dementia Rating–Sum of Boxes (CDR-SoB), and the Alzheimer's Disease Assessment Scale–Cognitive subscale (ADAS-Cog)

	Learning set		Evaluation set	
	CU (n = 668)		MCI (n = 680)	
	Normative data (n = 400)	CFA (n = 268)	Evaluation (n = 272)	CFA (n = 408)
Sex (% female)	54.5	59.3	44.5	40.7
Age (mean ± SD)	73.4 ± 6.9	72.6 ± 8.0	72.6 ± 8.2	72.8 ± 7.8
APOE ε4 (% carriers)	31.0	28.9	43.8	46.7
CDR-SoB (mean ± SD)	0.05 ± 0.17	0.05 ± 0.15	1.51 ± 1.04	1.50 ± 0.98
MMSE (mean ± SD)	29.2 ± 1.1	28.9 ± 1.2	27.9 ± 1.8	28.0 ± 1.7
ADAS-Cog (mean ± SD)	10.0 ± 4.7	11.1 ± 4.5	16.4 ± 6.8	15.0 ± 6.8

Abbreviations: ADAS-Cog, Alzheimer's Disease Assessment Scale–Cognitive subscale; *APOE*, apolipoprotein E; CDR-SoB, Clinical Dementia Rating–Sum of Boxes; CFA, confirmatory factor analysis; CU, cognitively unimpaired; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; SD, standard deviation.

(n = 400) was taken as normative data for subscore standardization, while the remaining 40% (n = 268) and 40% of the MCI sample (n = 272) were used to learn the parameters for calculating the composite scores with CFA. The evaluation set corresponded to the remaining 60% of MCI participants (n = 408), for which composite scores were calculated using the parameters from the learning set.

2.3 | Subscore standardization

Given the heterogeneous scales of neuropsychological tests, some of the scales were inverted to ensure that increasing values correspond to poorer performance. An initial set of 50 subscores (File A in supporting information) were transformed into standardized regression based (SRB) z-scores using the parameters learned from a normative sample. Specifically, each subscore was regressed on age and years of education with the group of 400 CU participants. Then, the regression coefficients and residual standard deviation were used to calculate the SRB z-scores for all participants, including CU and MCI patients.

The subscore from the naming test after a semantic clue (BMCUED) was dropped from further analysis because higher values, after scale inversion, can be associated with poor performance or perfect performance without the cue.

2.4 | Derivation of domain scores

The estimation of composite measures for six different domains was done by proposing and testing a factor model, which links a set of subscores from multiple tests with six domains: memory, language, visuospatial abilities, executive functioning, orientation, and attention (see Figure 1). Before establishing a factor model, variability of subscores and pairwise correlations were examined in the data partition used for CFA. Subscores whose variance was inflated by a few outliers were not included in the model, neither were the subscores showing no signifi-

cant correlation (greater than 0.25) with any other one and were not evaluating a similar task (see File A). The factor model was proposed taking into account what subscores evaluate, but also the number of previous works that performed factor analysis on similar neuropsychological test batteries.[11,12,22,23] The hypothesized model was tested and its parameters were estimated using CFA with the unweighted least squares estimator. Model fit was evaluated by the root mean square error of approximation (RMSEA) and the Tucker-Lewis index (TLI).

With the factor model parameters, the unobserved factor scores are calculated as linear combinations of standardized z-scores.[24] The weights were estimated by minimizing the portion of the observed variance that is not explained by the factors (see File B in supporting information). The resulting estimated factor values quantify dysfunction of the different domains included in the model. The learned set of weights can be used to calculate the domain-specific scores of new observations once they have been transformed into SRB z-scores.

2.5 | Clustering the MCI sample

By exploring the existence of MCI subgroups with an unsupervised clustering method, the six composite scores expose different cognitive profiles in the MCI sample. Specifically, the partition around medoids (PAM) method, also known as k-medoids, iteratively splits the data set in k clusters, being the k representative points the most central points (medoid) in each cluster, and the remaining points assigned to the cluster with the nearest representative point.[25]

Here, we incorporated the relations between domains by weighting the distance between two subjects with the estimated covariance between factors, thereby ensuring that the largest variance dimensions contribute more to the differences between subjects.

The number of clusters was set by revising a collection of 30 indices[26] for multiple options of k from 2 to 10. Cluster stability for these possible partitions (2 ≤ k ≤ 10) was also evaluated following a bootstrap approach. For a given partition in k subgroups, this process
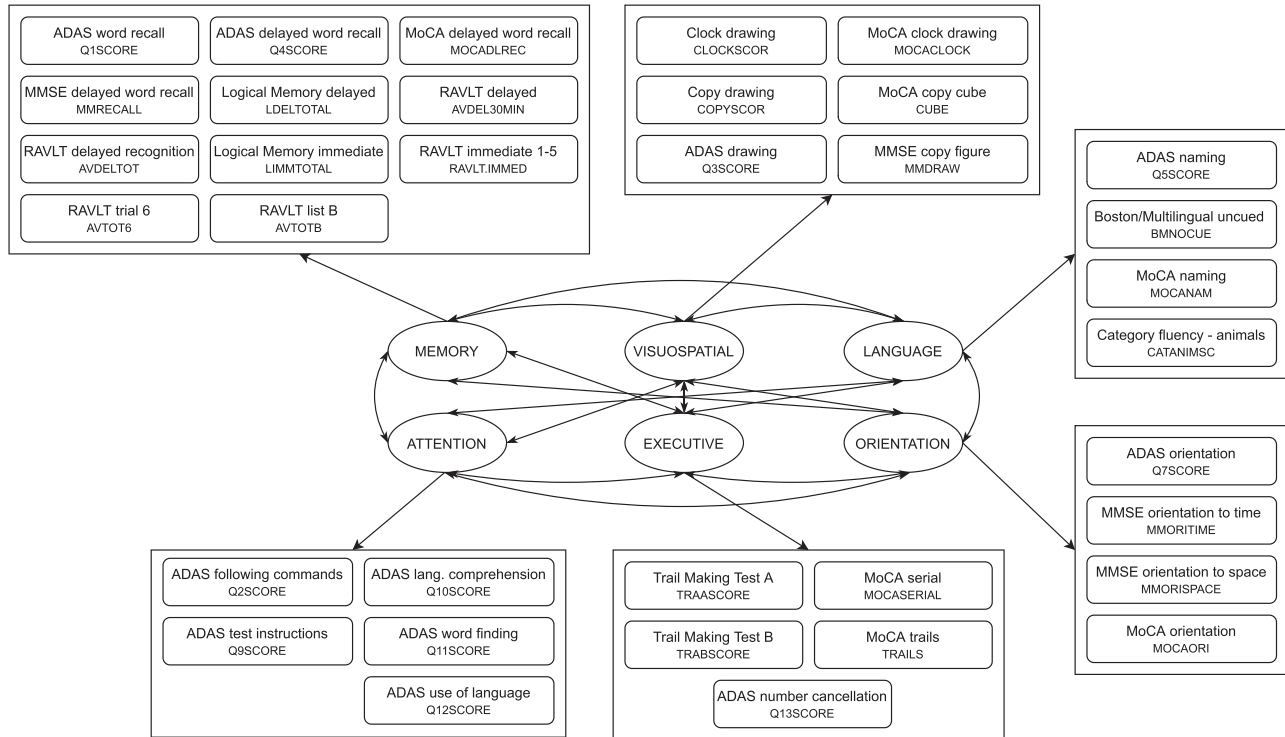
FIGURE 1 Proposed factor model connecting six cognitive domains with 35 subscores from nine different neuropsychological tests. A complete table with each subscore code and description can be found in File A in supporting information. ADAS, Alzheimer's Disease Assessment Scale; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; RAVLT, Rey Auditory Verbal Learning Test

consists of partitioning a subsample (80%) in k subgroups, calculating the overlap with the initial clusters, and repeating this procedure multiple times (1000 iterations). The mean value of the overlap across iterations measures cluster stability.

2.5.1 | Differences between MCI subgroups

Resulting subgroups of MCI participants were compared in terms of their composite scores per domain and their risk of progression to dementia. Pairwise domain score differences between subgroups were examined with Wilcoxon-Mann-Whitney *U* tests while applying the Bonferroni correction for multiple comparisons. A multivariate Cox proportional hazard regression model tested the subgroup effect in the progression from MCI to AD dementia while controlling for age, sex, and years of education. Kaplan-Meier survival curves illustrated progression to dementia of the different MCI subgroups, and curves were compared using omnibus and pairwise log-rank tests.

2.6 | Prediction of progression to dementia

Domain-specific scores were also evaluated in the automated prediction of progression from MCI to AD dementia. This evaluation consisted of classifying MCI patients as either stable or converters following the time window approach[27] fixing five different time periods: 12,

24, 36, 48, and 60 months. The six composite scores along with age, sex, and years of education were used to train random forest classifiers.[28] A random forest (RF) is an ensemble of decision trees constructed using a bootstrap aggregating approach. To create each decision tree, a new training set is generated by sampling, uniformly and with replacement, the original training set. This procedure ensures the collection of trees comes from independent identically distributed samples. The prediction is given by the majority voting of the decision trees in the ensemble, effectively improving the prediction accuracy.[28]

Classification performance was assessed by constructing the receiver operating characteristic (ROC) and calculating its area under the curve (AUC). Depending on the time window, data for training the classifier might be highly unbalanced. This was taken into account when designing the cross-validation scheme: at each iteration, a RF classifier was trained with a balanced subset by randomly selecting the 70% of the underrepresented class with an equal number of samples from the other class. The classifier was tested with the remaining observations, in some cases reaching a larger number of samples. This process was repeated 1000 times per time window.

2.7 | Implementation

All methods were implemented in R (version 3.6.3), code for processing ADNI data and reproducing the reported results is available in https://github.com/diagiraldo/neuropsycho_adni .

TABLE 2 Description of MCI subgroups, along with the CU sample for reference

	CU	MCI subgroup			
	reference	MCI 1	MCI 2	MCI 3	MCI 4
N	668	159	129	88	32
Age (mean ± SD)	73.1 ± 7.4	72.4 ± 7.8	72.2 ± 7.8	74.6 ± 7.5	72.9 ± 7.7
Sex (% female)	56.4	43.4	34.9	43.2	43.8
APOE ε4 (% carriers)	30.2	35.9	47.3	55.2	75
Memory	−0.59 ± 0.70	−0.04 ± 0.63	0.57 ± 0.72	0.95 ± 0.73	1.51 ± 0.67
Language	−0.26 ± 0.56	−0.18 ± 0.49	0.39 ± 0.72	0.50 ± 1.01	0.66 ± 1.00
Executive	−0.14 ± 0.26	−0.09 ± 0.26	0.13 ± 0.47	0.19 ± 0.44	0.31 ± 0.62
Visuospatial	−0.27 ± 0.96	−0.71 ± 0.43	0.83 ± 1.09	0.49 ± 1.40	1.22 ± 1.35
Orientation	−0.51 ± 0.82	−0.70 ± 0.41	−0.59 ± 0.45	1.60 ± 0.98	5.06 ± 1.84
Attention	−0.15 ± 0.35	−0.08 ± 0.38	0.11 ± 0.72	0.28 ± 1.00	0.22 ± 0.89
Mean CSI	—	0.97	0.94	0.9	0.85
Cox proportional HR	—	ref.	2.57 P ≤ .001	3.84 P ≤ .001	7.68 P ≤ .001
95% CI	—	ref.	1.59–4.20	2.33–6.30	4.32–13.70

Notes: Demographic information, mean and SD of domain composite scores (mean ± SD), mean CSI, and proportional HR with their 95% CI.
Abbreviations: APOE, apolipoprotein E; CI, confidence interval; CSI, cluster stability index; CU, cognitively unimpaired; HR, hazard ratio; MCI, mild cognitive impairment; SD, standard deviation.

3 | RESULTS

3.1 | Domain scores

Fit statistics given by CFA indicate a good model fit (RMSEA = 0.09, TLI = 0.95). After the factor model parameters are estimated, domain composite scores are obtained as linear combinations of subscores. Weights for this calculation are presented in File B.

3.2 | Subgroups of MCI patients

The cognitive state of MCI participants was characterized by the six domain scores and different impairment profiles were found in the MCI patient sample by cluster analysis. After the distance between subjects is estimated, there are multiple criteria to choose the number of clusters (k) into which data could be divided. After examining 30 different indices,[26] data partition in four clusters was suggested by 13 of these indices. Additionally, the mean cluster stability index was checked for multiple values of k resulting in values above 0.85 for 2 ≤ k ≤ 4. PAM was applied to divide the MCI sample into four different subgroups. The description of these subgroups is presented in Table 2 along with the description of the group of CU participants as a reference. Figure 2 shows the distributions of domain dysfunction scores for each one of the MCI subgroups. A total of 60 pairwise tests was performed to compare domain composite scores between MCI subgroups and against the CU group, effect size r was computed for each test, and P-values were adjusted for multiple comparisons using the Bonferroni correction. Two profiles were observed at the extremes of the dysfunction spectrum: group 1 exhibits the lowest impairment in all domains, being comparable with the CU group,

and group 4 has the highest dysfunction scores in five out of six domains.

In the control-like subgroup, 49 out of 159 individuals progressed to dementia on the course of the follow-up; those participants converted on average 44.5 months after evaluation. This particular subgroup supports previous findings, which suggest a considerable number of false positives in the diagnosis of MCI in the ADNI database.[29,30]

Characterization of MCI participants with the six proposed domain dysfunction scores revealed four different cognitive profiles in the sample of ADNI participants diagnosed with MCI:

- MCI 1 shows the lowest scores for all six domains. Compared to controls, this group shows significantly higher memory dysfunction (effect size r = 0.31, P < .00005) and lower visuospatial dysfunction score (effect size r = 0.17, P = .00008). Indeed, these participants should have exhibited some memory impairment during the neuropsychological evaluation to be diagnosed with MCI according to the ADNI criteria.
- MCI 2 is more impaired than MCI 1 in memory (effect size r = 0.39, P < .00005), language (effect size r = 0.44, P < .00005), executive function (effect size r = 0.25, P = .0012), and visuospatial abilities (effect size r = 0.75, P < .00005). Although the attention dysfunction does not differ from MCI 1, it does differ with respect to CU (effect size r = 0.18, P = .00001).
- The third subgroup (MCI 3) differs from MCI 2 only in memory (effect size r = 0.25, P = .015) and orientation (effect size r = 0.81, P < .00005).
- The last subgroup (MCI 4) differs from MCI 3 in memory (effect size r = 0.32, P = .026) and orientation (effect size r = 0.73, P < .00005).
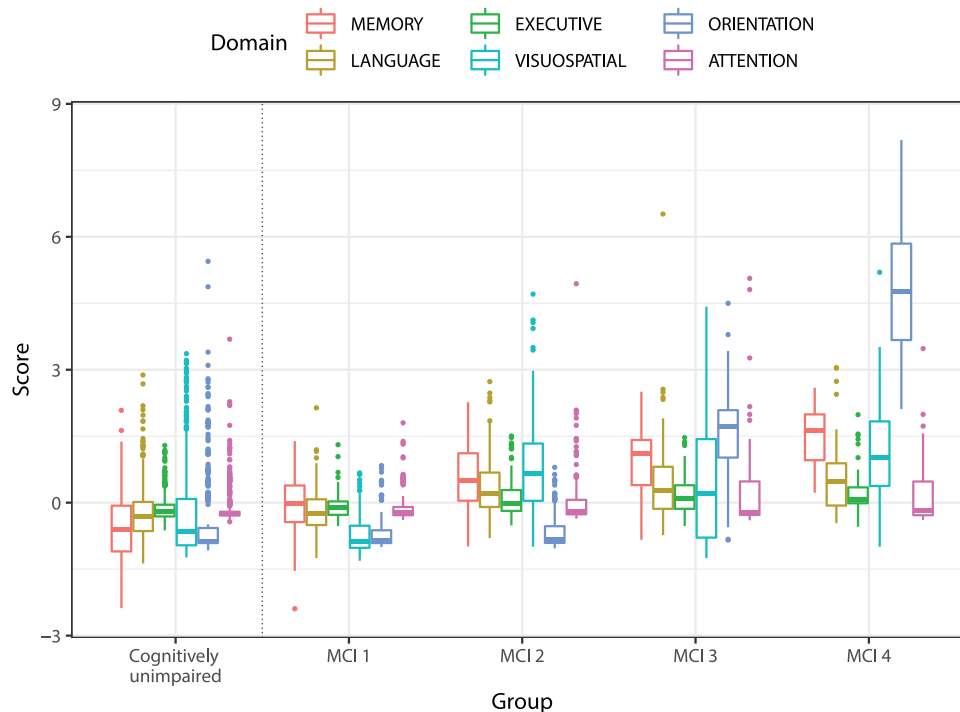
FIGURE 2 Distribution of domain dysfunction scores per mild cognitive impairment (MCI) subgroup including the complete group of cognitively unimpaired (CU) participants as reference

Kaplan-Meier survival curves for the four subgroups of MCI are illustrated in Figure 3. According to the pairwise comparison between curves, MCI subgroup 1 exhibits significantly lower progression probability than the other three subgroups, and subgroup 4 has significantly higher progression probability than the rest of subgroups. The resulting MCI subgroups show distinctive survival curves confirming that the different cognitive profiles are related with different progression risk.

The multivariate Cox models took the control-like subgroup (MCI 1) as reference and included sex, age, and years of education as covariates. The resulting proportional hazard ratios (HR) are presented in Table 2, HR estimates for MCI 2 and 3 compared to MCI 1 are 2.57 (95% confidence interval [CI: 1.59–4.20]), and 3.84 (95% CI [2.33–6.30]), respectively. Significantly higher HR results for MCI subgroup 4, which have a risk of progression to AD dementia around 7.7 (95% CI [4.32–13.70]) times higher than the risk for the control-like subgroup. From the Cox model, age, years of education, and sex had no effect.

3.3 | Automated prediction of progression to AD dementia

Classifiers were trained for each time window with the six domain scores, years of education, age, and sex. To compare with standard outcome measures, at each iteration of the validation scheme, two additional classifiers were trained while including the same covariates using the same sets of observations. The first one was trained with the scores of commonly used neuropsychological tests, namely the ADAS-Cog,

MMSE, MoCA, and AVLT, while the second was trained only with the ADAS-Cog. The number of trees for all RF was set at 200. The distributions of the AUC values per time period across the 1000 iterations for the three classifiers are shown in Figure 4, mean AUC for classification with domain scores are 0.68, 075, 0.74, 0.74, and 0.76 for prediction within 12, 24, 36, 48, and 60 months, respectively.

Classifier performance is significantly higher when trained with domain scores rather than with the set of test totals, including the ADAS-Cog. When predicting MCI conversion within 12 months, resulting mean AUCs are 0.68 and 0.63 (Cohen's $d = 0.73$, $P \leq .00001$) for classifiers trained with dysfunction scores and total tests, respectively. When the conversion prediction is done within 60 months, these mean AUC values are 0.76 and 0.69 (Cohen's $d = 1.59$, $P \leq .00001$), respectively. Although it might be counter-intuitive that prediction performance is better for the long term than for the short term, this is likely due to the varying number of cases used for training and testing at each time window (see File C in supporting information).

4 | DISCUSSION

This work has introduced a data-driven methodology to characterize the cognitive state of MCI patients by developing specific domain scores using subscores from the neuropsychological tests battery applied to the ADNI participants. These domain scores highlight subgroups of MCI patients who exhibit different risks of progression to AD dementia, and show better performance than standard outcomes when predicting conversion from MCI to dementia up to 5 years.
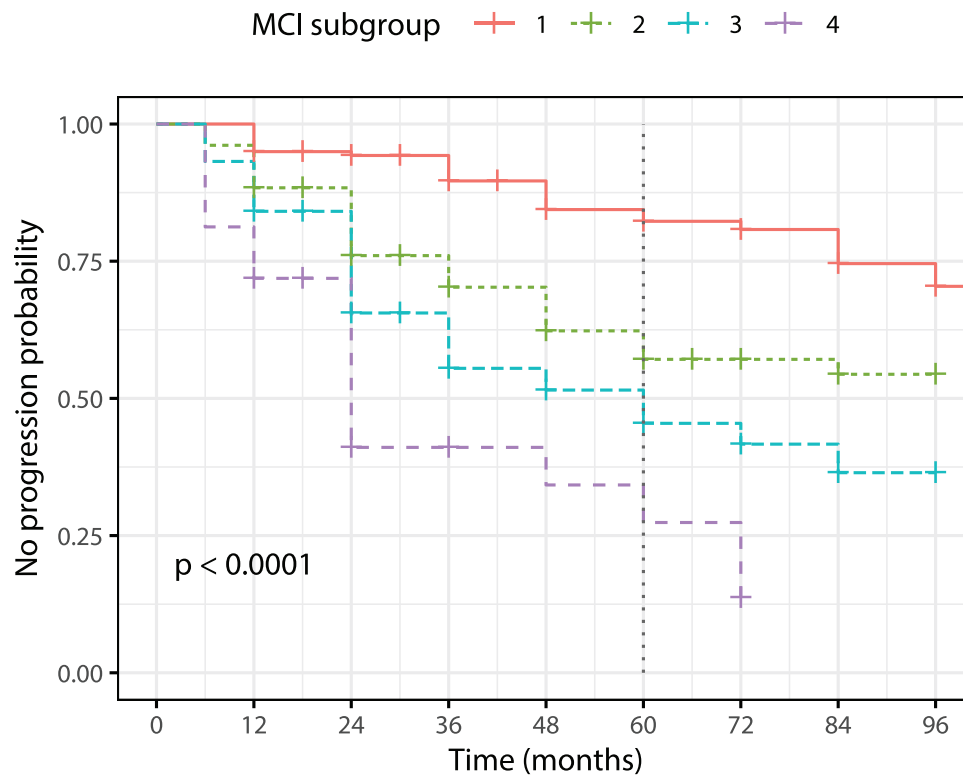
FIGURE 3   Kaplan-Meier curves for the mild cognitive impairment (MCI) subgroups
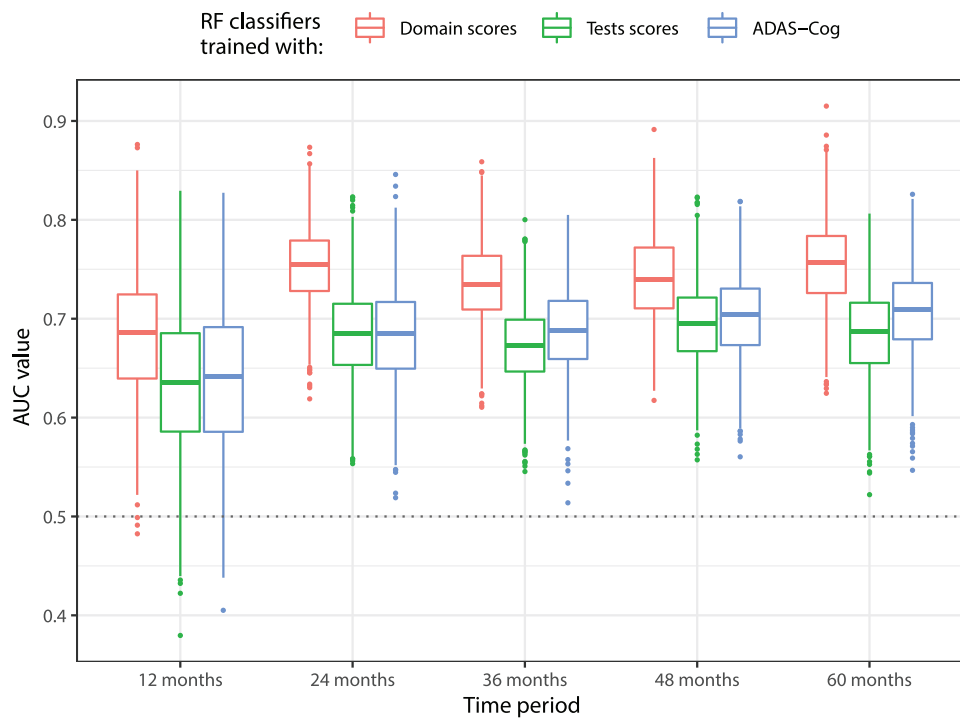

FIGURE 4   Distribution of area under the curve values for prediction of progression from mild cognitive impairment (MCI) to dementia within 12, 24, 36, 48, and 60 months. Classifiers trained with composite domain scores consistently outperform classifiers trained with the Alzheimer's Disease Assessment Scale –Cognitive subscale (ADAS-Cog), and with the set of total tests scores from ADAS-Cog, Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), and Rey Auditory Verbal Learning Test (AVLT)

A six-factor model estimates simultaneously composite scores for all the domains. By learning the weights for domain score calculation from a sample containing both CU and MCI in similar proportions, we can capture a more general statistical structure of the cognitive evaluation than if we had used a narrower sample within the spectrum of impairment. This is an extension of previous works that establish single-factor models to obtain a composite measure for particular domains such as memory[11] and executive functioning.[12] Memory composite score in this work strongly agrees with the one hypothesized for ADNI-Mem,[11] resulting therefore in highly correlated memory measures ($r = -0.943$, $P < .00005$). Executive function score proposed here is also correlated with ADNI-EF[12] ($r = -0.818$, $P < .00005$), even though subscores from ADAS-Cog and MoCA, not considered in ADNI-EF, were herein included.

This investigation evaluates the characterization with two complementary analyses with a sample of MCI patients: unsupervised clustering and automated prediction of future progression to dementia. In the first analysis, MCI heterogeneity is approximated by the six domain composite scores instead of multiple separate scores per domain. As weights for domain score calculation were obtained as a solution that minimizes the portion of the variance that is not explained by the factors,[24] the obtained composite scores do mitigate the effect of individual measurement errors, leading to more robust measures of impairment for each domain. This is a methodological advantage over previous works that studied MCI heterogeneity with separate neuropsychological scores per domain.[29–31] Another methodological advantage consists in adapting the notion of distance between subjects by including the domain covariance in the metrics. Most of the state-of-the-art research performs the cluster analysis[29,31–33] using the Euclidean distance to compare sets of cognitive variables between individuals. However, this distance relies on the assumption of orthogonality between dimensions and therefore each measure is considered independent from the other ones, an assumption hard to hold and far from the given nature of the data.

The cognitive characterization presented partitioned the MCI group into four different subgroups. Beyond the methodological differences, the obtained division is consistent with previous works investigating cognitive heterogeneity in MCI with ADNI data.[29–31,34] All these works also identified a subgroup of control-like individuals in the group of participants diagnosed with amnestic MCI according to ADNI criteria, and two or three MCI subgroups, which vary in the level of impairment of memory, executive function,[31] and language.[29] In this work, the separation between the remaining three MCI subgroups is guided by two domains that covary closely, memory and orientation, while showing relatively similar levels of impairment in language, executive functioning, visuospatial abilities, and attention. Examination of future progression to dementia for the different MCI subgroups in this study resulted in well-differentiated survival curves, providing evidence for the usefulness of the proposed characterization to stratify the risk of progression to dementia during the upcoming 5 years. Therefore, the progressive risk of progression from MCI 1 to MCI 4 seems to be driven by memory and orientation. Although the important role of orientation might be unexpected, it is coherent with previous works that have iden-

tified orientation subscores among the most sensitive measures of cognitive impairment.[4,35] The four MCI subgroups are similar in terms of age and sex distribution, but they exhibit differences in the percentages of *APOE* ε4 carriers (including the genotypes *APOE* ε2/ε4, *APOE* ε3/ε4, and *APOE* ε4/ε4). Although the relation between *APOE* status and risk of AD dementia is widely known, the fact that this known pattern was exposed by orientation impairment might be worthy of further analysis in future work.

The domain scores were also evaluated for automatically predicting future progression from MCI to AD dementia. Cross-validated results demonstrate that classifiers trained with our composite scores consistently outperform classifiers trained with the ADAS-Cog and multiple standard cognitive measures in addition to the ADAS-Cog, such as the MMSE, MoCA, and the AVLT. Prediction with domain scores also outperforms prediction with other cognitive composite scores in the literature.[4,35,36] When the domain scores are accompanied by the Clinical Dementia Rating (CDR) and the Functional Activities Questionnaire (FAQ), prediction performance is slightly better than the prediction with a set of 22 selected neuropsychological features.[37] Information about a fairer comparison with the state-of-the-art can be found in File C.

Considering that psychiatric conditions may play an important role in the development of cognitive impairment, we tested if the addition of psychiatric information improved the performance of progression prediction. Classification results indicate a very modest improvement of AUC values suggesting that psychiatric symptoms give little additional information that could be used to distinguish between MCI patients that will or will not progress to dementia. More details about this experiment and results are in File C.

One important limitation of this study is that only data from ADNI were used, so generalization to other samples of populations was not tested. The main reason for this is that the proposed methodology needs the subscores from neuropsychological tests and information with this level of detail is not available in other public databases.

The presented set of composite scores leads to a quantitative characterization of cognitive state for MCI patients. The presented results demonstrate that these composite domain scores are useful to stratify MCI patients and predict their future progression to dementia. Therefore, those scores could be easily included for patient monitoring or clinical trials. Future work should include longitudinal evaluation of domain dysfunction, along with AD biomarkers, that could improve understanding of the continuum between MCI and AD dementia.

ETHICAL STATEMENT

All procedures followed in ADNI 1/GO/2/3 were conducted according to Good Clinical Practice guidelines, the Declaration of Helsinki, US 21CFR Part 50 – Protection of Human Subjects, and Part 56 – Institutional Review Boards/Research Ethics Board, and pursuant to state and federal HIPAA regulations. Written informed consent was obtained from all participants and/or authorized representatives at each site. More details about the ethical considerations can be found in the clinical protocols at http://adni.loni.usc.edu/methods/documents/ (this article does not contain any studies with human participants performed by any of the authors).

CONFLICTS OF INTEREST

The authors report no conflicts of interest.

REFERENCES

1. Alzheimer's disease facts and figures. *Alzheimers Dement* 2020;16:391-460. https://doi.org/10.1002/alz.12068
2. Mehta D, Jackson R, Paul G, et al. Why do trials for Alzheimer's disease drugs keep failing? A discontinued drug perspective for 2010-2015. *Expert Opin Investig Drugs*. 2017;26:735-739. https://doi.org/10.1080/13543784.2017.1323868.
3. Sabbagh MN, Hendrix S, Harrison JE. FDA position statement "Early Alzheimer's disease: developing drugs for treatment, Guidance for Industry". *Alzheimers Dement*. 2019;5:13-19. https://doi.org/10.1016/j.trci.2018.11.004.
4. Raghavan N, Samtani MN, Farnum M, et al. The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimers Dement*. 2013;9:S21-S31. https://doi.org/10.1016/j.jalz.2012.05.2187.
5. Grochowalski JH, Liu Y, Siedlecki KL. Examining the reliability of ADAS-Cog change scores. *Aging, Neuropsychology, and Cognition*. 2016;23:513-529. https://doi.org/10.1080/13825585.2015.1127320.
6. Loewenstein DA, Ownby RMD, Schram L, et al. An evaluation of the NINCDS-ADRDA neuropsychological criteria for the assessment of Alzheimers disease: a confirmatory factor analysis of single versus multi-factor models. *J Clin Exp Neuropsychol*. 2001;23:274-284. https://doi.org/10.1076/jcen.23.3.274.1191. Neuropsychology, Development and Cognition: Section A.
7. Weintraub S, Wicklund AH, Salmon DP. The neuropsychological profile of Alzheimer disease. *Cold Spring Harb Perspect Med*. 2012;2. https://doi.org/10.1101/cshperspect.a006171.
8. Garcia-Alvarez L, Gomar JJ, Sousa A, et al. Breadth and depth of working memory and executive function compromises in mild cognitive impairment and their relationships to frontal lobe morphometry and functional competence. *Assessment & Disease Monitoring*. 2019;11:170-179. https://doi.org/10.1016/j.dadm.2018.12.010. Alzheimer's & Dementia: Diagnosis.
9. Tatsuoka C, Tseng H, Jaeger J, et al. Modeling the heterogeneity in risk of progression to Alzheimer's disease across cognitive profiles in mild cognitive impairment. *Alzheimer's Research & Therapy*. 2013;5:14. https://doi.org/10.1186/alzrt168.
10. Goyal D, Tjandra D, Migrino RQ, et al. Characterizing heterogeneity in the progression of Alzheimer's disease using longitudinal clinical and neuroimaging biomarkers. *Assessment & Disease Monitoring*. 2018;10:629-637. https://doi.org/10.1016/j.dadm.2018.06.007. Alzheimer's & Dementia: Diagnosis.
11. Crane PK, Carle A, Gibbons LE, et al. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav*. 2012;6:502-516. https://doi.org/10.1007/s11682-012-9186-z.
12. Gibbons LE, Carle AC, Mackin RS, et al. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging Behav*. 2012;6:517-527. https://doi.org/10.1007/s11682-012-9176-1.
13. Rosen W, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry*. 1984;141:1356-1364. https://doi.org/10.1176/ajp.141.11.1356.
14. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12. https://doi.org/10.1016/0022-3956(75)90026-6.
15. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a Brief Screening Tool For Mild Cognitive Impairment. *J Am Geriatr Soc*. 2005;53:695-699. https://doi.org/10.1111/j.1532-5415.2005.53221.x.
16. Wechsler D, *Wechsler Memory Scale-Revised: Manual*. Harcourt Brace Jovanovich; 1987.
17. Goodglass H, Kaplan E. *The Assessment of Aphasia and Related Disorders*. Lea & Febiger. 1983.
18. Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assesment of Alzheimer's disease. *Neurology*. 1989;39. https://doi.org/10.1212/WNL.39.9.1159.
19. Reitan RM, Wolfson D. *The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation*. Neuropsychology Press; 1985.
20. Kaplan E, Goodglass H, Weintraub S. *The Boston Naming Test. Lea and Febiger*. 1989.
21. Gollan TH, Weissberger GH, Runnqvist E, et al. Self-ratings of spoken language dominance: a Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition*. 2011;15:594-615. https://doi.org/10.1017/s1366728911000332.
22. Park LQ, Gross AL, McLaren DG, et al. Confirmatory factor analysis of the ADNI neuropsychological battery. *Brain Imaging Behav*. 2012;6:528-539. https://doi.org/10.1007/s11682-012-9190-3.
23. Giraldo DL, Sijbers J, Romero E. Quantifying cognition and behavior in normal aging, mild cognitive impairment, and Alzheimer's dis-

ease. 13th International Conference on Medical Information Processing and Analysis. 10572, 2017, 125-134. https://doi.org/10.1117/12.2287036

24. Bartlett MS. The statistical conception of mental factors. *British Journal of Psychology General Section.* 1937;28:97-104. https://doi.org/10.1111/j.2044-8295.1937.tb00863.x.

25. Leonard K. *PJR Finding Groups in Data: An Introduction to Cluster Analysis.* 1st ed.. Wiley-Interscience; 2005.

26. Charrad M, Ghazzali N, Boiteau V, et al. NbClust: an R package for determining the relevant number of clusters in a data set. *J Stat Softw.* 2014;61:1-36. https://doi.org/10.18637/jss.v061.i06. Articles.

27. Pereira T, Lemos L, Cardoso S, et al. Predicting progression of mild cognitive impairment to dementia using neuropsychological data: a supervised learning approach using time windows. *BMC Med Inf Decis Making.* 2017;17:110. https://doi.org/10.1186/s12911-017-0497-2.

28. Breiman L. Random forests. *Machine Learning.* 2001;45:5-32. https://doi.org/10.1023/A:1010933404324.

29. Edmonds EC, Delano-Wood L, Clark LR, et al. Susceptibility of the conventional criteria for mild cognitive impairment to false-positive diagnostic errors. *Alzheimers Dement.* 2015;11:415-424. https://doi.org/10.1016/j.jalz.2014.03.005.

30. Eppig JS, Edmonds EC, Campbell L, et al. Statistically derived subtypes and associations with cerebrospinal fluid and genetic biomarkers in mild cognitive impairment: a latent profile analysis. *J Int Neuropsychol Soc.* 2017;23:564-576. https://doi.org/10.1017/S135561771700039X.

31. Bondi MW, Edmonds EC, Jak AJ, et al. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *J Alzheimers Dis.* 2014;42:275-289. https://doi.org/10.3233/JAD-140276.

32. Peter J, Abdulkadir A, Kaller C, et al. Subgroups of Alzheimer's disease: stability of empirical clusters over time. *J Alzheimers Dis.* 2014;42:651-661.

33. Edmonds EC, McDonald CR, Marshall A, et al. Early versus late MCI: improved MCI staging using a neuropsychological approach. *Alzheimers Dement.* 2019;15:699-708. https://doi.org/10.1016/j.jalz.2018.12.009.

34. Nettiksimmons J, DeCarli C, Landau S, et al. Biological heterogeneity in ADNI amnestic mild cognitive impairment. *Alzheimers Dement.* 2014;10:511-521. https://doi.org/10.1016/j.jalz.2013.09.003.

35. Huang Y, Ito K, Billing CB, et al. Development of a straightforward and sensitive scale for MCI and early AD clinical trials. *Alzheimers Dement.* 2015;11:404-414. https://doi.org/10.1016/j.jalz.2014.03.008.

36. Llano DA, Laforet G, Devanarayan V. Derivation of a new ADAS-cog composite using tree-based multivariate analysis. *Alzheimer Dis Assoc Disord.* 2011;25:73-84. https://doi.org/10.1097/wad.0b013e3181f5b8d8.

37. Pereira T, Ferreira FL, Cardoso S, et al. Neuropsychological predictors of conversion from mild cognitive impairment to Alzheimer's disease: a feature selection ensemble combining stability and predictability. *BMC Med Inf Decis Making.* 2018;18:137. https://doi.org/10.1186/s12911-018-0710-y.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.