Spatial Navigation as a Digital Marker for Clinically Differentiating Cognitive Impairment Severity

Giorgio Colombo[1,2,*], Karolina Minta[1,3,*], William R. Taylor[1,4], Jascha Grübel[5,6], Eddie Chong[3,7], Joyce R. Chong[3,7], Mitchell K. P. Lai[3,7], Christopher P. Chen[3,7], Victor R. Schinazi[1,8]

[1]Future Health Technologies, Singapore-ETH Centre, Campus for Research Excellence And Technological Enterprise (CREATE), Singapore
[2]Chair of Cognitive Science, ETH Zurich, Zürich, 8092, Switzerland
[3]Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[4]Institute for Biomechanics, Department of Health Sciences and Technology, ETH Zurich, Zürich, Switzerland
[5]Center for Sustainable Future Mobility, ETH Zurich, Zürich, 8092, Switzerland
[6]Laboratory of Geo-information Science and Remote Sensing, Wageningen University, Wageningen, 6708 PB, The Netherlands
[7]Memory, Aging and Cognition Centre, National University Health Systems, Singapore.
[8]Department of Psychology, Bond University, Gold Coast, Queensland, Australia

* these authors contributed equally

Corresponding author:
Giorgio Colombo
giorgio.colombo@sec.ethz.ch

Abstract

Navigation impairments emerge early in Alzheimer's disease, but assessments targeting these deficits remain underutilised or impractical for cognitive screening. The Spatial Performance Assessment for Cognitive Evaluation (SPACE) is a digital tool that evaluates spatial navigation deficits associated with cognitive impairment. In a memory clinic and community cohort (n = 300), SPACE accurately discriminated across Clinical Dementia Rating levels with higher accuracy than models including demographic variables. Including SPACE significantly increased the AUC for distinguishing normal cognition from mild impairment (0.76 to 0.94), normal cognition from moderate impairment (0.79 to 0.95), and very-mild from mild (0.70 to 0.91), all with high sensitivity and specificity. A short version of SPACE (< 11 mins) reduced administration time by 40% while maintaining high diagnostic accuracy. All models were confirmed through cross-validation. These findings highlight the potential of digital navigation assessments to advance early detection, contributing to scalable and accessible healthcare.

Introduction

Dementia affects approximately 55 million people worldwide, and this number is projected to reach 150 million by 2050[1,2]. The global economic burden of dementia is estimated at $16.9 trillion, imposing substantial financial and social strain on individuals, families, and healthcare systems[3]. Alzheimer's Disease (AD) is the leading cause of dementia and is characterised by progressive cognitive decline. AD typically unfolds across distinct stages, from Mild Cognitive Impairment (MCI), where subtle memory and navigation difficulties emerge, to moderate and severe stages characterised by profound impairments in multiple cognitive domains, including memory, spatial orientation, and daily functioning[2,4-8]. Identifying individuals with MCI provides an opportunity to delay disease progression through targeted interventions[7-10].

Episodic memory impairments have long been the hallmark of MCI and AD, forming the cornerstone of diagnostic criteria. Current neuropsychological assessments typically focus on the detection of memory deficits alongside impairments in other cognitive domains (e.g., attention, executive function). These assessments, however, are often only capable of detecting AD after significant neurodegeneration has already occurred. As a result, most diagnoses are made when the disease has progressed to a more severe stage in which interventions are less effective[11,12]. While blood biomarkers simplify early detection compared to Positron Emission Tomography (PET) and Cerebrospinal Fluid (CSF) methods, their reliance on clinic-based sampling and interpretation does not fully overcome the limitations of in-clinic, late-stage diagnosis.

The early stages of AD are characterised by neurodegenerative changes in the hippocampus and the entorhinal cortex, including neuronal loss as well as the accumulation of amyloid-beta (Aβ) peptides and phosphorylated tau (p-tau) protein[13-17]. Critically, these subcortical regions also play an essential role in coding spatial information in the environment[18] and in assisting individuals to keep track of changes in position and orientation during navigation[19]. Indeed, an accumulating body of evidence from real-world[20-22] and Virtual Reality (VR)[23-28] studies has identified spatial navigation deficits as a promising marker of genetic risk for sporadic AD (APOE ε4-carriers)[29-31] and a reliable means of distinguishing individuals across the spectrum of cognitive impairment[11,32]. Here, experimental paradigms that require individuals to learn and recall the configuration of landmarks in an environment (i.e. cognitive mapping)[20,27,33], maintain orientation and track self-motion (i.e. path integration)[23,24,28,34], as well as infer landmark positions from different viewpoints (i.e. perspective taking)[35,36] have all exhibited sensitivity to early cognitive impairment. Among these tasks, path integration deficits have been consistently associated with hippocampal[20,27,33] and entorhinal cortex[24,30] atrophy, along with reduced activity of these brain areas[29,30].

Advances in digital technologies have enabled the development of scalable, gamified, spatial navigation assessments that offer engaging[37], ecologically valid[38], and accessible tools for early cognitive screening[39-41]. Such digital platforms offer remote deployment, reducing reliance on specialised clinical settings and costly visits, while enabling continuous monitoring of disease progression[7,8,42,43]. Despite strong theoretical and empirical support for spatial navigation as an early marker of cognitive decline, most existing tools have yet to be validated against a gold standard diagnosis. As a result, few, if any, digital navigation tools have made the transition into clinical practice, representing a missed opportunity for early detection and intervention.

In this study, we investigate whether the Spatial Performance Assessment for Cognitive Evaluation (SPACE)[44], deployed on a tablet-based platform (i.e. iPad), can discriminate between patients with various levels of cognitive impairment, as independently assessed by clinicians using the Clinical Dementia Rating (CDR) scale. SPACE assesses spatial navigation abilities through five tasks, including path integration, egocentric pointing, mapping, associative memory, and perspective taking. The different tasks in SPACE require participants to navigate a virtual environment, keep track of their position and orientation, reconstruct spatial layouts, recall object-location associations, and estimate spatial relationships from various perspectives. We assessed 300 individuals referred from memory clinics to examine the diagnostic ability of SPACE in discriminating their levels of cognitive impairment as measured by the CDR.

Results

Group differences in spatial navigation performance

Table 1 | Sample characteristics. Demographic breakdown of the sample by age, gender, education level, and CDR status.

	Gender	N	Mean	SD	Minimum	Maximum
Age	Male	124	74	7.02	53	89
	Female	176	75	6.13	61	89

	Level	N	Percentage
Education	No formal education	13	4%
	Elementary school	39	13%
	Middle school	100	33%
	High school	68	23%
	University	80	27%
CDR	Normal (CDR=0)	153	51%
	Very-mild (CDR=0.5)	96	32%
	Mild (CDR=1)	34	11%
	Moderate (CDR=2)	14	5%
	Severe (CDR=3)	3	1%

We performed robust ANOVAs to assess differences in spatial navigation performance in a cohort of 300 participants (Table 1) across four CDR groups (0 = normal cognition; 0.5 = very-mild impairment; 1 = mild impairment; 2 = moderate impairment, 3 = severe impairment) for each of the tasks in SPACE (Table 2). Results revealed significant CDR group differences in training time ($F_{(3, 296)} = 14.7$, $p < 0.001$, $\xi = 0.657$), path integration distance error ($F_{(3, 296)} = 7.0$, $p = 0.004$, $\xi = 0.587$), and perspective taking error ($F_{(3, 296)} = 11.0$, $p < 0.001$, $\xi = 0.51$). No significant group differences were observed for egocentric pointing error ($F_{(3, 296)} = 2.61$, $p = 0.101$) or mapping accuracy ($F_{(3, 296)} = 1.33$, $p = 0.291$). Due to the small number of participants with a CDR score of 3 (n = 3), the moderate and severe impairment groups were combined for analysis. Excluding these three participants with CDR = 3 and rerunning the analysis yielded comparable results, indicating that their inclusion did not influence the result of the ANOVA (Supplementary Table 1).

Post-hoc comparisons revealed that participants with normal cognition completed the training phase significantly faster than those with very-mild ($p = 0.022$), mild ($p < 0.001$), and moderate impairment ($p < 0.001$). The very-mild group was also significantly faster than both the mild ($p = 0.001$) and moderate groups ($p < 0.001$). Finally, participants with mild impairment were faster than those with moderate impairment ($p = 0.022$). For path integration distance error, participants with normal cognition performed significantly better than those with mild ($p <$

0.001), but not better than those with very-mild impairment (p = 0.44). Participants with very-mild also outperformed those with mild impairment (p < 0.001). For perspective taking error, participants with normal cognition performed significantly better than those with mild (p < 0.001) and moderate impairments (p < 0.001) but did not differ from those with very-mild impairment (p = 0.572). Participants with very-mild impairment also outperformed those with mild (p < 0.001) and moderate impairments (p < 0.001). No differences emerged between participants with mild and moderate impairments (p = 0.476).

Table 2 | Group differences in SPACE task performance by CDR scale levels. ANOVA results and post-hoc comparisons for SPACE performance metrics across clinical groups, with effect sizes and bootstrap confidence intervals.

ANOVA						
Dependent Variable			**F**	***p***	**Variance Explained**	**ES (ξ)**
Training time			14.7	<.001	0.432	0.657
Path integration error			7	0.004	0.345	0.587
Pointing error			2.61	0.101	0.145	0.381
Map R^2			1.33	0.291	0.065	0.256
Perspective taking error			11	<.001	0.26	0.51
Post-hoc comparisons						
	Group 1	**Group 2**	**psi-hat**	***p***	**Lower CI**	**Upper CI**
Training time	Normal	Very-mild	-14.6	0.022	-33.5	2.52
	Normal	Mild	-50.4	<.001	-85.3	-23.28
	Normal	Moderate	-102.8	<.001	-152.7	-50.17
	Very-mild	Mild	-35.8	0.001	-71.5	-5.17
	Very-mild	Moderate	-88.3	<.001	-140.4	-35.55
	Mild	Moderate	-52.5	0.022	-109.3	7.76
Path integration error	Normal	Very-mild	-7.57	0.44	-31.5	15.9
	Normal	Mild	-89.35	<.001	-156.5	-37.4
	Very-mild	Mild	-81.78	<.001	-146.5	-27.8
Pointing error	Normal	Very-mild	-1.67	0.55	-7.87	4.77
	Normal	Mild	-13.23	0.03	-24.82	1.51
	Very-mild	Mild	-11.56	0.058	-24.33	3.54
Map R^2	Normal	Very-mild	0.013	0.826	-0.128	0.150
	Normal	Mild	-0.134	0.168	-0.304	0.101
	Very-mild	Mild	-0.147	0.151	-0.333	0.098
Perspective taking error	Normal	Very-mild	-1.67	0.572	-9.76	5.91
	Normal	Mild	-14.9	<.001	-24.5	-6.24
	Normal	Moderate	-18.28	<.001	-32.51	-7.43
	Very-mild	Mild	-13.22	<.001	-23.06	-4.55
	Very-mild	Moderate	-16.61	<.001	-31.07	-5.64
	Mild	Moderate	-3.38	0.476	-17.42	8.85

Note: Bootstrap method, modified one-step estimator, 5000 samples, projection distances

Diagnostic classification performance

We conducted logistic regression analyses to evaluate the diagnostic utility of the SPACE measures that showed significant effects in the previous ANOVAs (i.e. training, path integration, and perspective taking). These analyses aimed to classify participants across CDR stages by comparing between diagnostic groups. Here, we compared baseline models that included demographic variables (i.e. age, education, and gender) to models that additionally incorporated the significant SPACE metrics. We assessed improvements in model fit using likelihood-ratio tests (see Table 3 and Supplementary Table 2 for full model statistics) and evaluated classification accuracy using the Area Under the Curve (AUC), sensitivity, and specificity.

Table 3 | Model fit and classification performance. Comparison of model fit and predictive accuracy for distinguishing clinical groups, with and without the inclusion of SPACE measures.

						Overall Model Test			Predictive measures						
Model	**Dev**	**AIC**	**Brier**	**MCC**	**R^2N**	**χ^2**	**df**	***p***	**AUC**	**Upp CI.**	**Low CI**	**Opt Cutoff**	**Sens**	**Spec**	**Acc**
Normal *vs.* Very-mild															
Age, Edu, Gender	300	314	0.23	0.14	0.02	3.53	6	0.739	0.57	0.50	0.64	0.39	0.40	0.75	0.62
+ SPACE	295	315	0.22	0.27	0.05	8.41	9	0.493	0.61	0.53	0.69	0.40	0.47	0.79	0.67
Normal *vs.* Mild															
Age, Edu, Gender	103.4	117.4	0.13	0.37	0.17	15.4	6	0.017	0.76	0.68	0.84	0.19	0.74	0.71	0.72
+ SPACE	68.7	88.7	0.07	0.62	0.51	50.1	9	<.001	0.94	0.91	0.98	0.11	1.00	0.85	0.86
Normal *vs.* Moderate															
Age, Edu, Gender	86.6	100.6	0.08	0.36	0.23	19.3	6	0.004	0.79	0.71	0.88	0.11	0.94	0.65	0.68
+ SPACE	51.2	69.2	0.05	0.59	0.59	54.7	8	<.001	0.95	0.91	0.99	0.09	0.94	0.88	0.88
Very-mild *vs.* Mild															
Age, Edu, Gender	89.8	103.8	0.18	0.31	0.13	8.25	6	0.220	0.70	0.60	0.79	0.20	0.74	0.59	0.62
+ SPACE	56.9	76.9	0.09	0.60	0.54	41.14	9	<.001	0.91	0.86	0.97	0.11	0.95	0.73	0.77
Very-mild *vs.* Moderate															
Age, Edu, Gender	74.3	88.3	0.11	0.41	0.26	17.2	6	0.008	0.80	0.69	0.89	0.16	0.88	0.68	0.71
+ SPACE	52.8	70.8	0.07	0.59	0.52	38.7	8	<.001	0.91	0.85	0.98	0.18	0.88	0.85	0.86
Mild *vs.* Moderate															
Age, Edu, Gender	49.1	63.1	0.18	0.42	0.31	12.0	6	0.061	0.76	0.60	0.89	0.30	0.82	0.62	0.69
+ SPACE	44	62	0.15	0.62	0.42	17.11	8	0.029	0.83	0.69	0.97	0.45	0.69	0.91	0.83

Note: Model = predictors included (demographics only vs. demographics + SPACE); Dev = deviance; AIC = Akaike Information Criterion; Brier = Brier score; MCC = Matthews Correlation Coefficient; R^2N = Nagelkerke's R^2; χ^2 = model chi-square; df = degrees of freedom; *p* = significance; AUC = area under the ROC curve; Upp CI / Low CI = 95% confidence interval; Opt Cutoff = optimal threshold (Youden index); Sens = sensitivity; Spec = specificity; Acc = accuracy.

Incorporating SPACE measures improved model performance for all diagnostic contrasts except for normal *vs.* very-mild (Figure 1). Importantly, these measures improved classification not only for moderate impairment but also for early stages, including very-mild to mild cognitive impairment (Tables 3 & 4). Comparison between normal cognition and very-mild impairment produced the smallest improvements. Here, the demographic model yielded an AUC of 0.57, with sensitivity of 0.40 and specificity of 0.75. Adding SPACE raised the AUC slightly to 0.61(p_{DeLong} = 0.256), sensitivity to 0.47 and specificity to 0.79, but this change was not statistically significant ($\Delta\chi^2$ = 4.88, p = 0.181). Greater improvements were achieved for all the other five models. As expected, the addition of SPACE to contrasts involving the moderate impairment group resulted in considerable gains in diagnostic performance. More importantly, however, for the comparison between normal cognition and mild impairment, the demographic model achieved an AUC of 0.76, with sensitivity of 0.74 and specificity of 0.71. Critically, adding SPACE significantly improved model fit ($\Delta\chi^2$ = 34.7, p < 0.001), increasing the AUC to 0.94 (p_{DeLong} < .001), sensitivity to 1.00, and specificity to 0.85. Of similar clinical importance was the differentiation between very-mild to mild impairment, where the demographic model produced an AUC of 0.70, with sensitivity of 0.74 and specificity of 0.59. Again, the addition of SPACE significantly improved the

model ($\Delta\chi^2$ = 32.9, p < 0.001), increasing the AUC to 0.91 (p_{DeLong} <.001), sensitivity to 0.95 and specificity to 0.73.

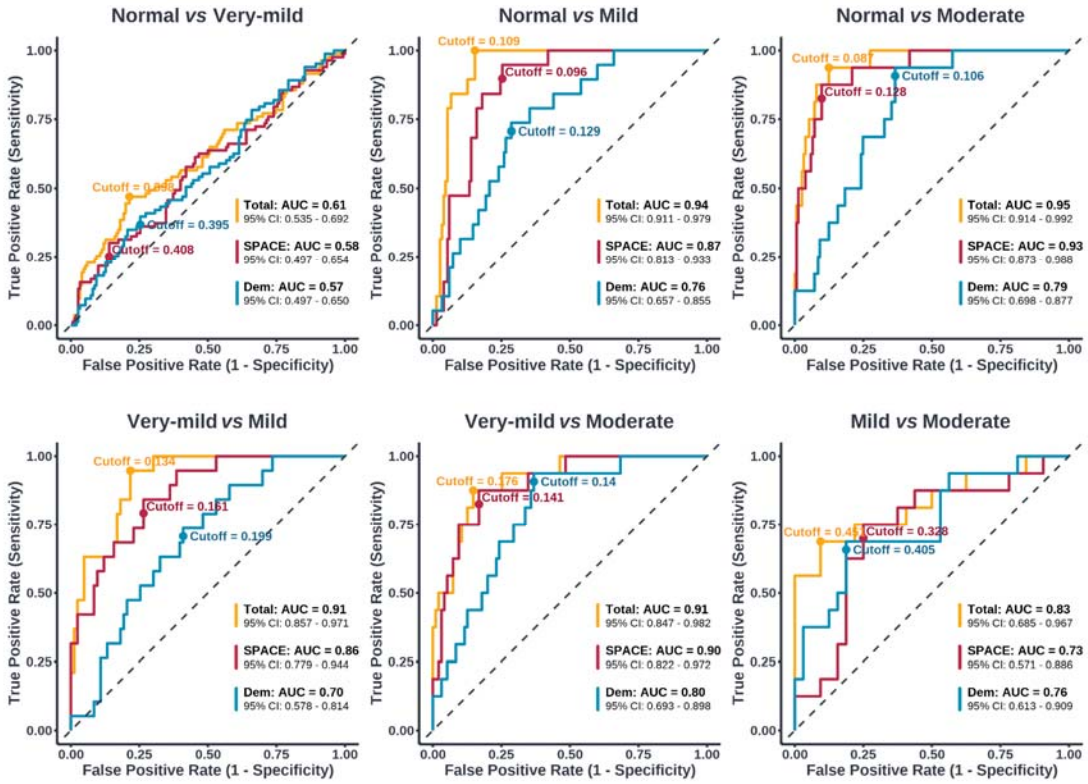

Figure 1 | Receiver Operating Characteristic (ROC) curves comparing pairwise classification performance across contrasts between cognitive stages: Dem (sociodemographic variables, blue), SPACE (performance on training, path integration, and perspective taking tasks, red) and Total (Dem + SPACE, orange). When the moderate group was included in the comparison, the SPACE models excluded the path integration task, as only two participants in this group completed it. AUC values and 95% confidence intervals (CI) are reported for each model. Cutoff values reflect optimal thresholds based on the Youden Index. The diagonal dashed lines indicate chance-level performance.

Table 4 | Contribution of SPACE measures to model fit. Likelihood ratio tests comparing baseline demographic models to extended models including SPACE measures across clinical group contrasts.

Model	Comparison	$\Delta\chi^2$	df	p	p_{DeLong}
Normal *vs.* Very-mild					
	Age, Edu, Gender + SPACE	4.88	3	0.181	0.256
Normal *vs.* Mild					
	Age, Edu, Gender + SPACE	34.7	3	<.001	<.001
Normal *vs.* Moderate					
	Age, Edu, Gender + SPACE	35.4	2	<.001	<.001
Very-mild *vs.* Mild					
	Age, Edu, Gender + SPACE	32.9	3	<.001	<.001
Very-mild *vs.* Moderate					
	Age, Edu, Gender + SPACE	21.5	2	<.001	0.001
Mild *vs.* Moderate					
	Age, Edu, Gender + SPACE	5.07	2	0.079	0.138

Cross-validation performance

To further evaluate the classification performance of SPACE, we performed stratified k-fold cross-validation (k = 10) and Leave-One-Out Cross-Validation (LOOCV) for each of the contrasts (Table 5). In both analyses, SPACE retained high diagnostic accuracy across most group comparisons. As expected, classification between normal cognition and very-mild impairment produced the weakest results. However, performance remained strong in all other clinically relevant contrasts. For distinguishing normal cognition from mild impairment, k-fold cross-validation produced an AUC of 0.87 with sensitivity of 0.79 and specificity of 0.83. LOOCV yielded similar results with an AUC of 0.85, sensitivity of 0.79, and specificity of 0.82. For the comparison between very-mild and mild impairment, k-fold cross-validation produced an AUC of 0.80, with sensitivity of 0.74, and specificity of 0.75. LOOCV obtained an identical AUC of 0.80, sensitivity of 0.63, and specificity of 0.75. Importantly, in the comparison between normal cognition and moderate impairment, k-fold cross-validation and LOOCV reached AUCs of 0.89 and 0.88 respectively, with sensitivity higher than 0.69 and specificity above 0.86 in both cases.

Table 5 | Cross-validation performance of SPACE-based models. Classification accuracy of SPACE models across clinical groups using K-fold and LOOCV, reporting AUC, sensitivity, and specificity.

Outcome	Cross Validations									
	K-fold					Leave One Out				
	AUC	Low CI	Upp CI	Sens	Spec	AUC	Low CI	Upp CI	Sens	Spec
Normal *vs.* Very-mild	0.50	0.42	0.59	0.39	0.67	0.52	0.44	0.60	0.31	0.68
Normal *vs.* Mild	0.87	0.77	0.97	0.79	0.83	0.85	0.75	0.96	0.79	0.82
Normal *vs.* Moderate	0.89	0.81	0.97	0.69	0.86	0.88	0.80	0.96	0.75	0.87
Very-mild *vs.* Mild	0.80	0.69	0.92	0.74	0.75	0.80	0.68	0.92	0.63	0.75
Very-mild *vs.* Moderate	0.84	0.74	0.94	0.63	0.84	0.84	0.74	0.94	0.63	0.82
Mild *vs.* Moderate	0.71	0.54	0.88	0.56	0.88	0.67	0.49	0.84	0.50	0.84

Short SPACE

On average, participants took 281 seconds to complete the SPACE training, 465 seconds for path integration, and 386 seconds for perspective taking tasks. A repeated measures ANOVA revealed significant differences in time to completion for these tasks ($F_{(2, 498)} = 75.28$, $p < 0.001$, $\eta^2p = 0.232$). Participants completed the training phase significantly faster than both the path integration and perspective taking tasks. Additionally, participants were faster at completing the perspective taking task than the path integration task. All Bonferroni post-hoc comparisons were significant ($p_{Bonferroni} < 0.001$).

Table 6 | Classification performance of the short SPACE version (sSPACE). Predictive performance of sSPACE models using K-fold and LOOCV procedures across clinical group comparisons.

				Cross Validation – sSPACE					
				K-fold			LOOCV		
Outcome	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Normal *vs.* Very-mild	0.61	0.49	0.72	0.48	0.39	0.69	0.51	0.39	0.63
Normal *vs.* Mild	0.91	0.91	0.81	0.88	0.84	0.75	0.87	0.88	0.76
Normal *vs.* Moderate	0.95	0.94	0.88	0.89	0.69	0.86	0.88	0.75	0.87
Very-mild *vs.* Mild	0.82	0.84	0.78	0.75	0.59	0.75	0.75	0.56	0.75
Very-mild *vs.* Moderate	0.91	0.88	0.85	0.84	0.63	0.84	0.84	0.63	0.82
Mild *vs.* Moderate	0.83	0.69	0.91	0.71	0.56	0.88	0.67	0.50	0.84

To assess whether a more time-efficient and user-friendly version of SPACE could also effectively discriminate between CDR stages, we repeated the classification analyses using only the training phase and perspective taking task (Table 6). In this short version of SPACE (sSPACE), which required only an average of 667 seconds, stratified k-fold cross-validation for the normal *vs.* very-mild contrast yielded an AUC of 0.61 (sensitivity = 0.49, specificity = 0.72). Again, the clinically relevant normal *vs.* mild impairment contrast resulted in an AUC of 0.91, (0.91, 0.81), while the very-mild *vs.* mild comparison yielded an AUC of 0.82 (0.84, 0.78). Since the models with the moderate group did not include the path integration task, contrasts for sSPACE were limited to the normal, very-mild, and mild groups (Figure 2).

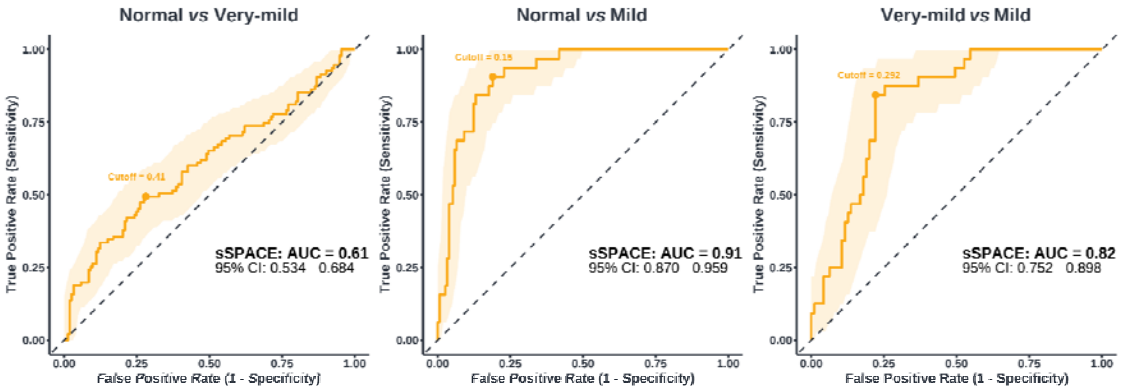

Figure 2 | ROC curves for the short version of SPACE (sSPACE). ROC curves showing the classification performance of sSPACE (comprising only the training and perspective taking tasks) for three contrasts: normal *vs.* very-mild, normal *vs.* mild, and very-mild *vs.* mild levels of impairment.

Discussion

We developed a tablet-based spatial navigation assessment (SPACE) and tested it in a clinical cohort of 300 individuals diagnosed with varying levels of cognitive impairment using the CDR scale. SPACE improved classification performance beyond demographic models across most group comparisons, including normal cognition *vs.* mild and moderate impairments, and very-mild *vs.* mild and moderate impairments. Classification accuracy was consistently high across these comparisons, with AUCs ranging from 0.91 to 0.95, sensitivities from 0.88 to 1.00, and specificities from 0.73 to 0.88. SPACE maintained high accuracy even when distinguishing between adjacent early stages, highlighting its potential for detecting subtle cognitive changes across various clinically diagnosed stages. We also evaluated a shorter version of SPACE (sSPACE), designed for rapid clinical screening and unsupervised deployment at home and in community settings. In under 11 minutes, sSPACE retained diagnostic accuracy comparable to the full version. Cross-validation analyses confirmed the reliability and robustness of these models, further highlighting their suitability for scalable cognitive screening. Altogether, these findings underscore the importance of incorporating spatial navigation tests into cognitive assessments and their sensitivity for detecting subtle differences in cognitive status.

While demographic models alone were able to achieve modest classification performance, the inclusion of SPACE consistently improved model accuracy, sensitivity, and specificity across all diagnostic contrasts, although gains were not significant when comparing normal cognition *vs.* very-mild impairment and mild *vs.* moderate impairment. As expected, SPACE increased the AUC

for the model distinguishing normal cognition *vs.* moderate impairment (AUC increase from 0.79 to 0.95), and when distinguishing very-mild *vs.* moderate impairment (AUC increase from 0.80 to 0.91). Across both of these contrasts, sensitivity and specificity remained high, ranging from 0.88 to 0.94, and 0.85 to 0.88, respectively. Notably, there was also a substantial improvement when differentiating normal cognition *vs.* mild impairment (AUC increase from 0.76 to 0.94) and very-mild *vs.* mild impairment (AUC increase from 0.70 to 0.91). Here, we also observed substantial gains in both sensitivity and specificity. For the contrast between normal cognition *vs.* mild impairment, sensitivity increased from 0.74 to 1.00 and specificity from 0.71 to 0.85. Similarly, for the contrast between very-mild *vs.* mild impairment, sensitivity rose from 0.74 to 0.95 and specificity from 0.59 to 0.73.

Critically, these findings were validated using stratified 10-fold cross-validation and LOOCV, with both methods yielding similarly high AUCs, sensitivity, and specificity across most contrasts. Previous research has shown that performance on virtual cognitive mapping tasks, and not episodic memory, can predict disease progression from normal cognition to impairment on the CDR, with reported diagnostic accuracy of AUC = 0.86[27]. Recently, an unsupervised mobile app focusing on episodic memory demonstrated promising cross-validated performance in distinguishing mild cognitive impairment from unimpaired individuals in memory clinic settings, with an AUC of 0.83, increasing to 0.87 when based on multiple sessions[40]. Building on this evidence, we show that a tablet-based spatial navigation test battery is able to achieve comparable or superior classification accuracies (AUCs = 0.94–0.95), confirmed through cross-validation, both for detecting mild and moderate impairments. Importantly, deficits in spatial navigation appear to be an excellent marker for discretising subtle clinical stages such as those from very-mild to mild cognitive decline. These results underscore the potential of spatial navigation assessments for complementing established memory-based tools.

SPACE classification performance for normal *vs.* very-mild impairment and mild *vs.* moderate impairment was modest. This result is likely due to a small overlap between the clinical classification of these disease stages, rather than a limitation of the navigation tests themselves[45]. The CDR classification between very-mild impairment and cognitively normal (CDR = 0.0-0.5) is based on the detection of subtle cognitive deficits, mainly relying on memory performance during the interview. In fact, previous investigations have sometimes sidestepped this challenge by simply combining the classifications of very-mild and mild impairment[46,47] or by completely omitting one of the categories[48]. In future, additional metrics (e.g. imaging and fluid biomarkers, APOE status, modifiable risk factors) could be administered together with SPACE to improve the efficacy of identifying the extremely subtle differences between disease stages[49]. However, even when addressing these marginal comparisons, SPACE consistently improved model performance compared to sociodemographic variables alone.

To address the challenges associated with typical, costly clinical assessments, it is critical that novel tools deliver efficient, scalable, and user-friendly digital solutions that can be deployed in community-based or other unsupervised screening contexts. By focusing on two of the most discriminatory tasks (i.e. training time and perspective taking), sSPACE considerably reduced the overall administration time (to < 11 min) of the full SPACE battery while maintaining strong diagnostic accuracy. Notably, our results revealed that sSPACE achieved AUCs of up to 0.91 in distinguishing cognitively normal participants from those with mild impairment, and 0.82 for differentiating very-mild from mild impairment. These values compare favorably to recent digital

drawing tests such as PENSIEVE-AI, which reported an AUC of 0.93 for distinguishing cognitively normal individuals from those with MCI/dementia in a large community sample[50]. Critically, the study grouped MCI and dementia cases together, and did not report classification performance for adjacent distinctions, such as between cognitively normal participants and MCI only. In addition, PENSIEVE-AI still requires some degree of supervision for more impaired cases and depends on a stylus-enabled tablet, which may limit scalability in non-clinical settings. Other unsupervised remotely administered tools, primarily based on episodic memory and recall, have demonstrated good, yet lower discriminatory power, with AUCs around 0.83[40]. In contrast, sSPACE combines strong diagnostic accuracy with the potential for fully unsupervised deployment, offering a practical and scalable solution for early cognitive screening. sSPACE can serve as an initial screening tool to flag individuals at risk, who may then undergo more detailed clinical assessment. This stepwise approach presents a viable opportunity for optimising resource management in healthcare.

Of all the spatial navigation tasks performed in SPACE, the training phase, path integration, and perspective taking were best able to discriminate levels of cognitive impairment. These navigation impairments likely reflect early pathology in regions critical for spatial orientation and memory[11,17,27] - deficits that are known to be more prominent in AD and MCI compared to normal aging[22,51]. A notable finding was the progressively longer times required to complete the SPACE training phase as the severity of cognitive impairment increased. This is likely due to deficits in both basic visuospatial integration[35,52,53] and fine motor skills for tracking objects during rotation and movement manoeuvers that are essential for completing the training phase. Similarly, a study using a smartphone-based app called Altoida showed that a tapping task added efficacy in distinguishing between MCI and early AD[41]. More recently, research has shown that reaching movements under visual interference are able to distinguish MCI from healthy aging[54]. Importantly, a novel cognitive-motor framework further implicates disrupted integration of motor and spatial systems as a driver of navigation errors in early-stage AD[53]. Thus, prolonged training time in cognitively impaired participants likely reflects not only spatial learning deficits but also deterioration of visuospatial and motor skills needed to interact with the software controls.

We found that deficits in the path integration task were sensitive indicators of early signs of cognitive impairment. Path integration has been widely studied and has consistently demonstrated a strong ability to discriminate across stages of cognitive impairment[29-32]. In our cohort, individuals with mild impairment exhibited significantly greater path integration errors than those with normal or very-mild cognition, aligning with prior evidence[23,24]. Path integration deficits have been shown to emerge in individuals at genetic risk for AD, appearing well before symptom onset and despite performing well in standard cognitive tests[29-31]. Importantly, deficits in path integration were associated with reduced grid-like representations[29,30] in the entorhinal cortex, as well as atrophy in both the entorhinal cortex[30] and hippocampus[23,24]. While these findings reinforce the interpretation of path integration tasks as sensitive markers of early AD, the inherent complexity of these tasks may limit their utility for widespread screening, particularly for elderly or unsupervised populations.

The perspective taking task in SPACE has emerged as a valuable marker, with individuals in the mild and moderate impairment groups exhibiting significantly greater errors compared to those with normal cognition or very-mild impairment. This task assesses the ability to mentally represent and manipulate spatial relationships from different viewpoints, tapping into both

visuospatial ability and executive function[55]. Indeed, research using immersive VR has found that early AD disrupts the ability to convert allocentric into egocentric information, highlighting compromised reference-frame transformations[56]. To our knowledge, only one study has specifically focused on perspective taking with amnestic MCI and AD participants, showing significant deficits in tasks requiring first-person viewpoint transformations[35]. Despite its simplicity and suitability for unsupervised deployment, perspective taking remains understudied in both research and clinical screening. In SPACE, we digitalised the Spatial Orientation Test[55], into a scalable, sensitive perspective taking assessment. However, its suitability for routine cognitive screening across diverse clinical populations requires further validation in broader clinical settings.

Despite the clear benefits of our digital assessment, several limitations should be acknowledged. First, the cross-sectional design limits conclusions about the ability of SPACE to track cognitive decline over time. Here, longitudinal studies are required to assess predictive validity. Second, although performance was robust under supervised conditions, the tools have not yet been evaluated in fully unsupervised settings, where variability in digital literacy and user familiarity with touchscreen interfaces may affect task performance. To mitigate these technological barriers, we previously conducted a detailed usability study that evaluated task controls, interface aids, and the configuration of landmarks in the virtual environment across age groups[44]. This work has allowed optimisation of the current design and reduced some of these confounds. Third, our findings were drawn from a single study cohort. As such, replication in other samples is necessary to support generalisability as well as differences in demographic characteristics. Finally, the absence of biomarker data restricts the interpretation of how well this tool correlates with other objective benchmarks of neuropathology.

We acknowledge that SPACE is not intended to replace comprehensive clinical or neuropsychological assessments. While SPACE does not offer a definitive diagnosis, this work has shown its clear potential for serving as a rapid, scalable tool for widespread screening of cognitive impairment. To maximise clinical relevance, SPACE and sSPACE metrics could also be integrated with complementary data sources such as neuropsychological assessments, fluid and imaging biomarkers, genetic profiles, and other clinical data[57]. A promising future direction is the continuous monitoring of spatial navigation ability, supported by mobile sensor technologies and integrated health data[43,58,59]. This approach could enable early detection of subtle cognitive decline, support longitudinal tracking of neurological changes, and contribute to digital phenotyping for MCI. Incorporating these tools into clinical workflows and public health systems may ultimately strengthen early detection strategies and support timely, targeted interventions.

In summary, this study provides clear evidence that spatial navigation tasks are a sensitive marker for detecting early cognitive impairment. Through immersive and engaging tasks that tap hippocampal and entorhinal function, the tablet-based SPACE tool provides a domain-specific and cost-effective alternative to traditional clinical assessments. Our results demonstrate strong diagnostic accuracy across cognitive levels, particularly across the very-mild and mild levels of impairment. The shortened version, sSPACE, maintains performance while reducing assessment time, hence, supporting its suitability for widespread and unsupervised deployment. This approach not only alleviates pressure on healthcare systems and extends reach to underserved populations. While additional validation in unsupervised environments is clearly needed, SPACE offers promising avenues for expanding access to cognitive screening and advancing early intervention strategies.

Methods

Participants

This study involved 300 patients (M_{age} = 74 years; 41% male) recruited either from the memory clinic (90%) or from the community (10%) at the National University Hospital (NUH) in Singapore. All recruited participants underwent an initial telephone screening to assess their eligibility for the study. Participants were eligible if they were over 50 years old and could walk 10 meters without using a walking aid. Participants were excluded if they manifested severe blindness or deafness, a history of seizure, epilepsy, or acute cardiac events. Signed informed consent was obtained from all participants or their legal representative prior to undergoing the experimental procedure. Ethical approval for this study was provided by the NHG Domain Specific Review Board (DSRB) in Singapore (2021/01160).

Sociodemographic and Health Questionnaire

Participants completed a sociodemographic and health questionnaire that included questions on their age, gender, education level, handedness, tablet experience, previous navigation training, vision impairments, chronic health conditions, history of head trauma and falls, depression, anxiety and stress, smoking status, alcohol consumption, sleep, and physical activity. Furthermore, participants were asked to self-evaluate their spatial navigation abilities using the Santa Barbara Sense of Direction (SBSOD) scale[55].

Global Clinical Dementia Rating Diagnosis

The global Clinical Dementia Rating (CDR) score was determined relative to six cognitive and functional domains: Memory, Orientation, Judgment & Problem Solving, Community Affairs, Home & Hobbies, and Personal Care. The CDR was calculated using the CDR-assignment algorithm based on the scoring rules established by Morris, 1993[60]. CDR scores were used to classify patients' impairment severity: 0 = no impairment, 0.5 = very mild, 1 = mild, 2 = moderate, and 3 = severe. Out of 300 participants, 153 had a CDR score of 0, 96 had a score of 0.5, 34 had a score of 1, 14 had a score of 2, and 3 had a score of 3. Due to the limited number of participants with a CDR score of 3 (n = 3), the moderate and severe groups were merged for analysis.

Spatial Performance Assessment for Cognitive Evaluation (SPACE)

SPACE is a serious game deployed on the iPad to assess spatial navigation abilities that may be indicative of cognitive impairment. In SPACE, participants assume the role of an astronaut tasked to explore and learn the position of different landmarks on a planet while completing a series of spatial navigation tasks (Figure 3). A comprehensive description of SPACE is provided in a previous publication[44].

SPACE begins with a training phase in which participants complete a series of manoeuvres designed to familiarise them with the controls while also assessing basic visuospatial skills. During this phase, participants learn how to rotate, move, and follow a robot from their starting point (rocket) to various destinations and back (Figure 3a). Following the training phase,

participants complete five spatial tasks that target different aspects of navigation. The path integration task evaluates how well participants can track changes in position and orientation as they navigate from the rocket to three landmarks. In each trial, participants follow the robot along two legs of a triangle and are then asked to return to the rocket via the third leg. At each landmark, the robot scans an item that participants will later be asked to recall (Figure 3b). The pointing task measures how accurately participants can indicate the direction of other landmarks from a first-person perspective. Over a series of trials, participants are placed in front of a landmark and asked to indicate the direction of other landmarks in the environment. The mapping task assesses their ability to form a top-down mental map of the environment by reconstructing the spatial layout of landmarks. The associative memory task tests their recall of the specific items scanned at each landmark. Finally, the perspective taking task measures participants' ability to calculate the spatial relation between the landmarks from various perspectives. Here, the player is asked to imagine standing at a landmark facing another landmark by looking at the map of the planet. Participants are then required to indicate the correct bearing towards a third landmark (Figure 3c).

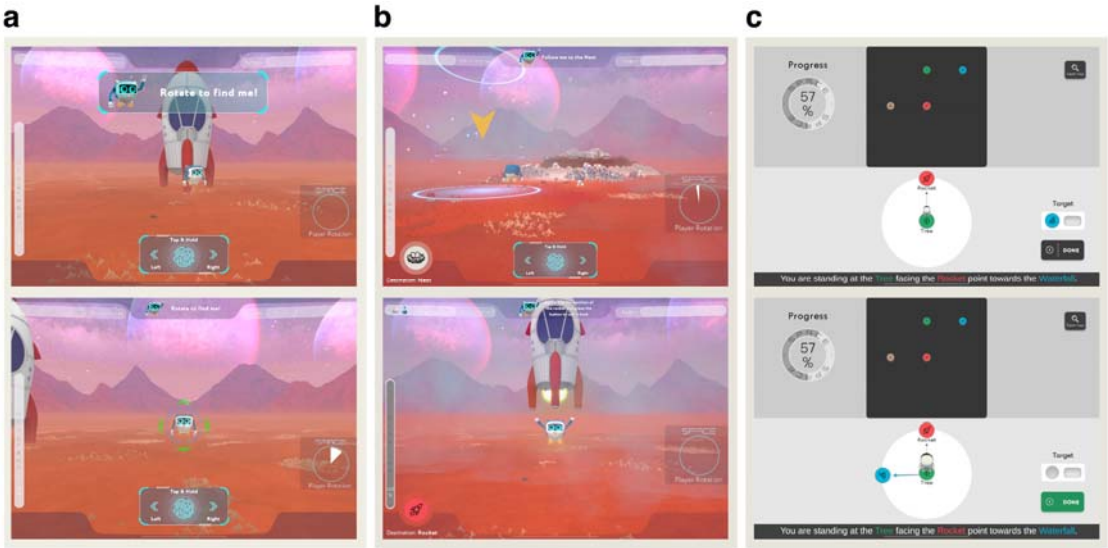

Figure 3 | Screenshots from SPACE tasks. Start (top images) and end (bottom images) screens from three core tasks within the SPACE assessment. (a) Training phase, rotation: Participants rotate their viewpoint by swiping left or right on the screen to find the robot. (b) Path integration: Starting from the rocket and following the robot to two landmarks, participants must navigate back to the starting location without assistance. (c) Perspective taking: Participants view a top-down map of the environment and are asked to imagine standing at one landmark (e.g., the tree), facing a second (e.g., the rocket), and pointing to a third (e.g., the waterfall) using the directional interface at the bottom of the screen.

Performance in SPACE is quantified using the following measures. Training time is the duration (in seconds) required for participants to complete the entire training task. Path integration distance error refers to the mean distance between the player's final position and the rocket's original position across all path integration trials. Pointing error is the mean absolute angular deviation (in degrees) from a starting position to the target location. Mapping accuracy is determined using bidimensional regression[61,62], a statistical method for assessing the degree of association (R^2) between the reference map and the map generated by the participants. The associative memory score is the percentage of correct pairings between scanned items and corresponding landmarks. Perspective taking error is the mean absolute angular deviation (in degrees) between the participant's directional estimate and the target landmark across all trials.

This study is part of a wider data collection effort to validate SPACE in community and clinical samples. Since the primary focus of this paper is to assess the performance of SPACE relative to the CDR scores, various other measures were excluded from the study. A complete list of all measures collected in the study is reported in the Supplementary Materials.

Study procedure

Upon arriving at the study site at the NUH, participants were provided with an overview of the study procedures. Once written informed consent was obtained, participants completed a series of questionnaires in which we collected sociodemographic data, self-reported navigation ability, and health information. Participants then completed a battery of neuropsychological tests and SPACE. After completing SPACE, blood and saliva samples were collected for biomarker analyses. Participants also completed a short gait assessment using single and dual tasks. The entire testing procedure lasted approximately 3 hours (see Supplementary Materials).

Statistical analyses

Descriptive statistics were computed for key demographic and clinical variables, including age, gender, education, and CDR scores. Group differences for performance on the spatial navigation tasks of SPACE were examined using one-way robust ANOVAs using a modified one-step estimator and bootstrap (5,000 samples). Post-hoc pairwise comparisons were conducted to evaluate the significance of group differences.

Logistic regressions were used to model the relationship between predictors and all combinations of binary outcomes from the CDR scale. To assess the contribution of SPACE, we compared models that included only demographic predictors (age, education, gender) with those that also incorporated the navigation tasks of SPACE and performed DeLong's test to evaluate whether the difference in AUCs was statistically significant. The optimal classification cutoff for each model was determined using Youden index, and assessed using the area under the ROC curve, sensitivity, specificity, and accuracy. Model performance was further evaluated using deviance, AIC, and the Brier score. Given that only two participants with moderate impairment completed the path integration task, this task was excluded from the analysis for this group.

We cross-validated our models using k-fold cross-validation and Leave-One-Out Cross-Validation (LOOCV). For both methods, we provide estimates of model performance, including AUC, sensitivity, and specificity, along with their corresponding confidence intervals. The k-fold cross-validation involved partitioning the dataset into k subsets ($k = 10$), training the model on k-1 subsets, and testing it on the remaining subset. LOOCV involves training the model on all but one data point and testing it on the remaining point, iterating this process for each data point in the dataset.

We also evaluated whether a short version of SPACE without the path integration task could discriminate between different CDR levels. A Repeated Measures ANOVA was employed to assess differences in the time taken to complete the three tasks (training, path integration, and perspective taking). Logistic regressions, ROC curve analysis, and cross-validation (stratified k-fold and LOOCV) were also performed on the short version of SPACE. All statistical analyses and plots were performed using Jamovi (v.2.6.26.0) and RStudio (v.2024.12.0+467), with p-values < 0.05 considered statistically significant.

References

1. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* **7**, e105–e125 (2022).
2. Livingston, G. *et al.* Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *Lancet* (2024) doi:10.1016/s0140-6736(24)01296-0.
3. Nandi, A. *et al.* Global and regional projections of the economic burden of Alzheimer's disease and related dementias from 2019 to 2050: A value of statistical life approach. *EClinicalMedicine* **51**, 101580 (2022).
4. Alzheimer's Association. 2024 Alzheimer's disease facts and figures. *Alzheimers. Dement.* **20**, 3708–3821 (2024).
5. Ngandu, T. *et al.* A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *Lancet* **385**, 2255–2263 (2015).
6. Gustavsson, A. *et al.* Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimers. Dement.* **19**, 658–670 (2023).
7. Niotis, K., Saperia, C., Saif, N., Carlton, C. & Isaacson, R. S. Alzheimer's disease risk reduction in clinical practice: a priority in the emerging field of preventive neurology. *Nat. Ment. Health* **2**, 25–40 (2024).
8. van der Flier, W. M., de Vugt, M. E., Smets, E. M. A., Blom, M. & Teunissen, C. E. Towards a future where Alzheimer's disease pathology is stopped before the onset of dementia. *Nat. Aging* **3**, 494–505 (2023).
9. Aisen, P. S., Jimenez-Maggiora, G. A., Rafii, M. S., Walter, S. & Raman, R. Early-stage Alzheimer disease: getting trial-ready. *Nat. Rev. Neurol.* **18**, 389–399 (2022).
10. Hampel, H. *et al.* Designing the next-generation clinical care pathway for Alzheimer's disease. *Nat. Aging* **2**, 692–703 (2022).
11. Coughlan, G., Laczó, J., Hort, J., Minihane, A.-M. & Hornberger, M. Spatial navigation deficits - overlooked cognitive marker for preclinical Alzheimer disease? *Nat. Rev. Neurol.* **14**, 496–506 (2018).
12. Lithfous, S., Dufour, A. & Després, O. Spatial navigation in normal aging and the prodromal stage of Alzheimer's disease: insights from imaging and behavioral studies. *Ageing Res. Rev.* **12**, 201–213 (2013).
13. Jack, C. R., Jr *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers. Dement.* **14**, 535–562 (2018).
14. Brier, M. R. *et al.* Tau and A□ imaging, CSF measures, and cognition in Alzheimer's disease. *Sci. Transl. Med.* **8**, 338ra66–338ra66 (2016).
15. de Calignon, A. *et al.* Propagation of tau pathology in a model of early Alzheimer's disease. *Neuron* **73**, 685–697 (2012).
16. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259 (1991).
17. Storandt, M., Mintun, M. A., Head, D. & Morris, J. C. Cognitive decline and brain volume loss as signatures of cerebral amyloid-beta peptide deposition identified with Pittsburgh compound B: cognitive decline associated with Abeta deposition. *Arch. Neurol.* **66**, 1476–1481 (2009).
18. O'Keefe, J. & Nadel, L. The hippocampus as a cognitive map. *Behav. Brain Sci.* **2**, 487–494 (1979).
19. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
20. deIpolyi, A. R., Rankin, K. P., Mucke, L., Miller, B. L. & Gorno-Tempini, M. L. Spatial cognition and the human navigation network in AD and MCI. *Neurology* **69**, 986–997 (2007).
21. Tangen, G. G., Nilsson, M. H., Stomrud, E., Palmqvist, S. & Hansson, O. Spatial navigation and its association with biomarkers and future dementia in memory clinic patients without dementia. *Neurology* **99**, e2081–e2091 (2022).

22. Hort, J. *et al.* Spatial navigation deficit in amnestic mild cognitive impairment. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 4042–4047 (2007).

23. Castegnaro, A., Ji, Z., Rudzka, K., Chan, D. & Burgess, N. Overestimation in angular path integration precedes Alzheimer's dementia. *Curr. Biol.* (2023) doi:10.1016/j.cub.2023.09.047.

24. Howett, D. *et al.* Differentiation of mild cognitive impairment using an entorhinal cortex-based test of virtual reality navigation. *Brain* **142**, 1751–1766 (2019).

25. Mapstone, M., Steffenella, T. M. & Duffy, C. J. A visuospatial variant of mild cognitive impairment: getting lost between aging and AD. *Neurology* **60**, 802–808 (2003).

26. Cushman, L. A., Stein, K. & Duffy, C. J. Detecting navigational deficits in cognitive aging and Alzheimer disease using virtual reality. *Neurology* **71**, 888–895 (2008).

27. Levine, T. F. *et al.* Spatial navigation ability predicts progression of dementia symptomatology. *Alzheimers. Dement.* **16**, 491–500 (2020).

28. Tu, S. *et al.* Lost in spatial translation - A novel tool to objectively assess spatial disorientation in Alzheimer's disease and frontotemporal dementia. *Cortex* **67**, 83–94 (2015).

29. Newton, C. *et al.* Entorhinal-based path integration selectively predicts midlife risk of Alzheimer's disease. *Alzheimers Dement* **20**, 2779–2793 (2024).

30. Bierbrauer, A. *et al.* Unmasking selective path integration deficits in Alzheimer's disease risk carriers. *Sci. Adv.* **6**, eaba1394 (2020).

31. Coughlan, G. *et al.* Toward personalized cognitive diagnostics of at-genetic-risk Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 9285–9292 (2019).

32. Segen, V., Ying, J., Morgan, E., Brandon, M. & Wolbers, T. Path integration in normal aging and Alzheimer's disease. *Trends Cogn. Sci.* **26**, 142–158 (2022).

33. Minta*, K. *et al.* SPACE: A novel digital tool for assessing hippocampal structural integrity in older adults. *Research Square* (2024) doi:10.21203/rs.3.rs-5484040/v1.

34. Mokrisova, I. *et al.* Real-space path integration is impaired in Alzheimer's disease and mild cognitive impairment. *Behav. Brain Res.* **307**, 150–158 (2016).

35. Marková, H., Laczó, J., Andel, R., Hort, J. & Vlček, K. Perspective taking abilities in amnestic mild cognitive impairment and Alzheimer's disease. *Behav. Brain Res.* **281**, 229–238 (2015).

36. Laczó, M. *et al.* Spatial navigation and visuospatial strategies in typical and atypical aging. *Brain Sci.* **11**, 1421 (2021).

37. Coutrot, A. *et al.* Global Determinants of Navigation Ability. *Curr. Biol.* **28**, 2861–2866.e4 (2018).

38. Goodroe, S. *et al.* Predicting real-world navigation performance from a virtual navigation task in older adults. *PLoS One* **20**, e0317026 (2025).

39. Thompson, L. I. *et al.* Remote and in-clinic digital cognitive screening tools outperform the MoCA to distinguish cerebral amyloid status among cognitively healthy older adults. *Alzheimers Dement. (Amst.)* **15**, (2023).

40. Berron, D. *et al.* A remote digital memory composite to detect cognitive impairment in memory clinic samples in unsupervised settings using mobile devices. *NPJ Digit Med* **7**, 79 (2024).

41. Meier, I. B. *et al.* Using a Digital Neuro Signature to measure longitudinal individual-level change in Alzheimer's disease: the Altoida large cohort study. *npj Digital Medicine* **4**, (2021).

42. Öhman, F., Hassenstab, J., Berron, D., Schöll, M. & Papp, K. V. Current advances in digital cognitive assessment for preclinical Alzheimer's disease. *Alzheimers Dement. (Amst.)* **13**, e12217 (2021).

43. Butler, P. M. *et al.* Smartwatch- and smartphone-based remote assessment of brain health and detection of mild cognitive impairment. *Nat. Med.* **31**, 829–839 (2025).

44. Colombo, G. *et al.* Detecting cognitive impairment through an age-friendly serious game: The development and usability of the Spatial Performance Assessment for Cognitive Evaluation (SPACE). *Comput. Human Behav.* **160**, 108349 (2024).

45. Wada-Isoe, K. *et al*. Global clinical dementia rating score of 0.5 may not be an accurate criterion to identify individuals with mild cognitive impairment. *J. Alzheimers Dis. Rep.* **3**, 233–239 (2019).

46. Kleiman, M. J., Barenholtz, E., Galvin, J. E. & Alzheimer's Disease Neuroimaging Initiative. Screening for early-stage Alzheimer's disease using optimized feature sets and machine learning. *J. Alzheimers. Dis.* **81**, 355–366 (2021).

47. Alim-Marvasti, A. *et al*. Validation of a rapid remote digital test for impaired cognition using clinical dementia rating and mini-mental state examination: An observational research study. *Front. Digit. Health* **4**, 1029810 (2022).

48. An, D., Shin, J. S., Bae, N., Seo, S. W. & Na, D. L. Validity of the tablet-based digital cognitive test (SCST) in identifying different degrees of cognitive impairment. *J. Korean Med. Sci.* **39**, e247 (2024).

49. Polk, S. E. *et al*. A scoping review of remote and unsupervised digital cognitive assessments in preclinical Alzheimer's disease. *NPJ Digit. Med.* **8**, 266 (2025).

50. Liew, T. M. *et al*. PENSIEVE-AI a brief cognitive test to detect cognitive impairment across diverse literacy. *Nat. Commun.* **16**, 2847 (2025).

51. Lester, A. W., Moffat, S. D., Wiener, J. M., Barnes, C. A. & Wolbers, T. The Aging Navigational System. *Neuron* **95**, 1019–1035 (2017).

52. Mitolo, M. *et al*. Visuospatial memory and neuroimaging correlates in mild cognitive impairment. *J. Alzheimers. Dis.* **35**, 75–90 (2013).

53. Hill, P. F. & Ekstrom, A. D. A cognitive-motor framework for spatial navigation in aging and early-stage Alzheimer's disease. *Cortex* **185**, 133–150 (2025).

54. Ilardi, C. R. *et al*. Deficits in reaching movements under visual interference as a novel diagnostic marker for mild cognitive impairment. *Sci. Rep.* **15**, 1901 (2025).

55. Kozhevnikov, M. & Hegarty, M. A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem. Cognit.* **29**, 745–756 (7 2001).

56. Morganti, F., Stefanini, S. & Riva, G. From allo- to egocentric spatial ability in early Alzheimer's disease: a study with virtual reality spatial tasks. *Cogn. Neurosci.* **4**, 171–180 (2013).

57. Berron, D. *et al*. Remote and unsupervised digital memory assessments can reliably detect cognitive impairment in Alzheimer's disease. *Alzheimers. Dement.* **20**, 4775–4791 (2024).

58. Marquardt, J. *et al*. Identifying older adults at risk for dementia based on smartphone data obtained during a wayfinding task in the real world. *PLOS Digit. Health* **3**, e0000613 (2024).

59. Ghosh, A., Puthusseryppady, V., Chan, D., Mascolo, C. & Hornberger, M. Machine learning detects altered spatial navigation features in outdoor behaviour of Alzheimer's disease patients. *Sci. Rep.* **12**, 3160 (2022).

60. Morris, J. C. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* **43**, 2412–2414 (1993).

61. Friedman, A. & Kohler, B. Bidimensional Regression: Assessing the Configural Similarity and Accuracy of Cognitive Maps and Other Two-Dimensional Data Sets. *Psychol. Methods* **8**, 468–491 (2003).

62. Tobler, W. R. Computation of the correspondence of geographical patterns. *Pap. Reg. Sci.* **15**, 131–139 (1965).

63. Chai, Y. L. *et al*. Apolipoprotein □4 is associated with dementia and cognitive impairment predominantly due to Alzheimer's disease and not with vascular cognitive impairment: A Singapore-based cohort. *J. Alzheimers. Dis.* **51**, 1111–1118 (2016).

Author Contributions

G.C., K.M., and V.R.S. conceived the study. G.C., J.G., and V.R.S. were responsible for software conceptualisation and development. E.C., J.R.C., and M.K.P.L. were responsible for participant recruitment and data collection. E.C. and C.P.C. provided clinical oversight and contributed to data collection coordination. G.C. and V.R.S. curated and analysed the data. G.C., K.M., and V.R.S. contributed to data interpretation. K.M. procured ethics approval. G.C. wrote the first draft of the manuscript and prepared the data visualisations. G.C., W.R.T., and V.R.S. contributed to critical revisions and substantive refinement of the manuscript. V.R.S. supervised the project. All authors reviewed and approved the final manuscript.

Acknowledgments

This research is supported by the National Research Foundation Singapore (NRF) under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.