



# Kent Academic Repository

Liu, Zhao, Soria, Daniele, Jie, Daniel Lai, Zhang, Jinbao, Shergill, Sukhi and Ang, Chee Siang (2026) *The use of fully immersive virtual reality for screening neurodegenerative diseases. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 18 (1). ISSN 2352-8729.

## Downloaded from

<https://kar.kent.ac.uk/112338/> The University of Kent's Academic Repository KAR

## The version of record is available from

<https://doi.org/10.1002/dad2.70244>

## This document version

Publisher pdf

## DOI for this version

## Licence for this version

CC BY (Attribution)

## Additional information

## Versions of research works

### Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

### Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal**, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

## Enquiries

If you have questions about this document contact [ResearchSupport@kent.ac.uk](mailto:ResearchSupport@kent.ac.uk). Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

## REVIEW ARTICLE

# The use of fully immersive virtual reality for screening neurodegenerative diseases: A systematic review of behavioral and diagnostic outcomes

Zhao Liu<sup>1</sup> | Daniele Soria<sup>1</sup> | Daniel Jie Lai<sup>2,3</sup> | Jinbao Zhang<sup>4</sup> | Sukhi Shergill<sup>5,6</sup> | Chee Siang Ang<sup>1,6,7</sup>

<sup>1</sup>School of Computing, University of Kent, Canterbury, UK

<sup>2</sup>School of Sciences, Psychology, Arts and Humanities, Computing, Engineering, and Sport, Canterbury Christ Church University, Canterbury, UK

<sup>3</sup>National Institute of Health and Care Research, Applied Research Collaboration Kent, Surrey and Sussex, Mill View Site, Hove, UK

<sup>4</sup>Care and Outcomes Research Centre, University of Kent, Canterbury, UK

<sup>5</sup>Department of Psychosis Studies, Institute of Psychiatry, Psychology Neuroscience, King's College London, Strand, London, UK

<sup>6</sup>Kent and Medway Medical School, University of Kent, Canterbury, UK

<sup>7</sup>National Heart & Lung Institute, Faculty of Medicine, Imperial College London, South Kensington Campus, London, UK

## Correspondence

Zhao Liu, School of Computing, University of Kent, University Road, Canterbury CT2 7NZ, Kent, UK.  
Email: [zl239@kent.ac.uk](mailto:zl239@kent.ac.uk)

Daniele Soria, Sukhi Shergill, and Chee Siang Ang are joint senior authors on this work.

## Funding information

China Scholarship Council (CSC), Grant/Award Number: 202208150024

## Abstract

Early detection of Alzheimer's disease (AD), Parkinson's disease (PD), and mild cognitive impairment (MCI) is crucial for timely intervention. Traditional cognitive screening tools lack ecological validity and sensitivity. Virtual reality (VR) provides realistic, controlled environments for assessing multidimensional cognition. This systematic review evaluated the diagnostic accuracy, feasibility, and applicability of immersive VR assessments for neurodegenerative screening. We searched PubMed, PsycINFO, and Embase for studies published June 2005 to April 2024. Eligible studies used head-mounted displays in adults with MCI, early AD/PD, or dementia. Ten studies ( $n = 472$ ) met criteria. Tasks targeted spatial memory, executive function, attention, and navigation. Several reported strong discriminations (area under the curve up to 0.89) and, when combined with machine learning, accuracies of 87% to 100%. Immersive VR shows promise as an ecologically valid, engaging, and scalable screening approach; however, standardization of tasks and outcomes, real-world validation, and robust longitudinal evidence are needed to support clinical adoption.

## KEYWORDS

Alzheimer's disease, cognitive screening, ecological validity, immersive virtual reality, machine learning, mild cognitive impairment, neuropsychology

## Highlights

- This review systematically describes the application of fully immersive virtual reality (VR) in the early screening of neurodegenerative diseases, with a focus on studies using head-mounted devices to simulate real-life tasks.
- Task types such as spatial memory, daily living simulations, and executive function assessments have demonstrated high sensitivity and specificity in diagnosing mild cognitive impairment (MCI) and early-stage Alzheimer's disease (AD).

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* published by Wiley Periodicals, LLC on behalf of Alzheimer's Association.

- Approximately one third of studies combined machine learning techniques to analyze multimodal behavioral data (e.g., path deviations, task duration, and language responses), significantly improving diagnostic accuracy.
- This study highlights methodological heterogeneity, small sample sizes, and the lack of longitudinal studies as current research limitations, and calls for future standardized, multicenter, and long-term follow-up studies to validate the predictive validity and real-world applicability of VR tools.

## 1 | INTRODUCTION

With the global population aging rapidly, neurodegenerative diseases (NDs), such as Alzheimer's disease (AD) and Parkinson's disease (PD), have emerged as some of the most pressing public health challenges of our time.<sup>1,2</sup> These disorders are characterized by a progressive deterioration of the structure and function of the nervous system, leading to a gradual decline in cognitive, behavioral, and motor abilities.<sup>3</sup> Beyond the profound physical and psychological toll on individuals, NDs also impose substantial caregiving demands and financial burdens on families and health-care systems.<sup>4</sup> It is estimated that > 55 million people worldwide are currently living with dementia, with nearly 10 million new cases diagnosed each year.<sup>5</sup> In the United Kingdom alone, the economic burden of dementia reached £34.7 billion in 2020, a figure expected to rise dramatically in the coming decades due to demographic shifts and increasing demand for long-term care services.<sup>6</sup>

Timely identification of neurodegenerative diseases is critical for implementing early interventions and optimizing clinical outcomes. However, conventional screening methods, such as the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA), present notable limitations.<sup>7</sup> Although these tools are widely used for their simplicity and clinical utility, they rely primarily on paper-and-pencil formats that may fail to capture the complexities of cognitive functioning in real-world environments. Moreover, these tools often lack sensitivity to early-stage cognitive impairments and are susceptible to practice effects with repeated administration, potentially compromising diagnostic accuracy over time.<sup>7</sup>

To address these limitations, researchers and clinicians have increasingly turned to digital technologies to modernize cognitive assessment practice.<sup>8</sup> An alternative approach is to use computerized tests as screening tools, such as Cogstate<sup>9</sup> and the Cleveland Clinic Cognitive Battery (C3B).<sup>10</sup> They not only include enhanced measurement precision and self-administration, but also cost effectiveness and reduced examiner bias.<sup>11</sup> Moreover, mobile applications and tablet-based assessments have gained traction in both clinical and community settings.<sup>12</sup> However, compared to performance on traditional neuropsychological testing (e.g., Wisconsin Card Sorting Test, Stroop Test), there is a weak correspondence with activities of daily living.<sup>13</sup> A more ecologically meaningful neuropsychological

testing approach involves constructing a function-led assessment.<sup>14</sup> thereby enabling direct observation of everyday behavior and facilitating the tracing of underlying neuropsychological mechanisms. Conducting such assessments in naturalistic settings, however, is expensive, time consuming, logistically complex, and difficult to standardize.<sup>15</sup>

Virtual reality (VR) offers a promising solution to these requirements. Basically, the VR system can be classified into three levels: non-immersive (e.g., computer displays), semi-immersive (e.g., projection screens), and fully immersive.<sup>16</sup> Full-immersive VR (FIVR) encompasses immersion, interaction, and visual realism through head-mounted displays (HMDs), surround sound, and other input devices.<sup>17</sup> FIVR can also evaluate participants' clinical, emotional, and social processing abilities in real time, and their approach is more closely aligned with real-world functional performance.<sup>18</sup> This enables cognitive testing tasks to simulate real-life daily activities for assessment purposes,<sup>19</sup> particularly in individuals with mild cognitive impairment (MCI) or early-stage ND.<sup>20</sup> Crucially, this immersion allows FIVR to uniquely target the cognitive domains most vulnerable in early neurodegeneration, such as complex executive functions and spatial navigation, which are often poorly captured by static 2D tests. For instance, the HMD-based six-domain battery (Cognitive Assessment by Virtual Reality [CAVIRE-2]) effectively distinguishes MCI patients from cognitively healthy adults (area under the curve [AUC] = 0.88; sensitivity = 88.9%; specificity = 70.5%), providing cutoff values stratified by age/education level.<sup>21</sup> Moreover, by casting tasks as instrumental activities of daily living (IADL), FIVR can simulate complex, lifelike settings while allowing for precise control over experimental variables and automatic data capture. Tasks such as navigating a virtual supermarket (e.g., VStore<sup>22</sup>), crossing a street,<sup>23</sup> or performing a series of routine household actions,<sup>24</sup> MCI patients wearing HMDs completed a point-of-interest (POI) task by physically walking within a virtual environment. The result was AUC = 0.90 for distinguishing "high/low AD risk MCI," significantly outperforming the AUC = 0.57 of the best paper-and-pencil test.<sup>25</sup>

The gamified elements were incorporated into the task, enhancing user engagement and motivation and reducing the anxiety associated with it.<sup>26</sup> These platforms can also capture rich behavioral data such as reaction times, eye movements, movement patterns, and task strategies, providing a multidimensional view of cognitive functioning.<sup>27</sup>

Recent studies have demonstrated the potential of VR-based assessments to differentiate between healthy controls and individuals with cognitive impairment, such as the Cognition Assessment in Virtual Reality (CAVIR)<sup>28</sup> and the RE@CH Assessment Module.<sup>8</sup> Preliminary evidence also indicates strong user engagement with interventions among MCI patients and suggests that performance in VR tasks is sensitive to cognitive status.<sup>26</sup>

However, the clinical utility of FIVR is moderated by critical user-level factors. Performance can be significantly influenced by age, education, and digital literacy, which may increase false positives or mask early deficiencies.<sup>29</sup> For instance, older adults with lower digital familiarity may experience anxiety or difficulty with controllers, affecting task outcomes independent of cognitive status.<sup>30</sup> Additionally, sensory limitations and susceptibility to cybersickness vary widely among older populations and are exacerbated by hardware heterogeneity.<sup>29</sup> These factors underscore the need to evaluate not only diagnostic accuracy but also the feasibility and usability of these systems across diverse geriatric populations.

A growing body of review literature has explored the application of VR technology in cognitive research. Still, most studies have confused the levels of immersion and focused on intervention effects rather than assessment accuracy. For example, recent reviews on memory or spatial navigation have mixed different VR modes, failed to limit studies to HMD tasks, and often categorized patients by a single diagnosis (e.g., AD) rather than encompassing a holistic research of early neurodegenerative diseases such as MCI, AD, PD, and frontotemporal dementia (FD).<sup>30,31</sup> In contrast, this study strictly limited the inclusion of fully immersive HMD assessment tasks and categorized patients by early cognitive impairment domains (e.g., navigation/path integration, executive control, attention, and social cognition), extracting diagnostic and psychometric indicators to clarify clinical applicability—a focus largely absent in previous research. This review addressed a critical gap by synthesizing evidence on the diagnostic applications and implementation potential of such systems for early neurodegenerative screening. Specifically, the objectives of this review are: (1) to evaluate the diagnostic effectiveness of FIVR-based cognitive assessments in detecting impairments associated with AD and PD; and (2) to assess the feasibility, usability, and clinical utility of FIVR systems in supporting health-care professionals' diagnostic processes. Through this analysis, we aim to evaluate the current state of FIVR technology in cognitive diagnostics and its potential to transform neuropsychological assessment in the era of digital health care.

## 2 | MATERIALS AND METHODS

### 2.1 | Search strategy

The following databases were searched: PubMed, PsycINFO, and Embase. Reference lists of key articles were also screened to identify additional eligible studies. Given the rapid development of VR-based assessments, studies dated before June 22, 2005 were considered out-

dated and therefore excluded. Only studies published between June 22, 2005 and April 8, 2024 were included. No language restrictions were applied. June 2005 was selected as a pivotal juncture, coinciding with the release of widely adopted game engines such as Unity,<sup>32</sup> which fundamentally transformed the research landscape by lowering development barriers. This timing validates the transition from high-end visual simulation to accessible VR technology, a convergence trend that Zyda<sup>33</sup> identified as the genesis of modern immersive applications.

The protocol for this review was not registered prospectively. However, the review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines.

The Population/Intervention/Comparison/Outcome (PICO) framework was used to guide the development of search terms; however, the Population/Intervention/Outcome (PIO) model was adopted for this review as no restrictions were placed on the study design. The comparator component was excluded from the search strategy to maximize article retrieval. The search terms were structured as follows:

- Population (P): individuals with NDs.
- Intervention (I): fully immersive VR-based assessment.
- Outcome (O): Task performance metrics (e.g., accuracy, response time), results of standard cognitive assessments (e.g., MMSE, MoCA), and diagnostic validity indicators such as AUC, classifier accuracy and where applicable, sensitivity and specificity.

The details of the search strategy were: (Neurodegenerative diseases\* OR "neurodegenerative diseases"[MeSH] OR mild cognitive impairment OR dementia OR Alzheimer OR Parkinson OR Huntington OR Amyotrophic lateral sclerosis OR motor neuron OR older) AND (cognitive assessment OR cognition testing OR cognition screening) AND (Game OR Gamification OR virtual reality OR immersive OR head-mounted display).

### 2.2 | Study inclusion and exclusion criteria

Eligible studies included: (1) Patients with a confirmed diagnosis of MCI or dementia who were aged  $\geq 60$  years (aligned with the World Health Organization definition<sup>34</sup>); healthy control groups of any adult age  $> 18$  were accepted for comparative analysis. Diagnostic confirmation was required using standard clinical guidelines (e.g., DSM-5) or validated neuropsychological assessments (e.g., MMSE, MoCA). (2) Interventions that were a fully immersive test run using VR HMDs. The specific task could be a cognitive test or a simulated work situation in everyday life. (3) Results of the experiment focused on a diagnosis or screening of the patient's cognitive state/condition. (4) Full text published in peer-reviewed journals.

Studies were excluded if (1) they only included a healthy participant group, (2) the age range and type of disease did not match, (3) they were qualitative studies only assessing user acceptability without cognitive evaluation, or (4) they were systematic reviews and meta-analyses, conference abstracts, and study protocols.

## 2.3 | Selection process

All retrieved records were imported into Rayyan,<sup>35</sup> a web-based systematic review platform. Duplicate records were automatically detected and resolved within Rayyan. The platform was subsequently used to facilitate blinded title/abstract screening and full-text review by the independent reviewers.

During the title and abstract screening stage, the lead author (Zhao Liu) conducted a preliminary screening of all records, and two independent reviewers (Daniel Jie Lai and Jinbao Zhang) each randomly assessed 50% of the studies for cross-validation. Agreement with Daniel Jie Lai was moderate (Cohen Kappa = 0.24, prevalence-adjusted and bias-adjusted kappa [PABAK] = 0.88, observed agreement = 93.9%), while the agreement with Jinbao Zhang was high (Kappa = 0.49, PABAK = 0.88, observed agreement = 94.2%).

The discrepancy between the relatively low Kappa values and the high observed agreement can be attributed to two main factors. First, statistically, this reflects the well-known “Kappa paradox”.<sup>36</sup> Second, in practice, discrepancies often arose from ambiguities in study abstracts, particularly regarding VR hardware (e.g., distinguishing HMDs from non-immersive screens).

During the full-text screening stage, Zhao Liu and Daniel Jie Lai independently reviewed six full-text articles and demonstrated perfect agreement (Kappa = 1.00, observed agreement = 100%). Zhao Liu and Jinbao Zhang reviewed 17 full-text articles and achieved substantial agreement (Kappa = 0.74, observed agreement = 88.2%). Two disagreements were resolved through discussion and subsequently included. All discrepancies were addressed through reviewer discussion.

## 2.4 | Data extraction

The following information was extracted from each included study:

1. Study characteristics, including author and year of publication, country, study design, participant characteristics, sample size (including presence or absence of a control group), and study duration. We extracted data as reported in the primary studies. When specific quantitative metrics (e.g., AUC) were missing, we reported available statistical outcomes (e.g., *p* values) and noted this limitation in the results table.
2. VR intervention characteristics, including the intervention description, type of VR intervention (e.g., assessment based, training based), and delivery format (e.g., VR devices, supporting systems).
3. To ensure a standardized synthesis of the heterogeneous outcomes, we adopted Bowen et al.’s framework for feasibility studies.<sup>37</sup> Data were mapped into four key domains:
4. Acceptability: How the intended recipients react to the intervention. We extracted data on tolerability (e.g., simulator sickness, adverse events), dropout rates due to discomfort, and user experience feedback.

5. Practicality: The extent to which the intervention can be delivered when resources are constrained. We extracted metrics on administration time, technical issues (e.g., software bugs, hardware failures), and ease of use.
6. Implementation: The extent to which the intervention can be successfully delivered to the target population. We extracted data on completion rates and adherence to the study protocol.
7. Limited efficacy: The promise of the intervention being successful with the intended population. We extracted diagnostic performance metrics, including sensitivity, specificity, AUC, and correlations with standard neuropsychological tests (e.g., MMSE, MoCA).

Data extraction was independently conducted by three reviewers (Zhao Liu, Daniel Jie Lai, and Jinbao Zhang). No automation tools were used for data extraction. Any discrepancies were resolved through discussion.

## 2.5 | Data synthesis and analysis

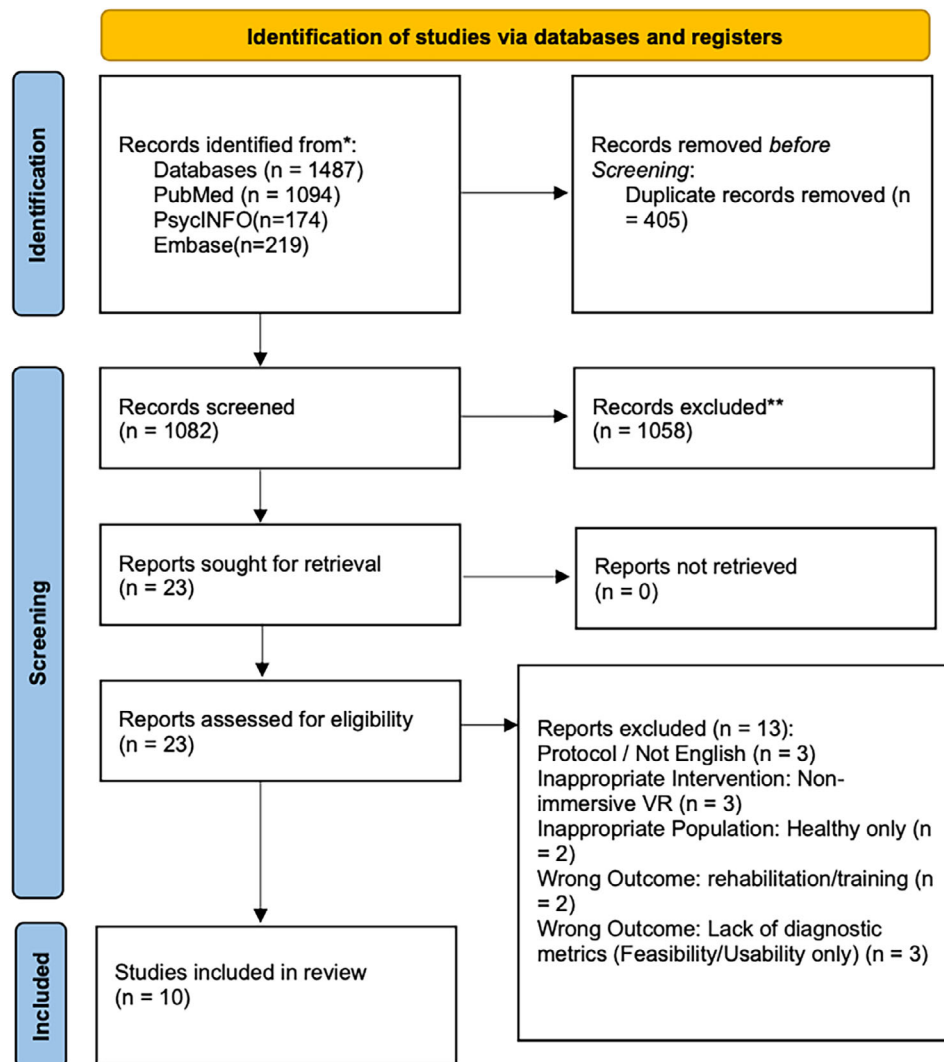
Given the heterogeneity of the included studies, a quantitative meta-analysis was not feasible. Therefore, we conducted a narrative synthesis structured around the Bowen et al.<sup>37</sup> framework outlined in section 2.4. Findings were systematically grouped into feasibility (implementation, practicality), acceptability, and diagnostic efficacy (limited efficacy) to facilitate cross-study comparison. Because of the qualitative nature of the meta-analysis, the effect indicators are presented using descriptive statistics from the original studies, including AUC, sensitivity, specificity, correlation coefficient (*r*), and *p* values for intergroup comparisons.

## 2.6 | Heterogeneity and bias assessment

Sources of heterogeneity (e.g., differences in VR hardware and control group selection) were qualitatively assessed by comparing study characteristics and outcomes. Sensitivity analysis was not performed because this review uses descriptive integration. Similarly, due to the limited number of included studies (*n* < 10 for any specific outcome measure), the risk of bias due to missing results (publication bias) was not formally assessed using methods such as funnel plots; this practice is in accordance with Cochrane guidelines.

## 2.7 | Risk of bias assessment

We assessed the methodological quality of the included studies using the QUADAS-2 tool (Quality Assessment of Diagnostic Accuracy Studies, version 2),<sup>38</sup> which is specifically designed to evaluate the risk of bias in diagnostic accuracy research. The tool comprises four domains: (1) patient selection, (2) index test, (3) reference standard, and (4) flow and timing. Each domain is assessed for risk of bias, while the first three domains are additionally evaluated for concerns regarding



**FIGURE 1** Flowchart of study selection based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. VR, virtual reality.

applicability. Assessment was performed independently by two reviewers (Zhao Liu and Daniele Soria), with any discrepancies resolved through discussion.

### 3 | RESULTS

Initial database searches identified a total of 1487 articles. After the removal of 405 duplicates, 1082 articles were retained for title and abstract screening. Based on predefined exclusion criteria, 1059 articles were subsequently excluded, leaving 23 studies eligible for full-text assessment. Upon detailed review, 10 articles met the inclusion criteria and were included in the current systematic review. A PRISMA flow diagram illustrating the detailed process of identification, screening, and inclusion of eligible studies is presented in Figure 1.

#### 3.1 | Study characteristics

This systematic review included 10 studies published between 2015 and 2023, conducted across multiple geographic regions, including Europe (UK, Spain), East Asia (China, Korea), North America (Canada), and South America (Brazil), see Table 1. A total of 472 participants were included, with individual sample sizes ranging from 12 to 108. The populations studied primarily consisted of individuals diagnosed with MCI, AD, or other early-stage neurodegenerative conditions. Study designs varied across the included studies, with 40% (4/10) being experimental studies,<sup>39–42</sup> 30% (3/10) comparative studies,<sup>20,43,44</sup> two feasibility studies,<sup>45,46</sup> and one validation study.<sup>39</sup> One of these studies additionally incorporated machine learning methods for predictive modelling.<sup>41</sup> The VR-based assessments evaluated a range of cognitive domains, including spatial orientation, object-location memory, executive function, language processing, and reaction time. VR



**TABLE 1** Summary of studies using immersive VR for cognitive assessment of MCI and early AD.

Author, year	Country	Study design	Population	Sample/control	Duration	Intervention	Type	Delivery format	Outcome measures	Main findings
Castegnaro et al., 2022 <sup>39</sup>	UK	Exp	aMCI vs. CG	23/24	50 min	Object-location memory task (OLT)	Ass	iVR (HTC Vive)	Spatial binding accuracy, AUC	aMCI showed impaired accuracy ( $p < 0.001$ ); AUC = 0.89 outperformed standard tests.
Wu et al., 2023 <sup>42</sup>	China	Exp	MCI vs. CG	44/42	N/A	VR speech task + EEG	Ass	Oculus Quest 2 + Muse 2	Multimodal data: speech, EEG, digital scores	SVM (RBF) accuracy 87%, voting fusion reached 89.8%.
Moussavi et al., 2022 <sup>43</sup>	Canada	Comp	AD/MCI/ Others	65/20/8	N/A	Spatial orientation task	Ass	Wheelchair-based VR	Spatial task scores	Differentiated AD and MCI ( $p < 0.001$ ); effective clinical screening.
Da Costa et al., 2022 <sup>47</sup>	Brazil	Vali	MCI vs. CG	19/29	N/A	SOIVET Maze and Route	Ass	Immersive VR maze	Accuracy, ROC	Moderate diagnostic accuracy; validated feasibility.
Campo-Prieto et al., 2023 <sup>45</sup>	Spain	Feas	PD	26	N/A	VR reaction time test	Ass	Immersive VR + TUG	RT, MMSE, TUG-Cognitive	RT negatively correlated with MMSE ( $p = -0.576$ ); useful fall risk indicator.
Park, 2022 <sup>44</sup>	South Korea	Comp	MCI vs. CG	36/56	N/A	SCT-VR spatial memory task	Ass	VR headset	SCT-VR, MoCA-K	SCT-VR: sensitivity 0.944, specificity 0.964; test-retest ICC = 0.982.
Jang et al., 2023 <sup>46</sup>	South Korea	Feas	MCI vs. CG	12/108	N/A	VR daily life simulation	Ass	Custom VR	VR score, MoCA, DHEA	AUC = 0.765; VR outperformed MoCA.
Serino et al., 2015 <sup>40</sup>	Italy	Exp	aMCI/AD/CG	13/14/14	3 trials	Spatial representation in VR	Ass	Custom immersive VR	Allocentric/ errors	aMCI had allocentric deficits; AD had integrative spatial impairment.
Bayahya et al., 2021 <sup>20</sup>	Saudi Arabia	Comp	Dem/MCI/CG	30/20/65	N/A	Smart 3D VR cognitive test	Ass	3D VR simulator	Mini-Cog, memory, visuospatial	Kappa agreement up to 100%; VR valid as screening tool.
Tsai et al., 2021 <sup>41</sup>	Taiwan, China	Exp	MCI/early AD vs. CG	6/6	N/A	VR supermarket + ML	Ass	HTC Vive VR + path tracking	46 behavioral & task indices	45/46 features significant; ML classifiers (e.g., SVM, XGBoost) accuracy = 100%.

Abbreviations: AD, Alzheimer's disease; aMCI, amnesic mild cognitive impairment; Ass, assessment; AUC, area under the curve; CG, control group; Comp, comparative; Dem, dementia; DHEA, dehydroepiandrosterone; EEG, electroencephalogram; Exp, experimental; Feas, feasibility; ICC, intra-class correlation; iVR, immersive virtual reality; MCI, mild cognitive impairment; ML, machine learning; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; PD, Parkinson's disease; ROC, receiver operating characteristic; RT, reaction time; SCT-VR, social cognition training with virtual reality; SOIVET, Spatial Orientation in an Immersive Virtual Environment Test; SVM, support vector machine; TUG, Timed Up and Go; Vali, validation; VR, virtual reality.

**TABLE 2** Assessment of feasibility and acceptability based on Bowen et al.'s framework.

Author, year	Acceptability (e.g., tolerance, satisfaction)	Practicality (e.g., time, tech issues)	Implementation (e.g., completion, attrition)
Castegnaro et al., 2022 <sup>39</sup>	No fatigue or motion sickness reported.	Total time $\approx$ 50 minutes (including set-up/practice).	100% completion ( $n=100$ ).
Wu et al., 2023 <sup>42</sup>	Not reported.	Used wearable EEG (MUSE 2) + Oculus Quest 2.	86 participants completed.
Moussavi et al., 2022 <sup>43</sup>	Wheelchair set-up reduced kinetosis; simple to use.	Quick administration (5–10 minutes).	Longitudinal retention low (30%), but cross-sectional $n=93$ analyzed.
Costa et al. 2021 <sup>47</sup>	Dropout rate: 11.4% (controls) and 14.8% (MCI) due to cybersickness.	Task shortened (18 turns) to reduce anxiety/difficulty.	8 participants dropped out (total $n=48$ analyzed).
Campo-Prieto et al. 2023 <sup>45</sup>	No adverse effects; 0 SSQ symptoms.	Short duration (60s test trials).	100% completion ( $n=26$ ).
Park 2022 <sup>44</sup>	Not reported.	Controller-based HMD set-up.	92 participants completed.
Jang et al. 2023 <sup>46</sup>	VRSQ scores indicated no severe sickness (mean $\approx$ 15–19).	Average time 18.9 minutes. 30 participants excluded due to software error.	120/150 completed (80% retention after tech exclusions).
Serino et al. 2015 <sup>40</sup>	1 aMCI patient failed to complete tasks (reason unspecified).	NeuroVirtual 3D platform used.	44/45 completed.
Bayahya et al. 2021 <sup>20</sup>	No withdrawal; reported as “user friendly.”	Time < 5 minutes.	100% completion ( $n=115$ ).
Tsai et al. 2021 <sup>41</sup>	No additional burden reported; enhanced immersion.	High-end PC required (RTX 2070) for rendering.	100% completion (small sample $n=12$ ).

Note: The “limited-efficacy” domain of Bowen et al.'s framework, which corresponds to diagnostic performance, is presented separately in Table 3 to allow for a detailed integration with quality assessment.

Abbreviations: aMCI, amnesic mild cognitive impairment; EEG, electroencephalogram; HMD, head-mounted device; MCI, mild cognitive impairment; SSQ, Simulator Sickness Questionnaire; VRSQ, Virtual Reality Sickness Questionnaire.

hardware predominantly involved commercially available immersive devices such as HTC Vive and Oculus Quest, and only one assessment reported that the duration was 50 minutes. All studies used FIVR systems to assess spatial memory, executive functions, or other cognitive domains associated with early neurodegenerative decline. The summaries of all studies are in Table 1.

### 3.2 | Feasibility and acceptability

A detailed feasibility and acceptability structure based on Bowen et al.'s<sup>37</sup> framework is presented in Table 2.

- **Acceptability:** The interventions were generally well tolerated, and participants reported high engagement, particularly in tasks with high ecological validity, such as the virtual supermarket.<sup>20,41</sup> Although some studies reported minimal adverse effects (e.g., Campo-Prieto et al.<sup>45</sup>), others noted dropout due to motion sickness, particularly in MCI groups (e.g., Da Costa et al.<sup>47</sup>).
- **Implementation:** Most studies demonstrated high feasibility. Completion rates were generally high across most studies with explicit rates > 90% in Castegnaro et al.<sup>39</sup> and Tsai et al.<sup>41</sup> and the full adherence to the task was reported by Campo-Prieto et al.<sup>45</sup> However, Moussavi et al.<sup>43</sup> reported that longitudinal retention was low (30%).

- **Practicality:** Most systems used commercially available headsets (e.g., HTC Vive, Oculus Quest), enabling practical deployment. A few studies didn't specify the assessment time, and most would take 5 to 20 minutes to complete the testing.
- **Limited efficacy:** The preliminary efficacy of these VR tools for diagnostic screening—including sensitivity, specificity, and discriminatory power—is detailed in section 3.3 and summarized in Table 3.

### 3.3 | Diagnostic performance and cognitive relevance

The diagnostic performance metrics and associated quality assessments (QUADAS-2) for all included studies are summarized in Table 3.

### 3.4 | Object localization and spatial memory tasks

Castegnaro et al.<sup>39</sup> developed three immersive VR subtasks, such as object location memory, object recognition, and object-context association, to evaluate object-space binding deficits in individuals with amnesic mild cognitive impairment (aMCI). The tasks were delivered via an HTC Vive platform and incorporated active navigation and multidimensional performance metrics (e.g., absolute distance error). Task performance was also correlated with cerebrospinal fluid



**TABLE 3** Summary of diagnostic performance and quality assessment (QUADAS-2).

Author, Year	Outcome measure	Comparison group	Diagnostic performance metrics	Patient selection	Index test	Reference standard	Flow and timing
Castegnaro et al., 2022 <sup>39</sup>	Object-location memory	aMCI vs. HC	AUC = 0.89	Low	Low	Low	Low
Wu et al., 2023 <sup>42</sup>	Multimodal VR features	MCI vs. HC	Accuracy = 89.8%	Low	Low	Low	Unclear
Moussavi et al., 2022 <sup>43</sup>	VR navigation	AD vs. MCI	Sig. diff. ( $p < 0.001$ )	Low	Unclear	Low	Low
Park, 2022 <sup>44</sup>	SCT-VR	MCI vs. HC	AUC = 0.99; Sens = 94%; Spec = 96%	Unclear	Unclear	Low	Low
Costa et al., 2021 <sup>47</sup>	SOIVET Maze & Route	MCI vs. HC	Maze AUC = 0.73; route AUC = 0.70	High	Low	Low	High
Jang et al., 2023 <sup>46</sup>	VARABOM	MCI vs. HC	AUC = 0.77; Sens = 83%; Spec = 72%	Unclear	Low	High	Unclear
Bayalhya et al., 2021 <sup>20</sup>	Smart health VR	Dementia vs. HC	Accuracy = 97.2%	Unclear	Unclear	Unclear	Unclear
Campo-Prieto et al., 2023 <sup>45</sup>	VR reaction time	Fallers vs. non-fallers	AUC = 0.74; Sens = 70%; Spec = 75%	Low	Low	Low	Low
Tsai et al., 2021 <sup>41</sup>	VR supermarket	MCI/AD vs. HC	Accuracy = 100%	High	Low	Unclear	Unclear
Serino et al., 2015 <sup>40</sup>	VR spatial memory	aMCI vs. HC	Sig. diff. ( $p < 0.01$ )	Low	Unclear	Low	Low

Abbreviations: AD, Alzheimer's disease; aMCI, amnesic mild cognitive impairment; AUC, area under the curve; HC, healthy control; MCI, mild cognitive impairment; QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies, version 2; Risk of Bias ratings: high, low, unclear; SCT-VR, social cognition training with virtual reality; Sens, sensitivity; Sig. diff., significant difference; SOIVET, Spatial Orientation in an Immersive Virtual Environment Test; Spec, specificity; VARABOM, virtual reality-based mild cognitive impairment monitoring; VR, virtual reality.

biomarkers (e.g., tau), improving sensitivity to AD pathology. The object location memory task demonstrated superior discriminative accuracy (AUC = 0.89) compared to conventional tools, while the association of object context was significantly associated with the volume of the lateral entorhinal cortex, highlighting the utility of the task in the detection of preclinical AD.

Park<sup>44</sup> introduced a VR-based spatial memory reconstruction task (social cognition training with VR [SCT-VR]) requiring participants to encode and reposition multiple objects within a virtual scene. Developed in Unity and executed on an HMD platform with controller-based input, the task captured spatial reconstruction accuracy and time metrics. The system yielded exceptionally high classification accuracy between MCI and healthy older adults (sensitivity = 0.944; specificity = 0.964), outperforming standard instruments such as the MoCA and supporting the diagnostic utility of spatial memory tasks in early cognitive screening. However, the QUADAS-2 assessment (Table 3) indicates an unclear risk of bias for Park due to insufficient details on patient selection.

Wu et al.<sup>42</sup> designed a multimodal VR system that integrated voice-based tasks with electroencephalogram (EEG) acquisition to assess executive and language function in MCI. Participants engaged in iterative question-and-answer and object-matching tasks within a virtual environment, while EEG and behavioral data (e.g., response latency, speech accuracy) were simultaneously recorded. Multiple machine learning (ML) models, including support vector machine (SVM), Random Forest, and XGBoost, were evaluated using leave-one-out cross-validation (LOOCV) to maximize robustness with a limited sample size. The SVM achieved the highest accuracy (AUC = 87%), and gamma-band power positively correlated with cognitive load, demonstrating the potential of VR-EEG integration for neurocognitive profiling.

### 3.5 | Spatial orientation and navigation tasks

Serino et al.<sup>40</sup> implemented a spatial orientation conversion task using immersive VR to investigate allocentric-to-egocentric transformation impairments in AD and aMCI. Conducted in a cave automatic virtual environment simulating 3D indoor spaces, the task used passive navigation and measured spatial conversion accuracy and reaction time. AD and aMCI participants showed significantly reduced accuracy in allocentric-to-egocentric transformations compared to controls, supporting the "mental frame syncing" hypothesis and confirming the task's sensitivity to early spatial processing deficits.

Da Costa et al.<sup>48</sup> developed the Spatial Orientation in an Immersive Virtual Environment Test (SOIVET) system, comprising two immersive tasks—maze navigation and route reproduction—to evaluate spatial learning impairments in MCI. The participants navigated autonomously using Oculus Rift headsets. Performance on both tasks correlated significantly with conventional spatial assessments (e.g., Mental Rotation Test) and demonstrated moderate discriminative capacity, underscoring the ecological value of VR-based spatial testing.

Moussavi et al.<sup>43</sup> created a spatial working memory task using HTC Vive in which participants navigated a complex virtual building and

returned to a starting point. The outcome measures included the accuracy and time of the reconstruction of the path, reflecting spatial updating and the navigational strategy. Participants in MCI exhibited greater deviation and recall delay in the path than controls, with results significantly correlated with MoCA scores, validating the clinical utility of the task for early detection of spatial impairment.

3.6 | Simulated activities of daily living and executive function tasks

Tsai et al.<sup>41</sup> developed an immersive VR-based supermarket task to assess spatial memory and multitasking in MCI. Implemented on Oculus Quest 2, the system involved navigation, item selection, and pricing, with controller input and voice feedback. A multimodal feature set—including task duration, error rates, and path efficiency—was modelled using SVM and XGBoost classifiers, achieving 100% accuracy in distinguishing MCI from controls. However, this perfect accuracy should be interpreted with caution, given the small sample size ( $n = 12$ ) and high risk of bias in patient selection (Table 3).

Campo-Prieto et al.<sup>45</sup> introduced a VR-based reaction wall task targeting dynamic attention and motor responsiveness in PD. Activated on Oculus Quest 2 with motion-tracked input, the task recorded reaction time and strike accuracy in response to visual stimuli. Performance was moderately negatively correlated with MMSE scores ( $\rho = -0.576$ ,  $p = 0.002$ ) and positively with Timed Up and Go (TUG) scores, indicating both cognitive and motor predictive value.

Jang et al.<sup>46</sup> created a functional VR scenario simulating home-care routines to evaluate daily task execution in older adults. Tasks such as medication retrieval and caregiver calls were completed within HTC Vive environments, with real-world object affordances and path tracing. MCI participants showed lower task adherence and operational sequencing compared to healthy controls. The results show that the sensitivity of simulated ADL-based tasks to multidomain cognitive deficits.

Bayahya et al.<sup>20</sup> designed the “Smart Supermarket” immersive VR system, integrating route guidance, object selection, and verbal interaction for dementia risk screening. Built in Unity, the system collected behavioral indicators such as navigation efficiency and repetitive errors. Using a Random Forest classifier, the model yielded high agreement with Mini-Cog results ( $Kappa = 0.93$ ), highlighting the task’s dual advantage of ecological validity and diagnostic sensitivity.

3.7 | Comparison with clinical tools and artificial intelligence integration

Across studies, 7 of 10 compared VR task performance with standard clinical tools such as the MMSE, MoCA, or Mini-Cog, covering key cognitive domains such as memory, spatial ability, and executive function (see Table 4 for details). For example, Park<sup>44</sup> and Tsai et al.<sup>41</sup> demonstrated strong alignment with MoCA and MMSE, while Castegnaro et al.<sup>39</sup> achieved an AUC of 0.89 for MoCA-based differentiation. Da

TABLE 4 Comparison of VR tasks to standard clinical tools.

Author, Year	VR task	Compared to clinical tool	Tool(s)
Park, 2022 <sup>44</sup>	Spatial memory	Yes	MoCA-K
Tsai et al., 2021 <sup>41</sup>	Virtual supermarket + ML	Yes	MoCA, MMSE
Bayahya et al., 2021 <sup>20</sup>	VR navigation & memory	Yes	Mini-Cog
Campo-Prieto et al., 2023 <sup>45</sup>	Reaction time	Yes	MMSE
Serino et al., 2015 <sup>40</sup>	Allocentric-egocentric spatial transformation	Partially	MMSE (group classification only)
Da Costa et al., 2022 <sup>49</sup>	SOIVET Maze & Route	Yes	ACE-R, BJLO, Tower of London
Castegnaro et al., 2022 <sup>39</sup>	Object-location memory	Yes	MoCA
Jang et al., 2023 <sup>46</sup>	Grandchild care scenario	Partially	MoCA (AUC only)
Wu et al., 2023 <sup>42</sup>	VR speech task + EEG	Yes	MoCA, MMSE
Moussavi et al., 2022 <sup>43</sup>	Target localization in virtual building	No	None

Abbreviations: ACE-R, Addenbrooke's Cognitive Examination Revised; AUC, area under the curve; BJLO, Benton Judgment of Line Orientation test; EEG, electroencephalogram; Mini-Cog, brief cognitive screening tool combining memory and clock drawing; ML, machine learning; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; MoCA-K, Montreal Cognitive Assessment Korean version; SOIVET, Spatial Orientation in an Immersive Virtual Environment Test; Tower of London, neuropsychological test of planning ability; VR, virtual reality.

Costa et al.<sup>49</sup> incorporated multiple tools (Addenbrooke's Cognitive Examination Revised, Benton Judgment of Line Orientation, Tower of London) to broaden domain coverage. Serino et al.<sup>40</sup> and Jang et al.<sup>46</sup> performed only partial comparisons, and Moussavi et al.<sup>43</sup> did not apply clinical benchmarks. VR metrics generally showed moderate-to-strong correlations with these established instruments, while offering superior sensitivity for detecting subtle functional deficits (see Table 3).

Notably, three studies integrated artificial intelligence (AI)-based classifiers to enhance diagnostic performance. Tsai and Wu used SVM and XGBoost for behavior-EEG fusion, achieving 100% and 87% classification accuracies, respectively. Bayahya et al. applied a Random Forest model that matched Mini-Cog diagnoses ( $\text{Kappa} = 0.93$ ).<sup>20,41,42</sup> However, none of these studies reported external validation on independent datasets, relying instead on internal cross-validation (e.g., LOOCV in Wu et al.,<sup>42</sup> 10-fold CV in Tsai et al.<sup>41</sup>).

### 3.8 | Synthesis of evidence

In summary, a comprehensive analysis of existing evidence indicates that FIVR technology holds considerable potential for screening MCI and dementia, although research findings exhibit significant heterogeneity. Current trends suggest that VR tasks targeting key cognitive domains, such as spatial navigation and object-location memory, generally demonstrate superior diagnostic performance compared to traditional paper-and-pencil tests, with multiple studies reporting AUC values consistently  $> 0.85$ . Under conditions of appropriate task duration and manageable interactive burden, most studies report high task completion rates and low incidence of virtual motion sickness, broadly supporting the technology's feasibility and acceptability. Furthermore, some studies used case-control sampling or failed to detail blinding procedures and measurement protocols, resulting in unclear risk of bias in QUADAS-2 assessments. Substantial evidence gaps persist, particularly concerning the standardization of VR tasks and metrics, reporting of assessment duration, longitudinal retention rates, and the evaluation of real-world applicability. These cross-study patterns highlight both the potential of immersive VR technology in early cognitive screening and the fragmented developmental stage currently prevailing in this field. The risk of reporting bias (e.g., publication bias) could not be statistically assessed but remains a potential limitation, as studies with negative findings may be underrepresented.

## 4 | DISCUSSION

This systematic review synthesized current evidence regarding the application of FIVR in screening and early diagnosis of NDs, focusing primarily on AD, MCI, and PD. The review specifically evaluated the diagnostic effectiveness and clinical feasibility of immersive VR systems, aiming to clarify their role in modern cognitive diagnostics and potential implications for clinical practice.

### 4.1 | Diagnostic effectiveness of VR-based cognitive assessments

The findings of this review suggest that immersive VR technologies offer considerable promise for the early and accurate detection of cognitive impairments associated with NDs. Across the included studies, VR-based cognitive tasks demonstrated diagnostic accuracy that was not only comparable to, but in several cases surpassed, that of established neuropsychological tools such as the MMSE and the MoCA. However, rather than expecting FIVR to replace these traditional tools, our findings show that they can serve a complementary role. Although standard neuropsychological tests have excelled in quantifying specific cognitive performance, such as memory recall, VR tasks excel at capturing complex functional-behavioral data, such as navigational<sup>22,40,43,47</sup> and multitasking performance.<sup>20,41,45,46</sup> This ecological validity has profound implications for clinical decision making: compared to static paper-and-pencil test scores, patients' performance in naturalistic simulation settings better predicts their capacity for independent living. This facilitates an earlier rehabilitation plan or the implementation of safety assessments.<sup>48</sup>

A key trend identified is the integration of ML to enhance diagnostic precision. For instance, Tsai et al.<sup>41</sup> achieved 100% classification accuracy using SVM and XGBoost classifiers. Although promising, such results must be interpreted with significant caution. The application of complex ML algorithms to small datasets ( $n = 12$  in Tsai et al.) can lead to a high risk of overfitting, in which models memorize noise rather than learning generalizable patterns.<sup>50</sup> Furthermore, the lack of external validation in these studies limits their clinical translational value. Without validation of the model in independent cohorts, the reported "perfect" accuracy is likely to overestimate actual clinical diagnostic utility. Establishing transparent reporting standards for ML models such as the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD)<sup>51</sup> is crucial, as it will propel the field from proof-of-concept to clinical practice.

Another feature of VR technology is its ability to replicate real-world contexts while maintaining experimental control and allowing for repeated, low-cost testing under consistent conditions. This addresses the long-standing ecological validity limitations of traditional cognitive assessments. By embedding cognitive tasks within naturalistic environments, VR assessments can yield more detailed and representative behavioral data. Studies such as Kourtesis et al.<sup>52</sup> and Romero-Ayuso et al.<sup>53</sup> have demonstrated the alignment between behavioral responses in VR environments and those observed in real-world contexts. Others, such as Parsons<sup>54</sup> have highlighted the objectivity and standardization of behavior-based VR data capture.

Tasks embedded in VR environments, such as those conducted in a virtual supermarket<sup>22,41</sup> or a simulated home caregiving scenario<sup>46</sup> have shown particular utility in identifying early cognitive decline. This ecological validity not only enhances diagnostic precision but may also improve participant engagement and reduce evaluation-related stress, thereby contributing to the overall reliability of the screening process. Similar benefits have been reported in studies examining other

clinical populations in which ecological demands are integral to real-life functioning.<sup>52</sup> Taken together, these findings underscore the promise of VR as a practical and innovative approach to the early detection of cognitive impairment related to neurodegeneration.

## 4.2 | Feasibility, usability, and clinical implementation

The feasibility-related outcomes reported across the studies consistently underscored the practicality and acceptability of VR-based cognitive assessments. High task completion rates (> 90%) in several studies<sup>39,41</sup> suggested VR's ease of use and positive patient compliance. Usability and engagement were generally high, supported by intuitive interfaces and naturalistic task scenarios, facilitating immersive and meaningful interaction. The absence of significant adverse events in most studies further highlights VR's suitability for routine clinical deployment. However, mild adverse effects such as fatigue and cybersickness were reported in one study,<sup>45</sup> emphasizing the need for careful task design, especially for populations with potential physical or sensory limitations.

The adoption of commercially available immersive devices (e.g., Oculus Quest, HTC Vive) represents another critical strength. These devices, being cost effective, easily deployable, and user friendly, significantly reduce barriers to clinical implementation. Additionally, their compatibility with advanced analytic techniques, such as ML, further broadens their applicability and enhances diagnostic precision through the analysis of complex behavioral patterns.

Nevertheless, obstacles to routine clinical application remain. First, digital literacy remains a significant moderating factor; while existing research shows high engagement, the frequent recruitment of enthusiastic volunteers may mask anxiety or resistance within a broader older population.<sup>55</sup> Second, operating costs extend beyond hardware procurement, including software licensing, troubleshooting, and the need for dedicated physical space in clinics.<sup>56</sup> Third, training clinicians is a critical bottleneck. Although VR can standardize data collection, the interpretation of these data and the integration into concrete clinical guidelines still rely heavily on expert clinical judgement<sup>57</sup>. Therefore, successful implementation necessitates not only user-friendly software but also specialized data analysis capabilities.

## 4.3 | Limitations of current evidence and recommendations for future research

Although the aggregated data support FIVR as a potential cognitive assessment, the evidence was constrained by limitations identified in the QUADAS-2 assessment. A pervasive problem is the high risk of bias in patient selection, primarily stemming from the use of case-control study designs. By recruiting patients with a "clear diagnosis" of dementia and "healthy" controls, studies artificially amplify the diagnostic contrast, which may lead to spectrum bias. This may explain the near-perfect AUC values (0.99–100)<sup>41,44</sup> reported in some stud-

ies, which can significantly decrease when applied to the "complex" heterogeneous populations commonly found in primary care.

Furthermore, the lack of clear blinding in several studies poses a risk of operational bias; if the operator is aware of the participant's diagnosis, they might subtly guide the participant during the VR task. Small sample sizes further weaken the statistical power and the reliability of accuracy estimates.

Methodologically, few studies incorporated multimodal physiological data (e.g., EEG and eye-tracking), which could enhance diagnostic sensitivity. Participant-level factors—such as age, education, and digital literacy—were also rarely considered, despite their potential influence on usability and performance, particularly in older adults. These methodological flaws suggest that while the potential of VR is high, the current evidence base is likely overly optimistic.

Although many studies aimed to enhance ecological validity through realistic virtual scenarios and comparisons with standardized cognitive tools, none directly compared VR performance to real-world behavior. Similarly, no study systematically examined the effect of immersion level (e.g., HMDs vs. non-immersive platforms) on cognitive outcomes. Furthermore, most studies relied on cognitive screening tools (e.g., MoCA) as the reference standard, rather than robust biological markers such as positron emission tomography, magnetic resonance imaging, or cerebrospinal fluid biomarkers, limiting the ability to validate VR outcomes against the underlying pathology.

## 4.4 | Recommendations for future research

Future research should prioritize large-scale, multicenter studies using standardized VR protocols to enable meaningful cross-study comparisons and support robust evidence synthesis. Longitudinal designs are also critically needed to evaluate the predictive validity of VR-based cognitive assessments in tracking clinical progression and forecasting disease trajectories over time.

In detail, to advance the field from exploratory to clinical adoption, future research should address the following directions:

- **Standardization of protocols:** There is an urgent need for a standardized "core outcome set" for VR cognitive trials, defining essential reporting metrics (e.g., hardware specs, locomotion methods, cybersickness scores) to facilitate meta-analysis.
- **Robust study designs:** Future studies should move beyond case-control designs to consecutive cohort studies in primary care settings to determine the true predictive value of VR tools in undifferentiated populations.
- **External and cross-cultural validation:** As VR relies heavily on visual cues, tools developed in one culture (e.g., a Western supermarket) may not translate to others. Cross-cultural validation studies are necessary to ensure global applicability.
- **Multimodal integration:** Integrating VR with biomarkers (e.g., EEG, eye-tracking) could uncover latent neurocognitive signatures, thereby improving sensitivity to preclinical stages (e.g., MCI) in which behavioral deficits are subtle.



## 5 | SUMMARY AND CONCLUSIONS

In conclusion, this systematic review supports the effectiveness and clinical feasibility of immersive VR-based cognitive assessments for the early detection of neurodegenerative disorders. By addressing limitations inherent in conventional cognitive testing approaches, immersive VR offers transformative potential in neuropsychological assessment, promising earlier diagnosis, targeted intervention, and improved patient outcomes. As technology and digital health-care continue to evolve, VR is poised to become an integral component of cognitive diagnostic pathways, substantially shaping the future landscape of neuropsychological evaluation and patient management.

## ACKNOWLEDGMENTS

The authors thank the University of Kent for providing academic and administrative support. The China Scholarship Council funded this project. No. 202208150024.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## REFERENCES

- Choonara YE, Pillay V, Du Toit LC, et al. Trends in the molecular pathogenesis and clinical therapeutics of common neurodegenerative disorders. *Int J Mol Sci*. 2009;10(6):2510-2557.
- Gadhav DG, Sugandhi VV, Jha SK, et al. Neurodegenerative disorders: mechanisms of degeneration and therapeutic approaches with their clinical relevance. *Ageing Res Rev*. Published online 2024:102357.
- Ahmed N, Ahmed U, Thornalley PJ, Hager K, Fleischer G, Münch G. Protein glycation, oxidation and nitration adduct residues and free adducts of cerebrospinal fluid in Alzheimer's disease and link to cognitive impairment. *J Neurochem*. 2005;92(2):255-263.
- Aranda MP, Kremer IN, Hinton L, et al. Impact of dementia: health disparities, population trends, care interventions, and economic costs. *J Am Geriatr Soc*. 2021;69(7):1774-1783.
- Dementia WHO. *World Health Organ Retrived* <https://www.who.int/news-room/fact-sheets/detail/dementia>. Published online 2025.
- Wittenberg R, Hu B, Jagger C, et al. Projections of care for older people with dementia in England: 2015 to 2040. *Age Ageing*. 2020;49(2):264-269.
- Cullen B, O'Neill B, Evans JJ, Coen RF, Lawlor BA. A review of screening tests for cognitive impairment. *J Neurol Neurosurg Psychiatry*. 2007;78(8):790-799.
- Chua SIL, Tan NC, Wong WT, et al. Virtual reality for screening of cognitive function in older persons: comparative study. *J Med Internet Res*. 2019;21(8):e14821. doi:10.2196/14821
- Maruff P, Lim YY, Darby D, et al. Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer's disease. *BMC Psychol*. 2013;1:1-11.
- Rao SM, Galioto R, Sokolowski M, et al. Cleveland clinic cognitive battery (c3b): normative, reliability, and validation studies of a self-administered computerized tool for screening cognitive dysfunction in primary care. *J Alzheimer's Dis*. 2023;92(3):1051-1066.
- Björngrim S, van den Hurk W, Betancort M, Machado A, Lindau M. Comparing traditional and digitized cognitive tests used in standard clinical evaluation—a study of the digital application minnemera. *Front Psychol*. 2019;10:2327.
- Libon DJ, Swenson R, Price CC, et al. Digital assessment of cognition in neurodegenerative disease: a data driven approach leveraging artificial intelligence. *Front Psychol*. 2024;15:1415629.
- Bottari C, Dassa C, Rainville C, Dutil É. The criterion-related validity of the IADL Profile with measures of executive functions, indices of trauma severity and sociodemographic characteristics. *Brain Inj*. 2009;23(4):322-335. doi:10.1080/02699050902788436
- Burgess PW, Alderman N, Forbes C, et al. The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *J Int Neuropsychol Soc*. 2006;12(2):194-209. doi:10.1017/S1355617706060310
- Rand D, Rukan SBA, (Tamar) Weiss PL, Katz N. Validation of the Virtual MET as an assessment tool for executive functions. *Neuropsychol Rehabil*. 2009;19(4):583-602. doi:10.1080/09602010802469074
- Bamodu O, Ye X. *Virtual Reality and Virtual Reality System Components*. Published online 2013.
- Martín-Gutiérrez J, Mora CE, Añorbe-Díaz B, González-Marrero A. Virtual technologies trends in education. *EURASIA J Math Sci Technol Educ*. 2017;13(2). doi:10.12973/eurasia.2017.00626a
- Parsey CM, Schmitter-Edgecombe M. Applications of technology in neuropsychological assessment. *Clin Neuropsychol*. 2013;27(8):1328-1361. doi:10.1080/13854046.2013.834971
- Harvey PD, Depp CA, Rizzo AA, et al. Technology and mental health: state of the art for assessment and treatment. *Am J Psychiatry*. 2022;179(12):897-914. doi:10.1176/appi.ajp.21121254
- Bayahya AY, Alhalabi W, AlAmri SH. Smart health system to detect dementia disorders using virtual reality. *Healthcare*. 2021;9(7):810. doi:10.3390/healthcare9070810
- Pang JWM, Lim JE, Koh EYL, Tan NC. Validity and reliability of a virtual reality system as an assessment tool for cognitive impairment based on the six cognitive domains. *Virtual Real*. 2025;29(3):109. doi:10.1007/s10055-025-01187-0
- Porffy LA, Mehta MA, Mouchlianitis E, Shergill SS. VStore: feasibility and acceptability of a novel virtual reality functional cognition task. *Front Virtual Real*. 2022;3:875197.
- Wagner S, Joeres F, Gabele M, Hansen C, Preim B, Saalfeld P. Difficulty factors for VR cognitive rehabilitation training—Crossing a virtual road. *Comput Graph*. 2019;83:11-22.
- Tan NC, Lim JE, Sultana R, Quah JHM, Wong WT. A virtual reality cognitive screening tool based on the six cognitive domains. *Alzheimer's Dement Diagn Assess Dis Monit*. 2024;16(4):e70030.
- Howett D, Castegnaro A, Krzywicka K, et al. Differentiation of mild cognitive impairment using an entorhinal cortex-based test of virtual reality navigation. *Brain*. 2019;142(6):1751-1766. doi:10.1093/brain/awz116
- Park JS, Jung YJ, Lee G. Virtual reality-based cognitive-motor rehabilitation in older adults with mild cognitive impairment: a randomized controlled study on motivation and cognitive function. In: *Healthcare*. 8. MDPI; 2020:335.
- Clay V, König P, Koenig S. Eye tracking in virtual reality. *J Eye Mov Res*. 2019;12(1):10-16910.
- Miskowiak KW, Jespersen AE, Kessing LV, et al. Cognition assessment in virtual reality: validity and feasibility of a novel virtual reality test for real-life cognitive functions in mood disorders and psychosis spectrum disorders. *J Psychiatr Res*. 2022;145:182-189.
- Sayma M, Tuijt R, Cooper C, Walters K. Are we there yet? Immersive virtual reality to improve cognitive function in dementia and mild cognitive impairment. *The Gerontologist*. 2020;60(7):e502-e512.
- Saredakis D, Szpak A, Birkhead B, Keage HA, Rizzo A, Loetscher T. Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. *Front Hum Neurosci*. 2020;14:96.
- Kim O, Pang Y, Kim JH. The effectiveness of virtual reality for people with mild cognitive impairment or dementia: a meta-analysis. *BMC Psychiatry*. 2019;19:1-10.
- Unity Technologies. Unity. Published online June 2005. <https://unity.com>

33. Zyda M. From visual simulation to virtual reality to games. *Computer*. 2005;38(9):25-32. doi:10.1109/MC.2005.297
34. Organization WH. *World Report on Ageing and Health*. World Health Organization; 2015.
35. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210.
36. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol*. 1990;43(6):543-549.
37. Bowen DJ, Kreuter M, Spring B, et al. How we design feasibility studies. *Am J Prev Med*. 2009;36(5):452-457.
38. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.
39. Castegnaro A, Howett D, Li A, et al. Assessing mild cognitive impairment using object-location memory in immersive virtual environments. *Hippocampus*. 2022;32(9):660-678.
40. Serino S, Morganti F, Di Stefano F, Riva G. Detecting early egocentric and allocentric impairments deficits in Alzheimer's disease: an experimental study with virtual reality. *Front Aging Neurosci*. 2015;7:88.
41. Tsai CF, Chen CC, Wu EHK, et al. A machine-learning-based assessment method for early-stage neurocognitive impairment by an immersive virtual supermarket. *IEEE Trans Neural Syst Rehabil Eng*. 2021;29:2124-2132. doi:10.1109/TNSRE.2021.3118918
42. Wu R, Li A, Xue C, et al. Screening for mild cognitive impairment with speech interaction based on virtual reality and wearable devices. *Brain Sci*. 2023;13(8):1222.
43. Moussavi Z, Kimura K, Lithgow B. Egocentric spatial orientation differences between Alzheimer's disease at early stages and mild cognitive impairment: a diagnostic aid. *Med Biol Eng Comput*. Published online 2022:1-9.
44. Park JH. Can the virtual reality-based spatial memory test better discriminate mild cognitive impairment than neuropsychological assessment? *Int J Environ Res Public Health*. 2022;19(16):9950.
45. Campo-Prieto P, Cancela-Carral JM, Rodríguez-Fuentes G. Immersive virtual reality reaction time test and relationship with the risk of falling in Parkinson's disease. *Sensors*. 2023;23(9):4529.
46. Jang S, Choi SW, Son SJ, et al. Virtual reality-based monitoring test for MCI: a multicenter feasibility study. *Front Psychiatry*. 2023;13:1057513.
47. Mannan FA, Porffy LA, Joyce DW, Shergill SS, Celiktutan O. Automatic detection of cognitive impairment with virtual reality. *Sensors*. 2023;23(2):1026.
48. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PloS One*. 2019;14(11):e0224365.
49. Da Costa RQM, Pompeu JE, Moretto E, et al. Two immersive virtual reality tasks for the assessment of spatial orientation in older adults with and without cognitive impairment: concurrent validity, group comparison, and accuracy results. *J Int Neuropsychol Soc*. 2022;28(5):460-472.
50. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *J Br Surg*. 2015;102(3):148-158.
51. Kourtesis P, Collina S, Doumas LA, MacPherson SE. Validation of the virtual reality everyday assessment lab (VR-EAL): an immersive virtual reality neuropsychological battery with enhanced ecological validity. *J Int Neuropsychol Soc*. 2021;27(2):181-196.
52. Romero-Ayuso D, del Pino-González A, Torres-Jiménez A, et al. Enhancing ecological validity: virtual reality assessment of executive functioning in children and adolescents with ADHD. *Children*. 2024;11(8):986.
53. Parsons TD. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front Hum Neurosci*. 2015;9:660.
54. Huygelier H, Schraepen B, Van Ee R, Vanden Abeele V, Gillebert CR. Acceptance of immersive head-mounted virtual reality in older adults. *Sci Rep*. 2019;9(1):4519.
55. Garrett B, Taverner T, Gromala D, Tao G, Cordingley E, Sun C. Virtual reality clinical research: promises and challenges. *JMIR Serious Games*. 2018;6(4):e10839.
56. Pieri L, Tosi G, Romano D. Virtual reality technology in neuropsychological testing: a systematic review. *J Neuropsychol*. 2023;17(2):382-399.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Liu Z, Soria D, Lai DJ, Zhang J, Shergill S, Ang CS. The use of fully immersive virtual reality for screening neurodegenerative diseases: A systematic review of behavioral and diagnostic outcomes. *Alzheimer's Dement*. 2026;18:e70244. <https://doi.org/10.1002/dad2.70244>