npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-024-01206-5

A portable and efficient dementia screening tool using eye tracking machine learning and virtual reality

Check for updates

Ying Xu[1,4], Chi Zhang[2,4], Baobao Pan[2], Qing Yuan[1] & Xu Zhang[3]

Dementia represents a significant global health challenge, with early screening during the preclinical stage being crucial for effective management. Traditional diagnostic biomarkers for Alzheimer's Disease, the most common form of dementia, are limited by cost and invasiveness. Mild cognitive impairment (MCI), a precursor to dementia, is currently identified through neuropsychological tests like the Montreal Cognitive Assessment (MoCA), which are not suitable for large-scale screening. Eye-tracking technology, capturing and quantifying eye movements related to cognitive behavior, has emerged as a promising tool for cognitive assessment. Subtle changes in eye movements could serve as early indicators of MCI. However, the interpretation of eye-tracking data is challenging. This study introduced a dementia screening tool, VR Eye-tracking Cognitive Assessment (VECA), using eye-tracking technology, machine learning, and virtual reality (VR) to offer a non-invasive, efficient alternative capable of large-scale deployment. VECA was conducted with 201 participants from Shenzhen Baoan Chronic Hospital, utilizing eye-tracking data captured via VR headsets to predict MoCA scores and classify cognitive impairment across different educational backgrounds. The support vector regression model employed demonstrated a high correlation (0.9) with MoCA scores, significantly outperforming baseline models. Furthermore, it established optimal cut-off scores for identifying cognitive impairment with notable sensitivity (88.5%) and specificity (83%). This study underscores VECA's potential as a portable, efficient tool for early dementia screening, highlighting the benefits of integrating eye-tracking technology, machine learning, and VR in cognitive health assessments.

Dementia refers to a spectrum of neurodegenerative syndromes characterized by progressive disturbances in various cognitive functions severe enough to interfere with the patient's activities of daily living. The rapid increase in the number of patients with dementia has become a global health challenge, with Alzheimer's disease (AD) being the most common, accounting for 60–70%[1]. Mild cognitive impairment (MCI) refers to the progressive decline of memory or other cognitive functions but does not affect the ability of daily life as dementia[2]. The increasing prevalence of cognitive impairment has caused enormous economic loss[3,4]. Accumulating evidence shows that early diagnosis and timely intervention can delay cognitive decline[5–7]. so early screening in the preclinical stage of dementia is necessary.

Currently, cerebrospinal fluid β-amyloid (amyloid β, Aβ) and tau, amyloid positron emission tomography (PET) and AD pathogenic gene carrying are diagnostic markers for AD[8]. Though accurate, these approaches could hardly perform as an early screening tool for high costs and surgical invasiveness. Dementia is essentially cognitive dysfunction, so assessment of cognition is an important part of the diagnosis. Neuropsychological paper-based instruments, such as the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA) are commonly used, which cover orientation, memory, attention, calculation, language ability, visuospatial ability, etc. Though valid and reliable, traditional neuropsychological tests must be administered by trained physicians, and are neither simple nor efficient enough to serve as large-scale dementia screening tools. Moreover, neuropsychological assessment results are affected by both the physician's subjective judgments and the testing environment. Table 1 presents a comparison of various widely used

[1]Shenzhen Bao'an Centre for Chronic Disease Control, Shenzhen, PR China. [2]Shenzhen Yiwei Technology, Shenzhen, PR China. [3]National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen, PR China. [4]These authors contributed equally: Ying Xu, Chi Zhang
e-mail: 554208737@qq.com; zhangxu880605@gmail.com

Table 1 | Comparison of common diagnosis or screening approaches

Type	Neuropsychological tests	Imaging	Laboratory
Content	MMSE, MOCA, etc.	Brain scans	Cerebrospinal fluid examination
Method	Paper-based	CT/MRI/ PET-CT	Laboratory examinations
Pros and cons	Long duration (30 min); Non-invasive; Assisted by others; Subjective	Radiation exposure; Objective; Uncapable for early screening; High cost	Invasive; Objective; High cost

diagnostic or screening methods. At present, there is no efficient solution for large-scale dementia screening.

Eye-tracking, virtual reality (VR), and artificial intelligence (AI) technologies have shown great potential in the diagnosis of cognitive disorders or AD, aiding in the improvement of early screening, diagnosis, and treatment. Eye tracking is a real-time sensor technology that measures gaze points and eye movements, providing information on cognitive function and attention; VR technology allows for cognitive testing in controlled scenarios, accurately assessing cognitive abilities; AI technology can precisely process vast amounts of clinical data, identify subtle changes, and assist in the formulation of diagnostic and treatment plans. Cognitive-related uses of eye-tracking technology have been exploded in recent years, enabling online cognitive activity to be recorded, physicians to connect learning outcomes to the cognitive processes of subjects[9], and detection of emotional and cognitive states[10]. Numerous studies have shown that analyzing eye movements with algorithms from eye trackers enables quantitative cognitive assessment. Visual Paired Comparison (VPC), which uses non-invasive eye tracking to identify how subjects view novel versus repeated visual stimuli. MCI patients have lower novelty preference scores on VPC compared to healthy controls[11,12]. Passive assessment of visuospatial memory using eye tracking, where participants are shown a set of pictures and then presented with modified versions of these pictures. Compared to MCI and dementia patients, healthy controls spend significantly more time looking at the changed parts of the pictures[13]. Visual search tasks test the ability to find a previously seen target stimulus among distractors. Eye movement parameters can identify visual search impairments in MCI patients and early identify individuals at high risk for AD[14]. Visual tracking tasks test stimuli as moving objects, where subtle changes in visual attention can be used for quick MCI detection[15]. Anti-saccade tasks test, where the amnestic MCI group shows more errors, more misses, and fewer corrections compared to control groups[16]. Short-term memory binding tests are used to assess the ability to integrate surface features (e.g., shape and color) as unique representations in memory[17]. Impaired short-term memory binding abilities are common in sporadic Alzheimer's disease[18], familial Alzheimer's disease, and asymptomatic carriers of familial Alzheimer's disease[19]. Patients with MCI exhibit impaired gaze ability and exploratory eye movement behavior, accompanied by poor short-term memory binding capabilities, which can be used to assess the risk of MCI transitioning to dementia[20]. Serial cognitive task tests aim to assess specific cognitive domains, including deductive reasoning, working memory, attention, and memory recall. The percentage of time spent gazing at the correct answer (target image) is used as a measure of cognitive scoring, correlating well with neuropsychological test scores and showing good diagnostic performance in detecting MCI and dementia patients[21]. No-instruction eye-tracking tests can detect dementia-related cognitive impairments as related eye movement biomarkers for patients who are easily confused, have language difficulties, or are in the later stages of the disease[22]. However, several issues need to be addressed: the lengthy duration required for some eye-tracking cognitive assessments is not efficient such as 30 min VPC[23]; most existing cognitive assessment paradigms assess a limited range of cognitive domains; the high cost and limited portability of traditional monitor and eye-tracker setups make large-scale application challenging.

Table 2 | Model performance comparison

Model	Median AE	MAE	RMSE	Corr
MLP	3.81 ± 0.82	4.13 ± 0.72	5.17 ± 0.84	0.59 ± 0.11
SVR	2.04 ± 0.18	2.84 ± 0.34	3.78 ± 0.57	0.90 ± 0.06
Lasso	2.38 ± 0.07	2.91 ± 0.24	3.76 ± 0.42	0.85 ± 0.05
GBRT	2.31 ± 0.28	2.93 ± 0.29	3.74 ± 0.42	0.77 ± 0.08
Baseline	10.61 ± 0.18	10.71 ± 0.1	11.59 ± 0.10	0.47 ± 0.06

SVR support vector regression, *MLP* multi-layer perceptron, *Lasso* Lasso regression, *GBRT* Gradient Boost Regression Tree, *Median AE* median absolute error, *MAE* mean absolute error, *RMSE* root mean square error, *Corr* correlation with MoCA score.

VR technology can create controllable virtual scenes for assessing cognitive function, distinguishing between patients with mild cognitive impairment and those at low or high risk for AD[24,25]. The combination of eye-tracking technology and VR, with the portability of VR head-mounted display equipment and VR immersive environment, is expected to be applied to large-scale cognitive impairment screening in multiple scenarios. However, two challenging issues have to be handled. First, these technologies should be integrated in a smart way so that efficiency is guaranteed. Moreover, subtle changes in eye movements can potentially serve as biomarkers for early MCI identification, interpreting the vast amount of eye-tracking data poses challenges.

Eye-tracking technology captures and quantifies eye movements related to cognitive behavior, forming an essential data stream for identifying subtle behavioral changes. Machine learning methods in AI are crucial for recognizing these changes. In VPC tasks, machine learning methods have improved the accuracy of AD detection[11]. In visual search tasks involving finding a previously seen target stimulus among distractors, machine learning methods can distinguish between control groups, MCI patients, and dementia subjects, effectively identifying MCI subjects with dementia-like eye movement characteristics[14]. Machine learning has been applied to tasks designed for single cognitive functions. Recent research has proposed a multimodal machine learning analysis framework, which is applied to the analysis of multimodal mental health digital biomarkers[26]. This study enhances our potential to explore the application of machine learning in the analysis of eye movement characteristics across multiple cognitive domains.

In this study, we integrated eye-tracking technology, machine learning, and VR to design and develop an efficient, portable, and quantitative early screening tool for dementia. By conducting a 5-min multi-domain cognitive task within a controlled VR environment, we collected relevant eye movement data and pre-trained classification models using supervised machine learning algorithms. This project aims to facilitate the early detection of cognitive impairment, slow the progression of dementia, and significantly reduce the psychological, caregiving, and economic burdens that dementia imposes on individuals and society.

Results

VR-AI model selection

The model performance was evaluated on the test set using the following metrics: Median Absolute Error (Median AE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Correlation with MoCA score. The model with the best performance was selected to predict paper-based MoCA score (hereinafter referred to as VR-AI model). Each model underwent 5-fold cross-validation and compared in the following metrics (Table 2). MAE and RMSE are common metrics which are widely used to describe regression performance. However, there was unpredictable noise in the target value due to human factors and lead to extreme bad cases in both train and test set. Instead, median AE could indicate model performance in a more stable way disregard of outliers. Moreover, absolute error shows more intuitive difference between predictions and labels. Therefore, MAE was weighed more as the metrics to compare models. Higher correlation coefficient between predictions and targets indicates better performance of screening. SVR overperformed other models and baseline under most of our metrics, especially under average median absolute error of 2.04 and correlation coefficient as high as 0.9 and was selected as VR-AI model (Fig. 1).

Model interpretability

Contributions of each cognitive task and demographic variable to the prediction were evaluated via SHapley Additive exPlanations (SHAP) analysis, sorted by mean absolute Kernel SHAP values which were computed by special weighted linear regression. Figure 2 illustrates normalized importance of all features for VR-AI model. Years of education was the strongest predictor (mean |SHAP| = 14.6% (s.d. = ±7.7%)), followed by calculation task 4 (mean |SHAP| = 11.5 (s.d. = ±8.5%)). Memory, execution and recall tasks also ranked high among other features.

Screening evaluation

In this study executed in Shenzhen, demographic characteristics of participants were analyzed (Table 3), with a particular focus on educational levels. Given the varying numbers of participants across different educational backgrounds, such as a smaller number of illiterate individuals, participants were systematically categorized into three more evenly distributed groups based on their education levels to facilitate a more effective screening evaluation. The grouping was as follows: Group 1 comprised individuals with 0–6 years of education, including those who are illiterate and those who attended primary school, totaling 61 participants; Group 2 included participants with 6–9 years of education, encompassing junior high school, with a total of 63 participants; and Group 3 consisted of individuals with an education level greater than 9 years, including those who completed high school and those holding a bachelor's degree or higher, totaling 78 participants.

Referencing the work of Lu et al., it was underscored that the MoCA norms were meticulously crafted to take into account key influential factors. This led to the establishment of distinct optimal cutoff points tailored to the

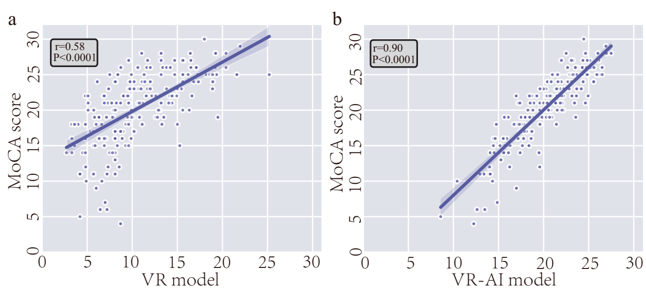


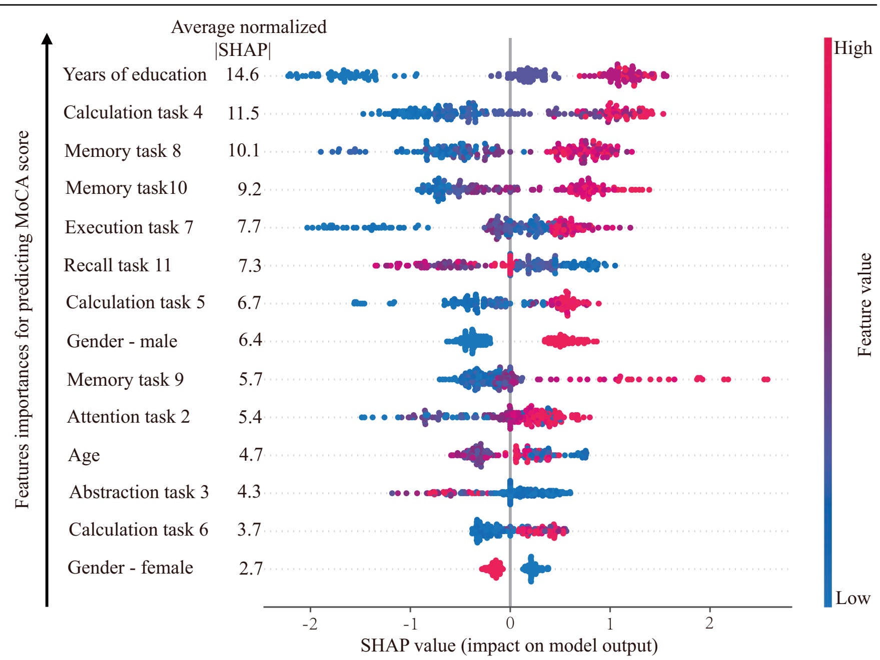
Fig. 1 | Correlation comparison of baseline VR model and VR-AI model. The coordinates of scatter points consist of model predicted scores (X-axis) and MoCA scores (Y-axis). Linear regression lines were also given to manifest linear relevance of models and MoCA score (**a** VR model vs. MoCA, $r = 0.58$, $p < 0.0001$; **b** VR-AI model vs. MoCA, $r = 0.90$, $p < 0.0001$).

Table 3 | Demographic characteristics of participants

Attribute	Count (*N*)	Pct.
Data source		
Fuan CRC	88	43.8%
Wanxiang CRC	113	56.2%
Age		
55–65	113	56.2%
65–75	79	39.3%
>75	9	4.5%
Gender		
Male	81	40.3%
Female	120	59.3%
Education level (years)		
Illiterate (0)	8	3.9%
Primary School (1–6)	53	26.4%
Junior High School (6–9)	62	30.8%
High School (9–12)	55	27.4%
Bachelor or above (>12)	23	11.4%

CRC Community Rehabilitation Center.

Fig. 2 | Contributions of cognitive tasks and demographic factors to the predictive model, ranked by their mean absolute SHAP values. The x-axis displays the contribution of each feature on the prediction. The y-axis lists the variables in order of decreasing importance. The color gradient from blue to red indicates the variables' relevance to the score, ascending from low to high. Each point represents an individual participant's SHAP value for that variable.



education levels of the individuals: 13/14 for those without any formal education, 19/20 for individuals with 1 to 6 years of education, and 24/25 for individuals educated beyond 7 years. These carefully determined cutoff points demonstrated a sensitivity rate of 83.8% for the detection of all forms of cognitive impairments[27].

Leveraging these distinct optimal cutoff points, which were directly correlated with the educational levels of individuals, the study's participants were efficiently categorized and labeled in alignment with Chinese educational standards as follows:

Group 1: 0–6 years of education: normal cognition (MoCA score 14–30), $n = 50$; cognitive impairment (MoCA score 0–13), $n = 11$.

Group 2: 6–9 years of education: normal cognition (MoCA score 20–30), $n = 44$; cognitive impairment (MoCA score 0–19), $n = 19$.

Group 3: education level greater than 9 years: normal cognition (MoCA score 25–30), $n = 47$; cognitive impairment (MoCA score 0–24), n = 31.

Classification performance of VR-AI model was evaluated due to ROC curves. Optimal thresholds of VR-AI score were given by F-β method which, in screening scenario, weighed more on sensitivity. The whole flowchart of the proposed model is shown in Fig. 3.

The screening performance was also evaluated in classification aspect. All subjects were grouped based on education levels and classified by their VR-AI scores and optimal thresholds to compute sensitivity and specificity. The AUCs of three groups were 0.88(95% CI: 0.78–0.97), 0.93(95% CI: 0.89–1.00), and 0.94(95% CI: 0.89–0.99) respectively (Fig. 4). The optimal screening cut-off scores for each group are given by F-β score: less than 6 years of education, 13/14; 6 to 9 years of education, 21/22; more than 9 years of education, 23/24. The VR-AI model showed excellent sensitivities of 89, 88, and 89% in screening for cognitive impairment and specificities of 82, 87, and 84%. The sensitivity of the overall screening of cognitive impairment and cognitively normal subjects can reach 88.5% with specificity of 83%.

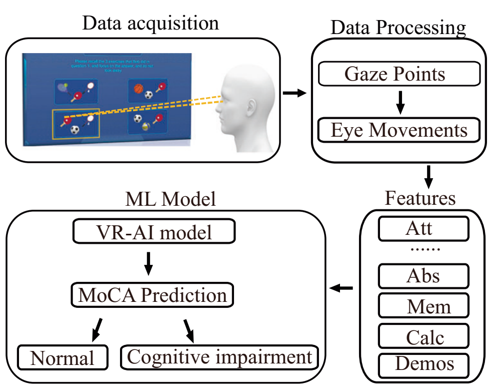


Fig. 3 | The flowchart of machine learning model used for MoCA prediction. The pipeline includes data acquisition, data processing, feature extraction (including memory (Mem), visual attention (Att), abstraction ability (Abs), and calculation (Calc), Demographics (Demos), machine learning (ML), and Classification.

Discussion

To retrospect the sample size of our research, the estimate of the sample size was given according to the general requirements of diagnostic research: (1) screening test sensitivity (Sen) ≥ 0.5, sensitivity tolerance error 0.10; (2) The specificity (Spe) of screening test was ≥0.5, the tolerance error of specificity was 0.10. (3) Significance test level two-sided $\alpha = 0.05$. The results are presented in Table 4. The sample size of the three groups (Group 1 $N = 61$, group 2 $N = 63$, group 3 = 78) in this study can meet the requirements of statistics.

The classification of participants into three groups based on their education levels aligns with the principles outlined by Lu et al. for adapting the MoCA for the Chinese population. The differentiation by education level is a critical aspect of this adaptation because educational attainment can significantly influence cognitive test performance. By creating distinct groups based on education levels, the VECA can more accurately reflect the cognitive abilities of individuals within those specific contexts, thereby enhancing the test's sensitivity and specificity for different segments of the population.

Eye movements were highly informative indicators of cognitive function. In our study, together with basic information—gender, age and education, encoded eye movements can excellently classify normal and cognitive impairment subjects within education-based groups via machine learning. As a result, we developed an efficient screening tool to identify subjects that would be labeled cognitive impairment by traditional paper-based instruments. The crucial part is regression of MoCA score. SVM regression outperformed and achieved absolute error of 2.04 and strong correlation with target variable. An important application of our screening tool is providing diagnosis suggestions. The AUCs of three education-based groups were 0.92(95% CI: 0.85–0.99), 0.95(95% CI: 0.89–1.00), and 0.94(95% CI: 0.89–0.99) respectively. Proposed optimal cut-off scores for each group achieve sensitivities of 81.8, 94.7, and 83.8% in screening for cognitive impairment and specificities of 98, 91.1, and 73.6%.

This study aims to employ a quantitative approach for large-scale early screening of dementia, specifically to aid physicians in detecting cognitive impairments across a broader spectrum. The use of VR devices, particularly when combined with eye-tracking technology, presents several advantages over traditional monitor and eye-tracker setups. Portability and applicability: This portability makes VR devices more suitable for large-scale early dementia screenings, allowing for early detection of cognitive impairments in a wider range of settings. Cost-effectiveness: As VR devices with eye-tracking capabilities become more common as consumer electronics, they present a cost-effective option. This affordability makes VR devices a viable choice for large-scale screening efforts. Interference Resistance: VR devices offer better resistance to interference compared to traditional setups involving monitors and eye trackers. Participants undergo assessments in a consistent virtual environment, minimizing the impact of external environmental variables and resulting in more objective and accurate test outcomes. Operational Convenience: In practical screening applications, VR

Fig. 4 | VR-AI model's ROC curves for three education-based groups. AUC values are 0.88 for less than 6 years of education (**a**), 0.93 for 6–9 years (**b**), and 0.94 for over 9 years (**c**), demonstrating the model's effectiveness across varying educational backgrounds. The green solid line represents the VR Screen Classifier, and the gray dashed line indicates Random Guess as a baseline.

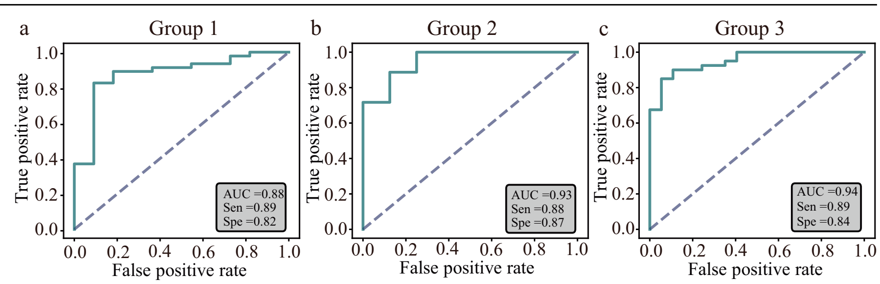


Table 4 | Sample size estimation of different sensitivity and specificity

Spe	Sen = 0.5		Sen = 0.6		Sen = 0.7		Sen = 0.8		Sen = 0.9	
	Nsen	Nspe	Nsen	Nspe	Nsen	Nspe	Nsen	Nspe	Nsen	Nspe
Spe = 0.5	68	68	65	68	57	68	86	68	86	68
Spe = 0.6	68	65	65	65	57	65	82	65	82	65
Spe = 0.7	68	57	65	57	57	57	72	57	72	57
Spe = 0.8	68	55	65	55	57	55	55	55	55	55
Spe = 0.9	68	32	65	32	57	32	32	32	32	32

Sen sensitivity, *Spe* specificity, *Nsen* number of Sen, *Nspe* number of Spe.

devices do not require a specific assessment location, such as an evaluation room. Moreover, VECA is less time-consumable (5-min test) and fully self-administered. This convenience significantly lowers the barriers to conducting cognitive impairment screenings, making early diagnosis and intervention more feasible.

In summary, this work introduces an approach to early detection of cognitive impairments by leveraging the portability, cost-effectiveness, resistance to interference, and operational convenience of VR devices. This not only assists physicians in detecting cognitive impairments on a larger scale, but also advances cognitive health screening technology, paving the way for future large-scale early dementia screenings.

While certain advancements have been made with the use of the VECA tool and satisfactory screening performance has been demonstrated, there are potential biases and limitations in our study.

Firstly, our participants were sourced from communities within Shenzhen, limiting the diversity of our sample. Information on other variables such as comorbidities, functional dependence, and medication usage could not be collected. This lack of comprehensive data may introduce bias in participant selection and confines the validity of our results to a specific cohort. Future studies should aim to collect more comprehensive data to fully account for the influencing factors that affect evaluation outcomes.

Moreover, the study did not thoroughly assess participants' acceptance of the VR intervention measures. While the VECA tool has shown potential for early screening of cognitive impairments, the subjective feelings and acceptance levels of participants towards such interventions remain unclear. Future research should include detailed assessments of participants' acceptance, incorporating both objective measurements and subjective feedback, to ensure that the intervention measures are broadly acceptable and applicable.

Another limitation is the absence of cross-training and testing across different clinical research centers, which restricts our evaluation of the model's generalizability and reliability. Future research should consider testing the VECA tool across a broader geographic location and population by establishing collaborations with multiple partners, thereby enhancing the model's generalizability and performance in diverse settings.

Additionally, the interactions between subjects and the VR headset are based on visual and literal capabilities. The absence of either capability could lead the system to misinterpret subjects' eye movements and produce deviated predictions. At the machine learning level, the validity and authenticity of the target variable (MoCA) are influenced by psychiatrists. Excessive noise in MoCA can distort our model's distribution and compromise its accuracy. To improve model performance, other informative features could be extracted from raw eye movements. Their validation and integration into the system should be explored in future studies.

Finally, we acknowledge these limitations as omissions in our preliminary research and look forward to addressing these issues through more comprehensive evaluations and broader testing to improve our research methodology.

In conclusion, this research suggests that VECA can be an efficient and portable tool for dementia screening. VECA has the potential to improve early diagnosis and treatment of dementia by helping physicians identify mild cognitive impairment in the early stages. This study highlights the application of eye-tracking technology, machine learning, and VR in cognitive evaluation. Further research with more data is needed to compare the effectiveness of VECA with traditional tests and dementia diagnosis outcomes.

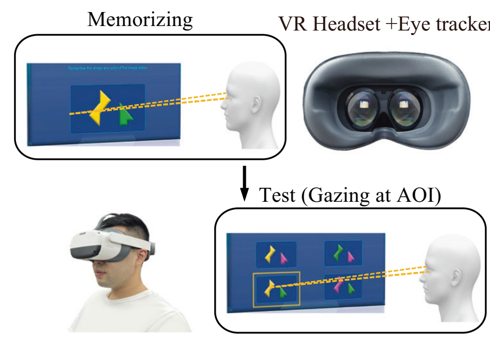


Fig. 5 | VR eye-tracking cognitive assessment based on VR head-mounted display and eye-tracking technology. In each task, participants were instructed to understand the task and fulfill related cognitive task. Multiple images including a correct answer (area of interest, AOI) and distractors (non-area of interest, non-AOI) were displayed in the VR scene, and subjects were instructed to identify and fixate on the correct answer.

Methods

This study has been approved by Shenzhen Baoan Chronic Hospital Ethics Committee and all subjects have signed written informed consent for participation. A total of 201 subjects were evaluated, including 88 from Fuan CRC (Community Rehabilitation Center) and 113 from Wanxiang CRC. Written informed consent was obtained from all participants involved in this study. Overview of participant demographics: (1) aged 65.5 ± 5.1 years; (2) 81 males (40.3%), female = 120 (59.7%); (3) years of education 9.4 ± 3.8 (Table 3). The study excluded participants who had psychiatric, ophthalmological, or hearing impairment disorders, as well as those who were unable to sit comfortably due to severe physical illness. Only participants with normal vision or corrected-to-normal vision without color blindness were included. A small number of participants were further excluded from the analysis if they encountered difficulties calibrating to the eye tracker or did not attempt to view the images. Exclusion criteria were applied in cases where the eye tracking equipment faced challenges in achieving proper pupil and corneal reflection due to physiological constraints or visual problems (e.g., droopy eyelid, cataracts, detached retinas, glaucoma, pupils too small), or if participants were unable to complete the eye tracking calibration procedure.

For data acquisition, demographic characteristics including age, gender, and education levels of the participants were collected and each of them underwent paper-based instruments of MoCA Chinese version. Figure 5 depicts the cognitive assessment system based on VR head-mounted display and eye-tracking technology used in the study, collectively referred to as

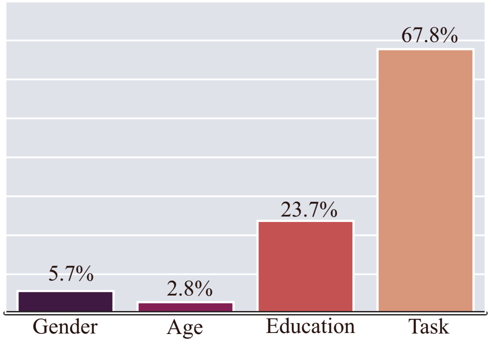


Fig. 6 | Feature importance analysis. The graph illustrates the relative importance of different features in the VRCA. The importance of each feature is expressed as a percentage. Gender accounts for 5.7%, age contributes 2.8%, education holds a weight of 23.7%, and task-related factors have the highest importance at 67.8%.

VECA, which takes 5 min only. Eye tracking calibration would be fulfilled before participants started the assessment to make sure gaze points captured by the eye tracker are valid. Cognitive tasks include memory (encoding, storage, and recall), visual attention (smooth pursuit), abstraction ability (ability to form abstract concepts), visuospatial working memory, visuospatial and executive function, calculation, language (understanding and execution of commands), and short-term memory binding. Detailed descriptions of each task are given in supplementary information (Supplementary Figs. 1–10). In each task, participants were usually given 3 s to understand the task and 5 or 8 s to fulfill related cognitive task. Multiple images including a correct answer (Area of interest, AOI) and distractors (non-Area of interest, non-AOI) were displayed in the VR scene, and subjects were instructed to identify and fixate on the correct answer. All eye movements were captured by the Tobii eye tracker embedded in Pico neo 3 pro eye, 95% of users can track better than 2° when their eyes are looking straight ahead (https://developer.tobii.com/xr/learn/eye-behavior/hardware-accuracy/).

For data processing, biological eye movements such as fixation and saccade were derived out of raw gaze points with velocity-threshold fixation identification (I-VT) algorithm (Supplementary Fig. 11), which automatically classifies excessive noise such as microsaccades as abnormal movements[28]. Gaze points that extremely deviate from their neighbors may be caused by winks or unexpected shivers and were also eliminated. The percentage of time the subject fixates on the AOI over total fixation time of a certain task is calculated and deemed as features. Previous research claimed that age is the biggest risk factor and made gender also a significant risk factor due to longer lifespan of female[29]. Therefore, demographic information such as age and gender of the subjects were supplemented to the feature set. Feature importances of Gender, age, education, cognitive task-related factors were calculated via Gradient Boosting Regression as in Fig. 6. Education level demonstrated nearly a quarter of importance (23.7%) and task-related factors have the highest importance at 67.8%.

Data preprocessing techniques such as one-hot and ordinal encoder were applied to categorical features. Standard scaling was also conducted for scale-sensitive models to eliminate the influence of various scales. All features and corresponding preprocess methods are summarized in Table 5. MoCA scores were categorized as the target variable so that our system could play as an efficient screening tool. In this way, the target variable better quantified the subjects' cognitive capability. The whole dataset was 70%–30% randomly split into train set and test set.

For modeling, Oyama et al. averaged each cognitive task score as the eye movement cognitive assessment score[21], (hereinafter referred to as VR model) which would be the baseline model. Support vector regression (SVR), Multi-layer perceptron (MLP), Lasso regression (Lasso), and Gradient Boost Regression Tree (GBRT) were adopted. Hyperparameters were optimized on train set via grid Search and 5-fold cross validations. Detailed model configurations are listed below:

Multi-layer perceptron (MLP): hidden layer sizes are configured as (16, 32, 16, 6), Rectified Linear Unit was used as activation function. Parameters were solved by Adam stochastic gradient-based optimizer. L2 penalty coefficient was set to 0.05. Max iterations of training was 1000.

Support vector regression (SVR): a linear kernel based support vector regressor with 0.5 L2 penalty is established. Insensitivity interval was set to 0.1 and shrinking option was activated to enhance model generalization performance.

Lasso regression: linear regression with 0.1 L1 penalized. Its maximum training iteration was set to 50.

Gradient Boosting Regression Tree (GBRT): 50 CART base decision trees with depth of 3 were ensembled and trained with learning rate of 0.1.

Table 5 | Features and corresponding preprocess methods

Feature name		Feature type	Preprocess methods
Gender		Categorical	One-hot encoding
Age		Numeric	Standard scaling
Education		Categorical	Ordinal encoding
Task score	vr_att vr_abs vr_calc4 vr_calc5 vr_calc6 vr_exec vr_mem8 vr_mem9 vr_mem10 vr_recall	Numeric	Standard scaling

Att attention, *abs* abstraction, *calc* calculation, *exec* execution, *mem* memory.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Non-identifiable patient data are available upon request to the corresponding author.

Code availability

The program of eye tracking data processing, feature extraction, modeling and evaluation based on the framework of VECA is openly available at https://github.com/zhangchi0923/VECA.

Received: 2 January 2024; Accepted: 29 July 2024;
Published online: 22 August 2024

References

1. WHO. Dementia. https://www.who.int/news-room/fact-sheets/detail/dementia (2021).
2. Petersen, R. C. Clinical practice. Mild cognitive impairment. *N. Engl. J. Med.* **364**, 2227–2234 (2011).
3. Jia, L. et al. Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study. *Lancet Public Health* **5**, e661–e671 (2020).
4. Jia, J. et al. The cost of Alzheimer's disease in China and re-estimation of costs worldwide. *Alzheimers Dement.* **14**, 483–491 (2018).
5. Blondell, S. J., Hammersley-Mather, R. & Veerman, J. L. Does physical activity prevent cognitive decline and dementia?: a systematic review and meta-analysis of longitudinal studies. *BMC Public Health* **14**, 1–12 (2014).

6. 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).
7. Park, H. et al. Combined intervention of physical activity, aerobic exercise, and cognitive exercise intervention to prevent cognitive decline for patients with mild cognitive impairment: a randomized controlled clinical study. *J. Clin. Med.* **8** https://doi.org/10.3390/jcm8070940 (2019).
8. Dubois, B. et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *Lancet Neurol.* **13**, 614–629 (2014).
9. Lai, M.-L. et al. A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educ. Res. Rev.* **10**, 90–115 (2013).
10. Skaramagkas, V. et al. Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE Rev. Biomed. Eng.* **16**, 260–277 (2023).
11. Lagun, D., Manzanares, C., Zola, S. M., Buffalo, E. A. & Agichtein, E. Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *J. Neurosci. Methods* **201**, 196–203 (2011).
12. Nie, J. et al. Early diagnosis of mild cognitive impairment based on eye movement parameters in an aging Chinese population. *Front. Aging Neurosci.* **12**, 221 (2020).
13. Haque, R. U. et al. VisMET: a passive, efficient, and sensitive assessment of visuospatial memory in healthy aging, mild cognitive impairment, and Alzheimer's disease. *Learn Mem.* **26**, 93–100 (2019).
14. Pereira, M. et al. Visual search efficiency in mild cognitive impairment and Alzheimer's disease: an eye movement study. *J. Alzheimers Dis.* **75**, 261–275 (2020).
15. Jiang, J. et al. A novel detection tool for mild cognitive impairment patients based on eye movement and electroencephalogram. *J. Alzheimers Dis.* **72**, 389–399 (2019).
16. Chehrehnegar, N., Shati, M., Esmaeili, M. & Foroughan, M. Executive function deficits in mild cognitive impairment: evidence from saccade tasks. *Aging Ment. Health* **26**, 1001–1009 (2022).
17. Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).
18. Cecchini, M. A. et al. Profiles of cognitive impairment in the continuum from normal cognition to Alzheimer's clinical syndrome: contributions of the short-term memory binding tests. *Int. J. Geriatr. Psychiatry* **35**, 1331–1340 (2020).
19. Cecchini, M. A. et al. Deficits in short-term memory binding are detectable in individuals with brain amyloid deposition in the absence of overt neurodegeneration in the Alzheimer's disease continuum. *Brain Cogn.* **152**, 105749 (2021).
20. Parra, M. A., Schumacher, M. & Fernández, G. A novel peripheral biomarker for mild cognitive impairment and Alzheimer's disease. *Alzheimers Dement.* **16** https://doi.org/10.1002/alz.044956 (2020).
21. Oyama, A. et al. Novel method for rapid assessment of cognitive impairment using high-performance eye-tracking technology. *Sci. Rep.* **9**, 1–9 (2019).
22. Mengoudi, K. et al. Augmenting dementia cognitive assessment with instruction-less eye-tracking tests. *IEEE J. Biomed. Health Inform.* **24**, 3066–3075 (2020).
23. Zola, S. M., Manzanares, C. M., Clopton, P., Lah, J. J. & Levey, A. I. A behavioral task predicts conversion to mild cognitive impairment and Alzheimer's disease. *Am. J. Alzheimers Dis. Other Demen.* **28**, 179–184 (2013).
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. Corriveau Lecavalier, N., Ouellet, E., Boller, B. & Belleville, S. Use of immersive virtual reality to assess episodic memory: A validation study in older adults. *Neuropsychol. Rehabil.* **30**, 462–480 (2020).
26. Jiang, Z. et al. Multimodal Mental Health Digital Biomarker Analysis From Remote Interviews Using Facial, Vocal, Linguistic, and Cardiovascular Patterns. *IEEE J. Biomed. Health Inf.* **28**, 1680–1691 (2024).
27. Lu, J. et al. Montreal cognitive assessment in detecting cognitive impairment in Chinese elderly individuals: a population-based study. *J. Geriatr. Psychiatry Neurol.* **24**, 184–190 (2011).
28. Salvucci, D. D. & Goldberg, J. H. In *Proceedings of the 2000 symposium on Eye tracking research & applications*. 71-78.
29. Beam, C. R. et al. Differences between women and men in incidence rates of dementia and Alzheimer's disease. *J. Alzheimer's Dis.* **64**, 1077–1083 (2018).

Acknowledgements
The authors gratefully acknowledge the financial support from the Shenzhen Fundamental Research Program, under Grant No. JCYJ20220531091006014, which made this research possible. Heartfelt gratitude is also extended to Mr. Lee Yick Kuen for his invaluable support and encouragement throughout the course of this research.

Author contributions
X.Z. organized the entire project, designed the contents of experiments and VR eye-tracking cognitive tests. Q.Y. is responsible for screening scenes, clinical suggestions, and other resources. Y.X. and C.Z. are responsible for data acquisition, data processing, and modeling. B.P. provided algorithm suggestions. All authors reviewed and approved the final version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-024-01206-5.

Correspondence and requests for materials should be addressed to Qing Yuan or Xu Zhang.