中山大學 中山眼科中心 ZHONGSHAN OPHTHALMIC CENTER, SUN YAT-SEN UNIVERSITY

Contents lists available at Lighthouse Scholar

EYE SCIENCE

journal homepage: journal.gzzoc.org.cn

doi: 10.12419/es26081401
View this article at: https://doi.org/10.12419/es26081401

• Original Article •

REMoCA: Rapid Eye-Tracking Assessment of MoCA-Defined Cognitive Impairment Across Visual-Acuity Strata

Xun Wang (王珣)[1#], Shuyi Zhang (张树意)[1,4#], Zhenzhe Lin (林桢哲)[1#], Jia Wang (王佳)[2#], Yanting Chen (陈彦婷)[2#], Dong Zheng (郑东)[3], Carol Y Cheung (张艳蕾)[4], Xiaoming Rong (容小明)[2], Xueer Tu (涂雪儿)[1], Wei Wang (王伟)[1], Yunjian Huang (黄运坚)[1], Lanqin Zhao (赵兰琴)[1], Duoru Lin (林铎儒)[1], Xiaohang Wu (吴晓航)[1], Yahan Yang (杨雅涵)[1], Dongni Wang (王东妮)[1], Xiaoming Zhou (周晓明)[3], Nannan Pan (潘楠楠)[3], Yayong Cui (崔雅泳)[3], Junhua Xu (许俊华)[3], Dongyuan Yun (云东源)[1], Wei Xiao (肖薇)[1], Xulin Zhang (张栩琳)[1], Meimei Dongye (东野枚枚)[1], Yi Li (李艺)[2], Zhenzhen Liu (刘臻臻)[1], Patrick Yu-Wai-Man[5,6,7,8], Jason C Yam (任卓昇)[4], Yamei Tang (唐亚梅)[2], Haotian Lin (林浩添)[1,9,10]

1. State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Ophthalmology and Vision Science, Guangdong Provincial Clinical Research Center for Ocular Diseases, Guangzhou 510060, China.
2. Department of Neurology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China.
3. Department of Neurology, The Affiliated Brain Hospital of Guangzhou Medical University, Guangzhou 510370, China.
4. Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China.
5. John van Geest Center for Brain Repair and MRC Mitochondrial Biology Unit, Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 0PY, United Kingdom.
6. Cambridge Eye Unit, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, United Kingdom.
7. Moorfields Eye Hospital NHS Foundation Trust, London EC1V 2PD, United Kingdom.
8. Institute of Ophthalmology, University College London, London EC1V 9EL, United Kingdom.
9. Hainan Eye Hospital and Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Haikou 570311, China.
10. Department of Genetics and Biomedical Informatics, Zhongshan School of Medicine, Institute for Frontier Interdisciplinary Research in Health Sciences and Technology, Sun Yat-sen University, Guangzhou 510080, China.

Received date: 2026-08-17; **Revised date:** 2026-08-31; **Accepted date:** 2026-09-08; **Published online:** 2026-09-14

# These authors contributed equally

Corresponding authors: Yamei Tang, E-mail: tangym@mail.sysu.edu.cn;
Haotian Lin, E-mail: linht5@mail.sysu.edu.cn.

Full Text

HIGHLIGHTS

1. Critical Discoveries and Outcomes

• Among five eye-movement paradigms evaluated in 1,212 older adults with measured near visual acuity, only the antisaccadic task discriminated participants with and without MoCA-defined cognitive impairment at every visual-acuity stratum, indicating that its discriminative signal was retained across the levels of near visual acuity examined.

2. Methodological Innovations

• REMoCA processes eye-movement trajectory videos directly with a pseudo-3D network rather than a predefined set of summary metrics, automating trial-level classification of eye-movement performance and combining it with demographic and clinical risk factors in a two-stage framework. Saliency mapping permits visual inspection of model-attended regions, and dominance analysis summarises the relative contribution of predictors.

3. Prospective Applications and Future Directions

• A shortened antisaccadic-task workflow was completed within three minutes in a prospective community-based proof-of-concept cohort, supporting technical feasibility outside hospital settings. Larger and more representative community studies, improved calibration, and explicit evaluation of decision thresholds are required before this approach can be considered for triage, referral, or population-level screening.

Abstract: **Purpose**: To examine whether eye movement (EM) performance discriminates between older adults with and without cognitive impairment across varying levels of near visual acuity (VA), and to develop and evaluate an antisaccadic-task-based framework (REMoCA) for detecting cognitive impairment. **Methods**: In this multicenter study, 1,428 participants aged 50 years or older completed video-based assessments comprising five EM tasks: fixation, saccade, antisaccade (AST), and the horizontal and vertical smooth pursuit. Cognitive status was assessed using the Montreal Cognitive Assessment (MoCA), with cognitive impairment defined as an education-adjusted MoCA score < 26 (MoCA-defined CI). Among the 1,212 participants with valid near VA and MoCA measurements, EM error rates were compared between those with and without MoCA-defined CI across three VA strata (logMAR ≤ 0.3, 0.3 to 0.5, and > 0.5) using linear regression adjusted for age, sex, educational level, hypertension, and diabetes. *P* values were corrected for multiple comparisons using the false discovery rate method. REMoCA, a two-stage framework combining a pseudo-3D network for trial-level AST classification with demographic and clinical risk factors, was developed and evaluated in an external test set and a prospective community-based proof-of-concept cohort. **Results**: Across all five EM tasks, the MoCA-defined CI group exhibited higher error rates, with between-group differences being most pronounced in the higher-VA stratum. Only the AST demonstrated significantly higher error rates in the MoCA-defined CI group across all VA strata (adjusted $P < 0.05$); the other tasks showed significant differences only in a subset of strata. REMoCA achieved an area under the curve (AUC) of 0.907 (95% CI: 0.850 to 0.963) in the external test set and 0.768 (95% CI: 0.694 to 0.842) in the prospective community-based proof-of-concept cohort, where the shortened AST-based workflow was completed within three minutes. **Conclusions**: Antisaccadic performance retained its discriminative ability for detecting MoCA-defined CI across the examined VA strata. These findings support the further evaluation of AST-based assessment as a rapid, VA-tolerant approach for detecting MoCA-defined CI, pending validation in larger, more representative

independent cohorts.

Keywords: eye movement; antisaccadic task; MoCA-defined cognitive impairment; visual acuity; rapid assessment

Cite this article as: Wang X, Zhang SY, Lin ZZ, Wang J, Chen YT, Zheng D, Cheung CYL, Rong XM, Tu XE, Wang W, Huang YJ, Zhao LQ, Lin DR, Wu XH, Yang YH, Wang DN, Zhou XM, Pan NN, Cui YY, Xu JH, Yun DY, Xiao W, Zhang XL, Dongye MM, Li Y, Liu ZZ, Yu-Wai-Man P, Yam JC, Tang YM, Lin HT. REMoCA: Rapid Eye-Tracking Assessment of MoCA-Defined Cognitive Impairment Across Visual-Acuity Strata. Eye Science. 2026, 0: 1-16. doi: 10.12419/es26081401.

Introduction

The integration of visual and cognitive functions is fundamental to human information processing.[1] Goal-directed behavior depends on both visual input and cognitive control.[2] Declines in either domain may alter observable performance,[3] yet isolating a cognition-specific signal from co-occurring visual decline remains challenging. This challenge is particularly relevant in aging populations, in whom sensory decline and cognitive impairment frequently coexist and can confound behavioral markers of cognition.[4-5]

Cognitive impairment is common in older adults: the pooled worldwide prevalence of mild cognitive impairment among community dwellers aged 50 years or older is an estimated 15.6%,[6] and affected individuals are at substantially increased risk of progression to dementia. Its early identification can therefore facilitate timely clinical assessment and intervention.[4] The Montreal Cognitive Assessment (MoCA) is among the most widely used screening instruments for cognitive impairment, offering high sensitivity for detecting mild cognitive impairment and coverage of multiple cognitive domains.[7] However, administration of the MoCA requires a trained examiner and a dedicated in-person session, which may limit accessibility in resource-limited settings.[8] Neuroimaging offers complementary structural information: white matter lesions (WMLs), a common feature of cerebral small vessel disease, are associated with cognitive decline, particularly in executive function and processing speed.[9] Yet, magnetic resonance imaging (MRI) is costly and impractical for population-level screening. These considerations underscore the need for rapid, scalable approaches that complement existing screening strategies.

Eye movement (EM) provides an objective, rapid, and automated behavioral readout of cognitive function.[10] Advances in eye-tracking technology enable high-resolution measurement of gaze position through near-infrared detection of pupillary and corneal reflections.[11] Because EM assessment inherently depends on visual input, however, the derived cognitive signal could be confounded by reduced visual acuity (VA). Most prior studies have focused on cohorts with normal VA,[11-12] despite the frequent co-occurrence of visual and cognitive impairment among older adults.[5] Establishing that EM measures retain their discriminative ability across varying levels of VA is therefore essential before clinical translation.

To address these gaps, we compared EM performance across five paradigms between older adults with and without cognitive impairment, operationally defined as an education-adjusted MoCA score below 26 (MoCA-defined CI), and examined whether any such differences were retained across levels of VA. On this basis, we selected the antisaccadic task (AST), which showed the most consistent between-group separation across the VA strata examined, to develop the Rapid Eye-Movement-based Cognitive Assessment (REMoCA), a two-stage framework for detecting MoCA-defined CI. We then evaluated REMoCA in an external test set and assessed its feasibility in a prospective, community-based proof-of-concept cohort.

Methods

Study design and participants

This multicenter study comprised three distinct phases: a prospectively collected development cohort (September 2020 to July 2025), a pre-enrolled multicenter test cohort (February 2021 to February 2022), and a prospective community-based proof-of-concept cohort (July 2025 to November 2025). The study examined the associations between VA and MoCA-defined CI and EM behavior, and developed a predictive model for MoCA-defined CI. Development participants were recruited from (1) neurology inpatients at Sun Yat-sen Memorial Hospital (SYMH) and (2) community-dwelling individuals recruited through an eye disease screening program advertised via social media. Eligible EM videos were partitioned at the participant level into training, validation, and internal test sets (7:1.5:1.5) for stage 1, which classified trial performance as correct or incorrect. Detailed criteria for this classification are provided in Supplementary File 1. After excluding 127 participants lacking the outcome data required for stage 2, the remaining participants were divided into training and internal test sets (7:3). Stage 1 used trial-level EM videos, whereas stage 2 used participant-level EM accuracy metrics and risk factors (RFs) to detect MoCA-defined CI. The stage 2 training and internal test sets were nested within the corresponding stage 1 training, validation and internal test sets, respectively, ensuring that no participant in the stage 2 internal test set had contributed data to stage 1 model training. Consequently, the participant-level EM-accuracy features used as stage 2 inputs were therefore derived from stage 1 predictions generated on held-out data for each participant, thereby preventing information leakage between the two stages. The external test set consisted of cataract outpatients at Zhongshan Ophthalmic Center (ZOC), neurology outpatients at SYMH, and neurology inpatients at Huiai Hospital. The final phase integrated the shortened REMoCA workflow into a community health screening program as a prospective proof-of-concept feasibility study.

Individuals aged 50 years or older were eligible. Participants unable to complete eye-tracking calibration or the EM protocol due to ocular, cognitive, physical, or systemic conditions were excluded from EM analyses. Individuals with other ocular diseases causing reduced vision were not routinely excluded, provided that calibration and testing could be successfully completed. Those without a valid MoCA outcome were ineligible for development of the stage 2 network for detecting MoCA-defined CI, and those who did not undergo MRI were excluded from the exploratory WML analysis. Written informed consent was obtained from all participants or their legal guardians in accordance with ethical guidelines.

This study was approved by the Institutional Review Board of ZOC, Sun Yat-sen University (Approval No. 2020KYPJ015), and adhered to the Declaration of Helsinki. It was prospectively registered with the Clinical Research Internal Management System of ZOC and ClinicalTrials.gov (NCT04236375). Participant demographics (age, sex, education, and clinical characteristics) are summarized in Table 1.

Data collection and reference standards

EMs were recorded by well-trained technicians using a standardized protocol, and all sessions were videotaped. The setup included a 1,280 × 720 pixel screen integrated with an eye tracker (EyeControl, QY-I, Shanghai Qingyan Technology Co., Ltd.) operating at a sampling rate of 100 Hz. Participants were instructed to stabilize their heads using a chinrest positioned 60 cm from the screen and perform EM tests in a fixed sequence: the fixation task (FT, 10 trials), the saccadic task (ST, 10 trials), AST (12 trials), the horizontal smooth pursuit task (HSPT, 8 trials), and the vertical smooth pursuit task (VSPT, 8 trials). This entire session lasted approximately 10 minutes, generating 48 videos per participant. EM performance on each trial was classified as correct or incorrect according to predefined criteria (Supplementary File 1, p. 22) and verified by video review. For example, a correct AST trial required participants to suppress the reflexive saccade toward the peripheral stimulus, compute its opposite direction (180° from the stimulus), and execute a voluntary

Table 1 Baseline characteristics of the internal dataset, external test set, and prospective community cohort

	Internal dataset	External test set	Prospective community cohort
No. of participants	1,326	102	124
No. of participants with MoCA <26	902 [a]	85	60
No. of participants with MoCA ≥26	297 [a]	17	64
Age, years			
Mean (SD)	65.96 (8.05)	65.77 (9.71)	69.88 (4.31)
Range	50–89	50–87	63–82
Sex			
Male, *n* (%)	658 (49.6)	57 (55.9)	47 (37.9%)
Female, *n* (%)	668 (50.4)	45 (44.1)	77 (62.1%)
Education, years			
Mean (SD)	11.04 (4.03)	10.59 (3.57)	13.10 (3.42)
Range	0–19	0–19	0–19
Participants with hypertension, *n* (%)	539 (40.6)	50 (49.0)	38 (30.6%)
Participants with diabetes, *n* (%)	270 (20.4)	28 (27.5)	12 (9.6%)
Near visual acuity, logMAR			
Median (IQR)	0.22 (0.24)	0.22 (0.20)	0.22 (0.19)
Range	−0.08–1.30	0–0.52	0–1.00
MoCA			
Median (IQR)	22 (8)[a]	20 (7.75)	26 (4)
Range	3–30	3–30	13–30
Fazekas scale among participants with MoCA <26			
Median (IQR)	2 (1)[b]	2 (3)[c]	NA
Range	0–6	0–6	NA

Data are presented as n, n (%), mean (SD), median (IQR), and range unless otherwise specified. The internal dataset includes all participants with valid EM data (n = 1,326); 1,199 were included in stage 2 after excluding 127 without the required outcome data. [a]A subset of 1,199 participants; [b]A subset of 556 participants; [c]A subset of 29 participants. SD, standard deviation; IQR, interquartile range; NA, not applicable; MoCA, Montreal Cognitive Assessment; WML, white matter lesion.

saccade accordingly. In the subsequent community-based proof-of-concept workflow, only the AST and RF were used, reducing the automated assessment time to approximately 3 minutes.[13]

Following the EM examination, participants were given a 10-minute rest period, after which near VA was assessed binocularly under standard lighting using the logarithm of the minimum angle of resolution (logMAR) near VA tumbling E chart by experienced technicians. Subsequently, cognitive function was evaluated using the Chinese version of MoCA, administered by neuropsychologists to ensure validity. The maximum score was 30; one additional point was added to the total score for individuals with 12 or fewer years of education. A score lower than 26 was used to define MoCA-defined CI among the study participants (i.e., aged ≥50 years) for model development and evaluation.[7] Finally, brain MRI was performed on 1.5T or 3T scanners (Magnetom Avanto/Vida/Skyra, Siemens, Erlangen, Germany), including axial fast spin-echo T1- and T2-weighted, and coronal fluid-attenuated inversion recovery (FLAIR) sequences. WMLs were evaluated using the Fazekas scale, scoring periventricular and deep white matter separately from 0 to 3 and summing them to yield a total score of 0–6.[14] Total scores >2 (i.e., 3–6) were classified as moderate-to-severe WMLs.[15] Two junior physicians (J.W. and S.Z. each with five years of clinical experience), blinded to clinical and EM data, independently evaluated all MRI scans (inter-rater $\kappa = 0.853$). Prior to formal scoring, both raters completed a calibration exercise on a set of training cases under the supervision of a senior neurologist (Y.T.) to standardize application of the Fazekas scale. Discrepancies were resolved by consensus between the two raters; when consensus could not be reached, a senior neurologist (Y.T., with more than 25 years of clinical experience) provided adjudication. RF, including age, sex, educational level, hypertension, and diabetes status, as well as neurological and systemic disorders, were extracted from electronic medical records or obtained by self-reported.

Development of REMoCA

The AST was selected for REMoCA because it showed significant between-group differences across all three VA strata and the most consistent discriminative performance among the five paradigms examined. REMoCA is a two-stage network for screening MoCA-defined CI (Figure 1e; Supplementary Table 1; Supplementary file 2, p. 23). Stage 1 used independent pseudo-3D (P3D) networks to automate the trial-level classification of EM performance as correct and incorrect. The P3D networks incorporated 2D spatial and 1D temporal convolutions within a residual network architecture to efficiently extract spatiotemporal features from EM videos, thereby significantly reducing computational and memory requirements.[16] Supplementary Figure 1 shows the architecture of the P3D networks, including their series and parallel fusion schemes. EM videos were preprocessed by resampling each video to a fixed, task-specific frame count (corresponding to one complete task cycle) and resizing the frames to a uniform resolution of 224 × 224 pixels. The specific frame counts for each task are detailed in Supplementary Table 1. The resulting frames were then standardized channel-wise using the dataset mean and standard deviation. Stage 2 consisted of a logistic regression model that integrated the stage 1 outputs and RFs to screen MoCA-defined CI. For the detection of moderate-to-severe WMLs, Stage 2 was trained exclusively on the subgroup of participants with a MoCA score < 26 (Supplementary Figure 2a). For benchmarking, three additional models were developed: a RF machine-learning model (MLM) based solely on RFs, and an EM MLM based solely on features from all EM tasks, and a hybrid MLM incorporating both RFs and all EM features.

To enhance model interpretability, EM videos from testing datasets were selected for saliency analysis using Eigen-CAM for the Stage 1 network.[17] This method generated visual heatmaps by aggregating and averaging convolutional activations across all frames of each trial. For stage 2, dominance analysis was used to rank the relative importance of each EM task in the model's performance in detecting MoCA-defined CI.

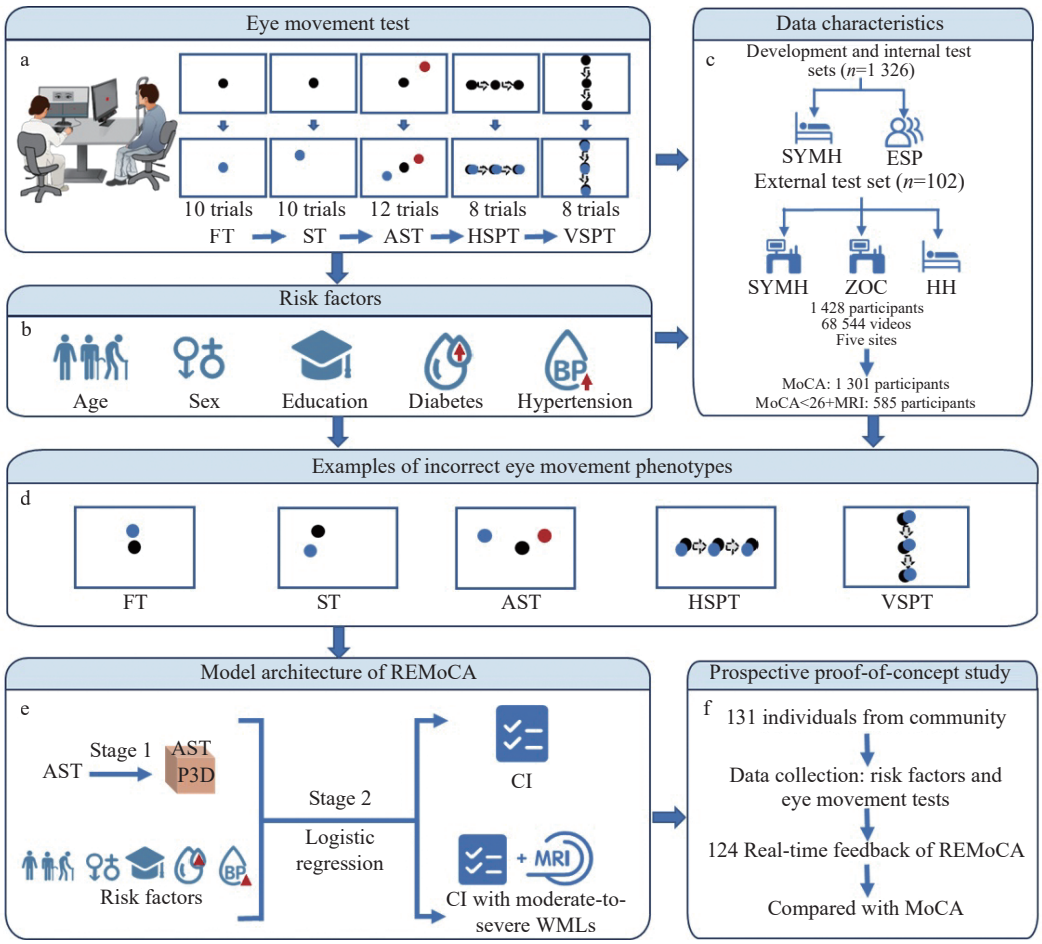

Figure 1 Overall study pipeline

(a) The five-task comparison protocol used a fixed sequence—FT (10 trials), ST (10), AST (12), HSPT (8), and VSPT (8)—and lasted approximately 10 minutes. (b) RF were extracted from health records. (c) A total of 68,544 videos from 1,428 participants across five centers were collected; 1,301 participants had valid MoCA results. The exploratory WML analysis included 585 participants with MoCA <26 and valid Fazekas ratings, of whom 145 had moderate-to-severe WMLs. (d) AST-derived error patterns discriminated between MoCA-defined cognitive groups across VA strata. (e) REMoCA used AST videos to classify trial-level behavior in stage 1 and combined participant-level AST performance with RF to predict MoCA-defined cognitive impairment in stage 2; WML classification was an exploratory secondary analysis in a restricted subset. (f) A prospective community proof-of-concept cohort evaluated a shortened AST plus RF workflow requiring approximately three minutes. EM, eye movement; FT, fixation task; ST, saccadic task; AST, antisaccadic task; HSPT, horizontal smooth pursuit task; VSPT, vertical smooth pursuit task; RF, risk factors; VA, visual acuity; MoCA, Montreal Cognitive Assessment; WML, white matter lesion.

Prospective proof-of-concept cohort in a community-based health screening program

REMoCA was prospectively integrated into an ongoing community-based health screening program to evaluate its feasibility in real-world settings (Figure 1f; Supplementary Figure 2b). This shortened workflow used AST videos and RF collected by three trained technicians and transmitted to a computer equipped with a graphics processing unit (GPU) for automated processing via the REMoCA framework. All three technicians completed standardized training on the study protocol, which covered positioning participant on the chinrest, delivering task instruction, and operating the eye tracker. When REMoCA classified a participant as screen-positive for MoCA-defined CI (i.e., above the predefined threshold), the system generated a real-time referral recommendation. The entire workflow, from initial participant instruction to automated result generation, required approximately 3 minutes per individual. For ethical

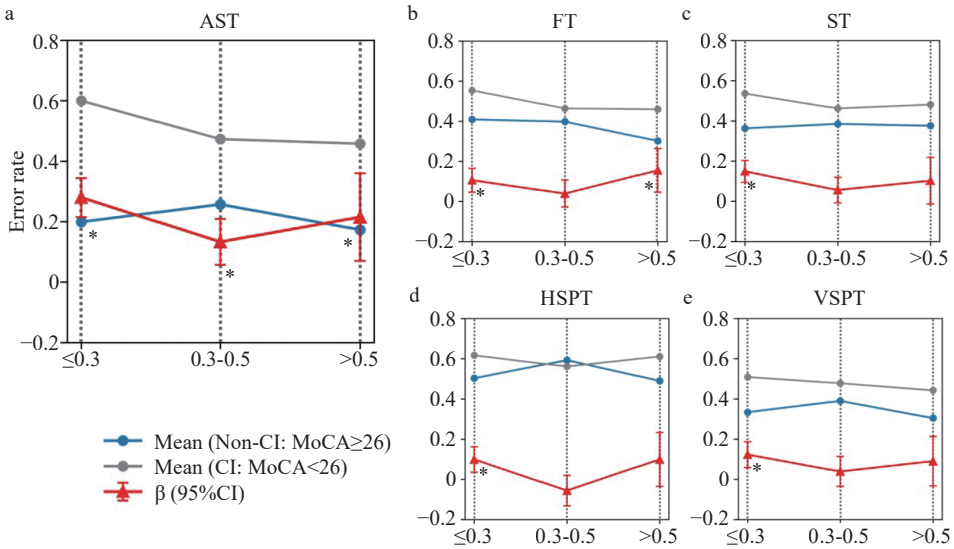

Figure 2 Mean error rates and between-group differences across five EM tasks stratified by cognitive status and VA

Blue lines indicate mean error rates for the non-CI group (MoCA ≥26), gray lines indicate mean error rates for the MoCA-defined CI group (MoCA <26), and red lines represent between-group differences estimated by linear regression (β) with 95% confidence intervals. VA was stratified by logMAR thresholds: the high-VA stratum (logMAR ≤0.3), the intermediate-VA stratum (logMAR 0.3–0.5), and the low-VA stratum (logMAR >0.5). LogMAR is a standard measure of VA, with higher values indicating poorer VA. (a) Antisaccadic task; (b) Fixation task; (c) Saccadic task; (d) Horizontal smooth pursuit task; (e) Vertical smooth pursuit task. *Asterisks indicate statistically significant differences in error rates between the MoCA-defined CI and non-CI groups based on linear regression analysis. *P* values were corrected for multiple comparisons using the false discovery rate method. EM, eye movement; VA, visual acuity; CI, MoCA-defined cognitive impairment.

reasons, the referral recommendations generated by REMoCA were strictly advisory and did not influence clinical management; all participants subsequently underwent MoCA assessment administered by qualified personnel as part of the standard protocol.

Statistical analysis

To characterize EM task performance across different VA levels, a VA-stratified analysis was conducted. Participants were stratified into three VA strata based on logMAR values (logMAR ≤0.3, 0.3< logMAR ≤0.5, and logMAR >0.5). Within each stratum, EM task error rates were compared between participants with and without MoCA-defined CI using linear regression models adjusted for age, sex, educational level, hypertension, and diabetes status. *P* values were corrected for multiple comparisons using the false discovery rate method.

For model development, the training set (n = 839, comprising 631 with MoCA-defined CI and 208 without) achieved an events-per-variable ratio of approximately 20 for the 10 candidate predictors. The 10 predictors included the error rates of five EM tasks (FT, ST, AST, HSPT, and VSPT) and five RF (age, sex, education level, history of hypertension, and history of diabetes). Accordingly, a total of 1,199 participants (902 with and 297 without MoCA-defined CI) were included in the model development dataset, comprising the training set (n = 839) and internal test set (n = 360) in a 7:3 split. This sample size met recommended criteria for predictive modeling, supporting model complexity while mitigating the risk of overfitting. For the prospective proof-of-concept cohort, the sample size was estimated before recruitment to ensure sufficient power to assess REMoCA performance. Although the estimated prevalence of cognitive impairment in elderly Chinese populations is 14.7%,[18] the prevalence in our screening program was expected to be higher owing to the potential association between poorer VA and cognitive impairment, as well as our broader study definition based on MoCA <26.[7] The prevalence of

cognitive impairment was therefore assumed to be 30%. We estimated that at least 23 individuals with MoCA-defined CI and 53 without MoCA-defined CI (76 in total) would be required to detect an AUC of 0.7 or greater with >80% power at a two-sided α of 0.05.

For the EM behavior classification model (Stage 1), a default probability threshold of 0.5 was used. For the MoCA-defined CI screening model (Stage 2), the optimal cutoff was determined using the Youden index, determined separately within the internal and the external test sets. Additionally, a cutoff corresponding to 85% sensitivity was derived in the internal test set, locked before recruitment of the prospective cohort, and subsequently applied to the prospective cohort for MoCA-defined CI.[19–21] AUCs were reported for all datasets. Accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and the F1 score were reported for the external test set and the prospective cohort at their respective predefined thresholds. Metrics were reported with Wald 95% confidence intervals (95% CIs), and AUCs were compared using DeLong's test. Subgroup analyses were stratified by age (≤65 *vs* >65 years), sex (female *vs* male), education (≤9 *vs* >9 years), hypertension status, diabetes status, and near vision impairment status (defined as logMAR near VA >0.3).[22-23] All statistical tests were two-tailed, and a *P* value of less than 0.05 was considered statistically significant. *P* values were corrected for multiple comparisons using the false discovery rate (FDR) method. All analyses were performed using SPSS version 24 (IBM Statistics, IBM Corp.).

Results

Dataset characteristics

A total of 1,539 participants were assessed for eligibility for EM performance analysis, model development, and multicenter testing. The EM assessment included five tasks (FT, ST, AST, HSPT, VSPT). Of these, 111 (7.21%) were excluded due to ocular diseases or poor physical condition that precluded completion of EM tasks (including ptosis or strabismus in either eye, severe CI, and cerebrovascular diseases). The remaining 1,428 participants provided valid EM data. Among them, 127 declined MoCA testing or MRI and were included only in Stage 1 model training (EM-behavior classification from EM videos). For stage 2 (detection of MoCA-defined CI), 1,301 participants were included, comprising 1,199 for model development and 102 for external testing. Among these, 987 (75.86%) had MoCA <26. Of those with MoCA <26, 585 had valid Fazekas scores, of whom 145 (24.79%) had moderate-to-severe WMLs. For the prospective proof-of-concept cohort, 131 individuals were assessed during a community health screening program, and 124 individuals (94.7%) were enrolled, of whom 60 (48.39%) had MoCA scores <26 (Figure 1, Table 1, and Supplementary Table 2). The participant recruitment process is shown in Supplementary Figure 2, and the distribution of neurological disorders is summarized in Supplementary Table 3.

Cognitive-group differences in AST performance persist across VA strata

A total of 1,212 participants with valid VA and MoCA measurements were included. Participants with an education-adjusted MoCA scores <26 were classified into the MoCA-defined CI group, whereas those with a score ≥26 were classified into the non-CI group.[7] Across all five EM tasks, the MoCA-defined CI group exhibited higher error rates. Descriptively, between-group differences were most pronounced in the highest-VA stratum (Figure 2). Among the EM tasks evaluated, only the AST demonstrated significantly higher error rates in the MoCA-defined CI group across all VA stratum (adjusted *P* <0.05); the other tasks showed significant differences only in a subset of strata (Figure 2). Baseline characteristics stratified by VA stratum and cognitive status are summarized in Supplementary Table 4.

AST-derived REMoCA enables MoCA-defined CI screening across datasets

For trial-level classification of the AST in the Stage 1 network, REMoCA achieved an AUC of 0.967 (95% CI, 0.958–0.975), a sensitivity of 0.922 (95% CI, 0.909–0.934), and specificity of 0.938 (95%

CI, 0.927–0.949) in the internal test set, and an AUC of 0.986 (95% CI, 0.980–0.991), a sensitivity of 0.966 (95% CI, 0.957–0.974), and a specificity of 0.924 (95% CI, 0.911–0.936) in the external test set (Figures 3a and 3b; Supplementary Table 5). For classification of MoCA-defined CI in the Stage 2 network, REMoCA achieved an AUC of 0.802 (95% CI, 0.761–0.843), a sensitivity of 0.860 (95% CI, 0.824–0.896), and a specificity of 0.506 (95% CI, 0.454–0.557) in the internal test set, and an AUC of 0.907 (95% CI, 0.850–0.963), a sensitivity of 0.918 (95% CI, 0.864–0.971), and a specificity of 0.824 (95% CI, 0.750–0.898) in the external test sets (Figures 3d and 3e; Table 2). In the external test set, REMoCA outperformed both the risk factor (RF) machine learning model (MLM) and the all-EM MLM, and achieved discriminative performance comparable to that of the hybrid MLM (combining all-EM and RF features) (Figure 3e; Table 2; Supplementary Tables 6 and 7).

As an exploratory secondary analysis, we assessed the classification of moderate-to-severe WML (total Fazekas score >2) within a subgroup restricted to participants with MoCA <26 and valid Fazekas ratings (Supplementary Table 8).[14-15] Subgroup results are detailed in Supplementary Tables 9 and 10. Saliency heatmaps were generated to visualize the regions attended by the Stage 1 network, and dominance analysis (DA) was used to estimate the relative contributions of individual EM tasks and RFs to the Stage 2 model (Supplementary Figure 3).

REMoCA shows lower but feasible performance in the prospective community-based cohort

In the prospective community-based proof-of-concept cohort, all participants completed AST examinations within 3 minutes and underwent real-time cognitive assessment using REMoCA. The development dataset was enriched for participants with MoCA-defined CI (75.86%), whereas the prospective community cohort had a lower prevalence

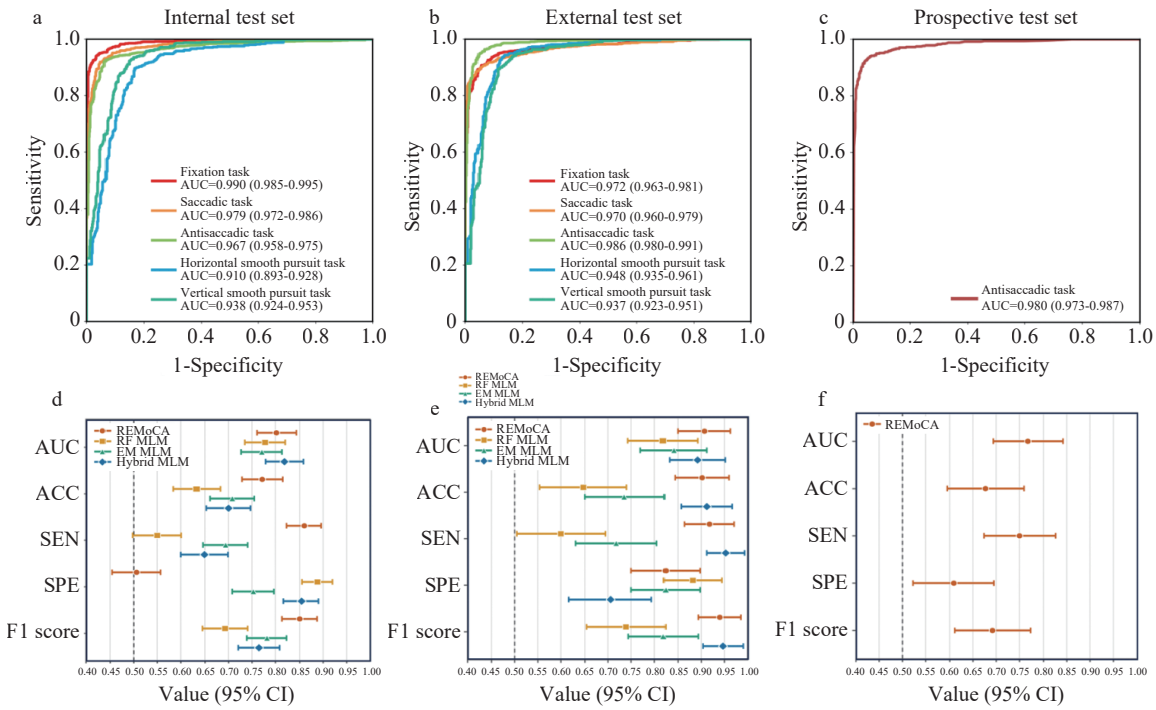

Figure 3 Performance of the MLMs

(a–c) AUCs for stage 1 EM-performance classification in the internal test set, external test set, and prospective community dataset. (d–f) Performance of stage 2 models for MoCA-defined cognitive impairment across the internal test set, external test set, and prospective community dataset. AUC, area under the receiver operating characteristic curve; RF, risk factors; EM, eye movement; MLM, machine-learning model; ACC, accuracy; SEN, sensitivity; SPE, specificity.

Table 2 Performance of REMoCA for predicting MoCA-defined cognitive impairment

	AUC (95% CI)	ACC (95% CI)	SEN (95% CI)	SPE (95% CI)	PPV (95% CI)	NPV (95% CI)	F1 score (95% CI)
Internal test set	0.802 (0.761–0.843)	0.772 (0.729-0.816)	0.860 (0.824-0.896)	0.506 (0.454-0.557)	0.841 (0.803-0.879)	0.542 (0.491-0.594)	0.850 (0.814-0.887)
External test set	0.907 (0.850–0.963)	0.902 (0.844-0.960)	0.918 (0.864-0.971)	0.824 (0.750-0.898)	0.963 (0.926-1.000)	0.667 (0.575-0.758)	0.940 (0.894-0.986)
Prospective cohort	0.768 (0.694–0.842)	0.677 (0.595-0.760)	0.750 (0.674-0.826)	0.609 (0.523-0.695)	0.643 (0.559-0.727)	0.722 (0.643-0.801)	0.692 (0.611-0.774)

Notes: AUC, area under the receiver operating characteristic curve; ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; 95% CI, 95% confidence interval. Threshold-dependent metrics (accuracy, sensitivity, specificity, PPV, NPV, and F1 score) are reported at dataset-specific operating points. For the internal and external test sets, this was the Youden-derived threshold determined separately within each dataset. For the prospective cohort, this was the 85%-sensitivity threshold, derived in the internal test set and locked before recruitment.

(48.39%). In the Stage 1 network, REMoCA achieved an AUC of 0.980 (0.973–0.987), a sensitivity of 0.924 (95% CI, 0.912–0.937), and a specificity of 0.958 (0.948–0.968) for AST trial classification (Figure 3c). In the Stage 2 network, REMoCA achieved an AUC of 0.768 (0.694–0.842), a sensitivity of 0.750 (95% CI, 0.674–0.826), and a specificity of 0.609 (95% CI, 0.523–0.695) for the real-time screening of MoCA-defined CI: this AUC was lower than that observed in the external test set (Figure 3f and Table 2). Confusion matrices for the prospective cohort are presented in Supplementary Figure 4.

Discussion

In this multicenter study, participants with MoCA-defined CI exhibited higher EM error rates across VA strata, although the magnitude of the between-group difference varied among strata. The AST discriminated most consistently between the MoCA-defined CI and non-CI groups across the VA levels examined. Therefore, we developed REMoCA as an AST-centered framework for detecting MoCA-defined CI and evaluated it in an external test set and a prospective community-based proof-of-concept cohort. Together, these findings identify the AST as a VA-tolerant EM measure for MoCA-defined CI and demonstrate its preliminary feasibility in a proof-of-concept setting, providing a basis for further development and evaluation.

Across VA strata, participants with MoCA-defined CI generally showed higher EM error rates than those without CI, suggesting a contribution of cognitive status to EM performance. This is consistent with the established role of higher-order cognitive control in EM tasks, particularly executive and inhibitory processes.[12,24] The between-group separation was most pronounced in the high-VA stratum and attenuated, but remained significant for the AST, in the lower-VA strata. This attenuation at poorer acuity more likely reflects the smaller and clinically more heterogeneous sample in the low-VA stratum (n = 98 in the MoCA-defined CI group and n = 41 in the non-CI group; Supplementary Table 4) than a direct effect of visual impairment on EM performance. Among the five paradigms, only the AST retained significant discriminative ability across all three VA strata.

AST performance depends on top-down inhibitory control, working memory, rule maintenance, and sufficient visual input to detect the target,[25] a profile that favors cognitive-driven over acuity-driven variation. Whereas discrimination by the other paradigms attenuated at reduced VA, the AST

retained significant discrimination across all near-VA strata. Together, this cognitively driven profile and the consistent cross-stratum discrimination of AST motivated its selection as the core REMoCA task, supporting its use as a preferred basis for rapid cognitive screening in older adults. Notably, REMoCA achieves this VA-tolerance through the selection of a task whose discriminative performance is intrinsically robust across acuity levels, rather than through explicit statistical adjustment for VA within the model itself. Therefore, near VA was not included as a model predictor. This design choice reflects the intended screening context: near VA assessment requires a trained examiner.[26] Based on our experience administering it in this study, near VA testing also requires a cooperative, sustained response from the participant. This demand can be particularly difficult to sustain among individuals with potential CI, the very population this screening tool is intended to identify, and can therefore add appreciably to assessment time. Requiring a VA measurement as a model input would reintroduce the accessibility and time burden that REMoCA is intended to obviate, particularly in settings where trained examiners or additional testing time are not readily available.

Previous studies have investigated EM-based features, such as latency and saccadic velocity, as potential markers of CI.[27-28] These studies established EMs as informative behavioral markers, typically by deriving predefined summary metrics, often in focused or disease-specific cohorts.[29] Building on this foundation, our approach processes EM-derived video sequences with a pseudo-3D network, preserving spatial and temporal information across each task. Because it operates directly on EM-trajectory videos rather than a predefined set of summary metrics, this approach preserves richer spatial and temporal detail and yields interpretable saliency heatmaps that make the model-relevant EM patterns directly visualizable, offering a more comprehensive characterization of dynamic EM behavior.

Retinal imaging has been investigated as an ocular approach for cognitive screening, but retinal biomarkers primarily capture structural and vascular features of the eye.[30-31] EM assessment is complementary in focus: rather than static structure, it probes visually elicited behavior that is actively shaped by cognitive control, thereby capturing a functional dimension of visual processing that complements these structural markers. REMoCA operationalizes this functional perspective through a two-stage network that separates automated trial-level classification from participant-level detection of MoCA-defined CI. Saliency maps allow visual inspection of model-attended regions, and DA summarizes predictor contributions. Building on this design, the prospective community cohort assessed whether a shortened AST-centered workflow could be completed by trained non-specialist personnel within three minutes.[32] Its lower AUC than the external test set indicates both reduced transportability and the lower MoCA-defined CI prevalence in this cohort, and supports interpreting this cohort as a proof-of-concept feasibility evaluation. Larger and more representative community studies, improved calibration, and explicit threshold evaluation are required before REMoCA can be considered for triage, referral, or population-level deployment.

This study has several limitations. First, the model detects MoCA-defined CI (defined by MoCA <26) rather than a clinical diagnosis of mild cognitive impairment or dementia. Although it is widely used for cognitive screening, MoCA is not a diagnostic reference standard. It is influenced by education, culture, visual input, motor demands, and comprehension of instructions. Nonetheless, its high sensitivity for detecting cognitive impairment makes MoCA-defined CI an appropriate target for a screening-oriented tool such as REMoCA, which is intended to identify individuals for further assessment rather than to establish a diagnosis.[7] Second, near VA was the only visual-function measure available. Because the EM targets were high-contrast stimuli presented in black and white at a near viewing distance, near VA was considered the visual-function measure most directly relevant to resolving the targets in this paradigm. Assessing additional dimensions, including contrast sensitivity, visual fields, glare, ocular alignment, ocular disease severity, and oculomotor control, would have been time-consuming

and costly and was not feasible in the present study. Our findings therefore relate specifically to near VA and remain to be extended to other dimensions of visual function. Third, the development dataset of REMoCA combined hospital-based and screening-program participants with heterogeneous disease spectra. This heterogeneity, together with the high prevalence of MoCA <26 (75.86%), may introduce spectrum bias and limit transportability to lower-prevalence community settings. Fourth, the use of a dichotomized MoCA total score also provides a coarser characterization of cognition than domain-level assessment, and the relationship between antisaccadic performance and the individual MoCA domains remains to be examined. Finally, the WML/Fazekas analysis was exploratory, restricted to participants with MoCA <26 who underwent MRI examination. It cannot be generalized to the entire cohort or interpreted as a WML diagnostic model. The extensive subgroup analyses were likewise exploratory, given the limited number of participants available within subgroups. In the future, representative cohorts with comprehensive visual and neuropsychological assessment, external validation in fully independent settings, and explicit evaluation of calibration and decision thresholds are needed before clinical deployment.

In conclusion, among the five EM paradigms examined, the AST most consistently discriminated between the MoCA-defined CI and non-CI groups, retaining significant separation across the VA strata examined. Building on this finding, we developed REMoCA, an AST-centered framework that detected MoCA-defined CI in an external test set and remained feasible in a prospective community-based proof-of-concept cohort. Together, these findings position rapid, AST-based assessment as a candidate, VA-tolerant basis for advancing cognitive screening in older adults, pending validation in larger, independent cohorts.

Acknowledgments

The authors would like to express their sincere thanks to Shanghai Qingyan Technology Co., Ltd. for providing the eye trackers used in this research, and to the National Supercomputer Center in Guangzhou for making high-performance computational resources available.

Author contributions

(I) Conception and design: Haotian Lin, Yamei Tang, Xun Wang, Shuyi Zhang, Zhenzhe Lin, Lanqin Zhao
(II) Administrative support: Haotian Lin, Yamei Tang, Dongni Wang, Xulin Zhang, Meimei Dongye, Yi Li
(III) Provision of study materials or patients: Shuyi Zhang, Jia Wang, Yanting Chen, Dong Zheng
(IV) Data Collection/Aggregation: Jia Wang, Yanting Chen, Xueer Tu, Yunjian Huang, Xiaoming Zhou, Nannan Pan, Yayong Cui, Junhua Xu
(V) Data analysis and interpretation: Xun Wang, Shuyi Zhang, Zhenzhe Lin, Carol Yim-Lui Cheung, Wei Wang, Lanqin Zhao, Duoru Lin, Xiaohang Wu, Yahan Yang, Dongyuan Yun, Zhenzhen Liu, Jason C. Yam
(VI) Manuscript writing: All authors
(VII) Final approval of manuscript: All authors

Fundings

This study was funded by the National Supercomputer Center in Guangzhou for making high-performance computational resources available. This study was funded by the National Natural Science Foundation of China (82441003), the Basic Scientific Research Projects of Sun Yat-sen University (23ykcxqt002), the Science and Technology Planning Project of Guangdong Province (2025A1515012870), the Guangdong Basic and Applied Basic Research Foundation (2024A1515010801), the Guangdong Basic Research Center of Excellence for Major Blinding Eye Diseases Prevention and Treatment, the UK National Institute of Health Research (NIHR301696), the Rosetrees Trust (PGL23/100048), the Moorfields Eye Charity (GR001376), the NIHR Cambridge Biomedical Research Centre (NIHR203312), and the NIHR Biomedical Research Centre based at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology (NIHR203322).

Conflict of interests

The authors disclose that the corresponding author, Haotian Lin, serves as Editor-in-Chief of *Eye Science*.

To avoid any potential conflict of interest, the editorial handling of this manuscript was assigned to an independent editor, and the authors were not involved in the peer-review or decision-making process. All editorial procedures were conducted in accordance with COPE guidelines and the journal's standard policies. The authors declare no other competing financial or non-financial interests relevant to this work.

Patient consent for publication

Written informed consent was obtained from all participants or their legal guardians in accordance with ethical guidelines.

Ethics approval and consent to participate

This study adhered to the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of Zhongshan Ophthalmic Center, Sun Yat-sen University(Approval No. 2020KYPJ015). It was prospectively registered with the Clinical Research Internal Management System of Zhongshan Ophthalmic Center and with ClinicalTrials.gov (NCT04236375).

Data availability statement

Due to patient privacy and ethical restrictions, the raw clinical data cannot be shared publicly. The de-identified demo data can also be requested for nonprofit academic research by the corresponding author (HL; linht5@mail.sysu.edu.cn), following review by the data access committee at the State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Centre. The source code for our model is available on GitHub (https://github.com/huapu4/REMoCA).

Supplementary materials

The supplementary materials are available online at: https://journal.gzzoc.org.cn/es/article/6324.

Declaration of generative AI use

In the preparation of this manuscript, the authors used ChatGPT (GPT-5.5) for English language polishing. After using this tool, the authors reviewed and edited the content as necessary and take full responsibility for the content of this publication.

References

1. Cortes N, Ladret HJ, Abbas-Farishta R, et al. The pulvinar as a hub of visual processing and cortical integration. Trends Neurosci. 2024, 47(2): 120-134. DOI: 10.1016/j.tins.2023.11.008.
2. Kay K, Bonnen K, Denison RN, et al. Tasks and their role in visual neuroscience. Neuron. 2023, 111(11): 1697-1713. DOI: 10.1016/j.neuron.2023.03.022.
3. Gilbert CD, Li W. Top-down influences on visual processing. Nat Rev Neurosci. 2013, 14(5): 350-363. DOI: 10.1038/nrn3476.
4. Livingston G, Huntley J, Liu KY, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. Lancet. 2024, 404(10452): 572-628. DOI: 10.1016/s0140-6736(24)01296-0.
5. Lim ZW, Chee ML, Da Soh Z, et al. Association between visual impairment and decline in cognitive function in a multiethnic Asian population. JAMA Netw Open. 2020, 3(4): e203560. DOI: 10.1001/jamanetworkopen.2020.3560.
6. Bai W, Chen P, Cai H, et al. Worldwide prevalence of mild cognitive impairment among community dwellers aged 50 years and older: a meta-analysis and systematic review of epidemiology studies. Age Ageing. 2022, 51(8): afac173. DOI: 10.1093/ageing/afac173.
7. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005, 53(4): 695-699. DOI: 10.1111/j.1532-5415.2005.53221.x.
8. Li L, Zhang B, Yang Y, et al. Diagnostic performance of self-administered unsupervised computerized cognitive tests for the identification of mild cognitive impairment (MCI) and dementia: a systematic review and meta-analysis. Aging Ment Health. 2025, 29(8): 1439-1455. DOI: 10.1080/13607863.2025.2495794.
9. Roseborough AD, Saad L, Goodman M, et al. White matter hyperintensities and longitudinal cognitive decline in cognitively normal populations and across diagnostic categories: a meta-analysis, systematic review, and recommendations for future study harmonization. Alzheimers Dement. 2023, 19(1): 194-207. DOI: 10.1002/alz.12642.
10. Verón GL, Juantorena GE, Keller G, et al. Eye tracking as a diagnostic tool in Alzheimer's disease, mild cognitive impairment, and related dementias: a systematic review. Alzheimers Dement Diagn Assess Dis Monit. 2026, 18(1): e70238. DOI: 10.1002/dad2.70238.

11. Fan H, Feng J, Ding Y, et al. Performance of antisaccades in patients with cerebral small vessel disease accompanied by white matter hyperintensities. Neurol Res. 2024, 46(10): 917-924. DOI: 10.1080/01616412.2024.2367934.
12. Tao M, Cui L, Du Y, et al. Analysis of eye movement features in patients with Alzheimer's disease based on intelligent eye movement analysis and evaluation system. J Alzheimers Dis. 2024, 102(4): 1249-1259. DOI: 10.1177/13872877241297416.
13. Eraslan Boz H, Koçoğlu K, Akkoyun M, et al. Uncorrected errors and correct saccades in the antisaccade task distinguish between early-stage Alzheimer's disease dementia, amnestic mild cognitive impairment, and normal aging. Aging Neuropsychol Cogn. 2024, 31(3): 457-478. DOI: 10.1080/13825585.2023.2198191.
14. Fazekas F, Chawluk JB, Alavi A, et al. MR signal abnormalities at 1. 5 T in Alzheimer's dementia and normal aging. Am J Roentgenol. 1987, 149(2): 351-356. DOI: 10.2214/ajr.149.2.351.
15. Muñoz Maniega S, Chappell FM, Valdés Hernández MC, et al. Integrity of normal-appearing white matter: Influence of age, visible lesion burden and hypertension in patients with small-vessel disease. J Cereb Blood Flow Metab. 2017, 37(2): 644-656. DOI: 10.1177/0271678X16635657.
16. Qiu Z, Yao T, Mei T. Learning spatio-temporal representation with pseudo-3D residual networks [EB/OL]. 2017: arXiv: 1711.10305. https://arxiv.org/abs/1711.10305.
17. Muhammad MB, Yeasin M. Eigen-CAM: class activation map using principal components//2020 International Joint Conference on Neural Networks (IJCNN). July 19-24, 2020. Glasgow, United Kingdom. IEEE, 2020: 1-7. DOI: 10.1109/ijcnn48605.2020.9206626.
18. Xue J, Li J, Liang J, Chen S. The Prevalence of Mild Cognitive Impairment in China: A Systematic Review. Aging Dis. 2018 Aug 1;9(4): 706-715. DOI: 10.14336/AD.2017.0928.
19. Mattsson N. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. Jama. 2009, 302(4): 385. DOI: 10.1001/jama.2009.1064.
20. Abughofah Y, Risacher SL, Unverzagt FW, et al. The Cognitive Change Index in the Alzheimer's Disease Research Center setting: Self- and informant-ratings for perceived cognitive decline. Alzheimers Dement. 2025, 21(10): e70754. DOI: 10.1002/alz.70754.
21. Mallett S, Halligan S, Thompson M, et al. Interpreting diagnostic accuracy studies for patient care. Bmj. 2012, 345(jul02 1): e3999. DOI: 10.1136/bmj.e3999.
22. Killeen OJ, De Lott LB, Zhou Y, et al. Population prevalence of vision impairment in US adults 71 years and older: the national health and aging trends study. JAMA Ophthalmol. 2023, 141(2): 197-204. DOI: 10.1001/jamaophthalmol.2022.5840.
23. World Health Organization. ICD-11 for mortality and morbidity statistics. World Health Organization, 2024. https://icd.who.int/browse/2025-01/mms/en#1103667651.
24. Lavie N, Tsal Y. Perceptual load as a major determinant of the locus of selection in visual attention. Percept Psychophys. 1994, 56(2): 183-197. DOI: 10.3758/BF03213897.
25. Anderson TJ, MacAskill MR. Eye movements in patients with neurodegenerative disorders. Nat Rev Neurol. 2013, 9(2): 74-85. DOI: 10.1038/nrneurol.2012.273.
26. Burton MJ, Ramke J, Marques AP, et al. The lancet global health commission on global eye health: vision beyond 2020. Lancet Glob Health. 2021, 9(4): e489-e551. DOI: 10.1016/s2214-109x(20)30488-5.
27. Chehrehnegar N, Nejati V, Shati M, et al. Behavioral and cognitive markers of mild cognitive impairment: diagnostic value of saccadic eye movements and Simon task. Aging Clin Exp Res. 2019, 31(11): 1591-1600. DOI: 10.1007/s40520-019-01121-w.
28. Russell LL, Greaves CV, Convery RS, et al. Novel instructionless eye tracking tasks identify emotion recognition deficits in frontotemporal dementia. Alzheimers Res Ther. 2021, 13(1): 39. DOI: 10.1186/s13195-021-00775-x.
29. Zhang X, Zhang F, Hou S, et al. Effectiveness of digital screening tools in detecting cognitive impairment among community-dwelling elderly in Northern China: a large cohort study. J Prev Alzheimers Dis. 2025, 12(3): 100080. DOI: 10.1016/j.tjpad.2025.100080.
30. Hao J, Kwapong WR, Shen T, et al. Early detection of dementia through retinal imaging and trustworthy AI. npj Digit Med. 2024, 7: 294. DOI: 10.1038/s41746-024-01292-5.
31. Cheung CY, Ran AR, Wang S, et al. A deep learning model for detection of Alzheimer's disease based on

retinal photographs: a retrospective, multicentre case-control study. Lancet Digit Health. 2022, 4(11): e806-e815. DOI: 10.1016/S2589-7500(22)00169-8.

32. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022, 28(5): 924-933. DOI: 10.1038/s41591-022-01772-9.