



Diagnostic Potential of Eye Movements in Alzheimer's Disease via a Multiclass Machine Learning Model

Jiaqi Song¹ · Haodong Huang^{2,5} · Jiarui Liu³ · Jiani Wu³ · Yingxi Chen¹ · Lisong Wang³ · Fuxin Zhong³ · Xiaoqin Wang³ · Zihan Lin⁴ · Mengyu Yan¹ · Wenbo Zhang³ · Xintong Liu³ · Xinyi Tang¹ · Yang Lü³ · Weihua Yu¹

Received: 10 April 2024 / Accepted: 12 August 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Early diagnosis plays a crucial role in controlling Alzheimer's disease (AD) progression and delaying cognitive decline. Traditional diagnostic tools present great challenges to clinical practice due to their invasiveness, high cost, and time-consuming administration. This study was designed to construct a non-invasive and cost-effective classification model based on eye movement parameters to distinguish dementia due to AD (ADD), mild cognitive impairment (MCI), and normal cognition. Eye movement data were collected from 258 subjects, comprising 111 patients with ADD, 81 patients with MCI, and 66 individuals with normal cognition. The fixation, smooth pursuit, prosaccade, and anti-saccade tasks were performed. Machine learning methods were used to screen eye movement parameters and build diagnostic models. Pearson's correlation analysis was used to assess the correlations between the five most important eye movement indicators in the optimal model and neuropsychological scales. The gradient boosting classifier model demonstrated the best classification performance, achieving 68.2% of accuracy and 66.32% of F1-score in multiclass classification of AD. Moreover, the correlation analysis indicated that the eye movement parameters were associated with various cognitive functions, including general cognitive status, attention, visuospatial ability, episodic memory, short-term memory, and language and instrumental activities of daily life. Eye movement parameters in conjunction with machine learning methods achieve satisfactory overall accuracy, making it an effective and less time-consuming method to assist clinical diagnosis of AD.

Keywords Alzheimer's disease · Mild cognitive impairment · Eye movement · Machine learning model · Multiclass diagnosis

Jiaqi Song and Haodong Huang contributed equally to this study.

✉ Yang Lü
yanglyu@hospital.cqmu.edu.cn

✉ Weihua Yu
yuweihua@cqmu.edu.cn

¹ Institute of Neuroscience, Chongqing Medical University, 1 Yixueyuan Road, Yuzhong District, Chongqing 400016, China

² Department of Obstetrics and Gynecology, Women and Children's Hospital of Chongqing Medical University, Chongqing 401147, China

³ Department of Geriatrics, The First Affiliated Hospital of Chongqing Medical University, 1 Youyi Road, Yuzhong District, Chongqing 400016, China

⁴ International Medical College, Chongqing Medical University, Chongqing 400016, China

⁵ Chongqing Municipal Health Commission Key Laboratory of Perinatal Medicine, Chongqing, 400016, China

Introduction

Approximately 57.4 million people were living with dementia worldwide in 2019. What is more, population aging has been having an enormous impact on the increasing prevalence of dementia. By the year 2050, the worldwide prevalence of dementia will grow to more than double, to 152.8 million, according to the forecast [1]. Dementia is generally understood as an acquired loss of cognitive ability caused by brain disease or injury that is sufficiently serious to interfere with functional activities of daily living [2]. Dementia due to Alzheimer's disease (ADD) is the most common form of dementia, accounting for about 60% of all dementia cases [3]. Between ADD and normal cognition, there can be a prodromal, intermediate stage called mild cognitive impairment (MCI). MCI is marked by lower performance in one or more cognitive domains than the person's previous level, although this does not fully interfere with independence of function

in daily life [4]. Based on statistical data, the prevalence of MCI is two times higher than dementia, with an estimated annual conversion rate of 10–15% towards dementia [5]. The rising prevalence of MCI and ADD results in a substantial economic burden on both individuals and society [6].

Early detection and diagnosis of ADD could motivate patients and caregivers to participate in early and effective therapeutic intervention, potentially improving quality of life and reducing the death rate of ADD [7]. However, the diagnosis of ADD and MCI has always been challenging in the clinic. At present, positron emission tomography (PET) targeting amyloid plaque deposition and tau tangle is the only approved test for Alzheimer's disease (AD) diagnosis by FDA [8, 9]. Nevertheless, these scans are costly. Neuropsychological assessment is most widely used to obtain the overall level of cognitive functioning and screen for cognitive impairment [10, 11]. However, the use of comprehensive neuropsychological assessment in clinical practice still faces many obstacles, including time-consuming and requiring a trained and qualified professional, which limits its application in clinical practice [12]. Furthermore, most screening tools based on language are affected by educational level and age [13]. Thus, the discovery of a reliable, noninvasive, and affordable marker would be an urgent need for subsidiary diagnosis of cognitive impairment in routine clinical work.

Eye movement, a sensitive, low-cost, non-invasive, and portable marker, has emerged as a potential tool for detecting cognitive change or deterioration [14, 15]. Eye movements are triggered and regulated by an intricate neural network involving numerous cortical and subcortical regions, which develop specific pathological changes beginning decades before clinical symptoms appear in patients with AD. Hence, eye movement has the potential for the timely identification of subtle cognitive deterioration [16]. Abnormalities in eye movements have been repeatedly observed in patients with AD. For example, patients with ADD have longer prosaccade and anti-saccade latencies and more anti-saccade errors than cognitively normal older adults [17, 18]. A critical issue then is whether eye movement has the potential to diagnose AD.

Several studies aim to distinguish patients with MCI and AD from normal cognition or distinguish AD from MCI using eye movements [19–21]. These studies achieve diagnostic results with the area under the receiver operating characteristic curve (AUC) reaching 0.752. Nonetheless, for effective clinical application, eye movement must be able to simultaneously distinguish ADD, MCI, and normal cognition, which needs further exploration. Moreover, the relationship between eye movements and specific cognitive subdomains is still unclear even prosaccade latency, anti-saccade errors, and saccadic intrusions have been found to correlate with Mini-mental State Examination (MMSE)

[22]. Therefore, the critical issue of how eye movements can diagnose AD remains inadequately addressed.

In this study, we collected demographic information, cognitive scores, and eye movement metrics from participants who completed the fixation, smooth pursuit, prosaccade, and anti-saccade tasks and screened the key eye movement markers that are capable of distinguishing ADD, MCI, and normal cognition. In addition, machine learning models were constructed to explore the diagnostic potential of eye movements. As a secondary objective, we aimed to determine the correlation of eye movement parameters with various specific cognitive subdomains through neuropsychological scales.

The main findings of our study are as follows: Firstly, 16 key eye movement features were identified for diagnosing cognitive impairment, including saccadic latency, error rate, and number of forward saccades. Secondly, 17 three-classification diagnostic models for ADD, MCI, and normal cognition were established, with the GBC model demonstrating superior performance, achieving an accuracy of 68.2% and an F1-score of 66.32%. Finally, the novel association between eye movement parameters in prosaccade and anti-saccade tasks and Free and Cued Selective Reminding Test (FCSRT)-immediate total recall, instrumental activities of daily life (IADL), and Boston Naming Test (BNT)-total score was observed.

Related Work

Jessica et al. summarized the researches that used eye-tracking technology and computational analysis to measure and compare eye movements in participants with different cognitive statuses. The authors introduced three tasks that are more complex than saccadic eye movement tasks, encompassing reading task, visual exploration task, and free viewing video task. At the same time, it is proposed that currently eye trackers have only been used in controlled laboratory environments, while by using computational technology to analyze recorded videos, eye tracking can be applied to a wider range of scenarios [23].

Opwonya et al. utilized eye movement parameters to generate logistic regression, random forest classifier, support vector machines, and extreme gradient boosting classification models to predict cognitive status, and logistic regression algorithm exhibited the highest AUC of 0.715 in normal cognition vs MCI classification [19]. Logistic regression algorithm was also implemented by Jiang et al. for selecting discriminatory features and classifying different models, achieving discrimination with an AUC of 0.807 in normal cognition vs MCI classification [20].

Liu et al. proposed a novel deep learning-based approach called Depth-induce Integrated Comparison serial attention

Network (DISCN). The multi-image integration module was used to integrate visual stimuli and eye movement heatmap into visual saliency map, which were then sent to serial attention module for diagnosis. The DISCN achieved an AUC difference of 0.81 in normal cognition vs AD classification [24].

The aforementioned studies achieved a high rate of accuracy in the binary classification due to clear distinctions between the two groups. However, tackling the three classifications of AD diagnosis proves to be more practical but formidable compared to binary classification [25, 26]. Currently, only a few diagnostic models have achieved an accuracy of over 60% in the three classifications of AD [27]. Tong et al. utilized a nonlinear graph fusion approach in the three-classification diagnosis of AD, achieving an accuracy of 60.2% [28]. Lin et al. achieved a three-classification diagnosis accuracy of 66.7% by using decision tree algorithm with multimodal data of magnetic resonance imaging (MRI), PET, cerebrospinal fluid markers, and genetic features [29]. In another study, the multiclass diagnosis of AD was performed using electroencephalography (achieved accuracies of 65.52%) and functional near-infrared spectroscopy (achieved accuracies of 58.62%) [30].

Methods

Participants and Diagnosis Procedures

In this study, 258 subjects were recruited from Memory Clinic, Department of Geriatrics, The First Affiliated Hospital of Chongqing Medical University and the surrounding community, including 111 patients with ADD, 81 patients with MCI, and 66 people with normal cognition. The detailed demographic information of subjects was provided in Table 1, and the flowchart is shown in Fig. 1, Step I. The diagnosis was performed by experienced geriatric psychiatrists according to the results of clinical evaluation and cognitive assessment. AD diagnosis was determined using the National Institute of Aging and the Alzheimer's

Disease Society for clinically probable AD in 2011[31]. MCI diagnosis was made according to MCI International Working Group consensus criteria [4]. Inclusion criteria included being 60 years or older, participating in the study voluntarily, and having sufficient hearing and vision to complete the neuropsychological test and eye-movement calibration. Patients with a history of alcohol or drug abuse, severe psychiatric illness, intracranial tumor, or stroke were excluded. The protocol was approved by the Medical Ethics Committee of The First Affiliated Hospital of Chongqing Medical University (approval number: 20212901; time of ethics approval: 10 May 2021). All participants or their legal guardians signed the informed consent after receiving a detailed explanation of the study.

Neuropsychological Evaluation

Neuropsychological batteries were used to assess general cognitive status, attention, executive function, language, visuospatial skills, and memory. General cognitive status was evaluated by MMSE and Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) [32, 33], attention was assessed by Trail Making Test-A (TMT-A) and Digit Span Test (DST) [34], executive function was assessed by TMT-B [35], language was assessed by BNT [36], visuospatial skills were assessed by Clock Drawing Test (CDT) [37], memory was assessed by FCSRT and Auditory Verbal Learning Test (AVLT) [38, 39], the abilities of daily life were assessed by physical self-maintenance scale (PSMS) and IADL scale [40].

Assessment of Eye Movement

Eye movement parameters were recorded during four visual tasks, including direct gaze toward the target (fixation task), smooth pursuit tasks, looking toward a jumping target (prosaccade task), and looking away from a jumping target (the anti-saccade task). A desktop-mounted eye-tracker (Tobii4L model, Ji Zhi Medical Technology Co. Ltd, China) monitored eye movements using the 250 Hz pupil-corneal

Table 1 Participants' demographic characteristics

Variable	Total <i>N</i> = 258	NC <i>N</i> = 66	MCI <i>N</i> = 81	ADD <i>N</i> = 111	<i>P</i> -value
Age	73.22(8.64)	69.50(15.00)	72.00(11.00)	76.00(9.00)	0.001 ^{a*}
Female, <i>N</i> (%)	164(63.60)	49(74.20)	49(60.50)	66(59.50)	0.064 ^b
Education, (y)	9.28(4.25)	10.50(4.00)	9.00(5.00)	9.00(7.00)	0.076 ^{a*}
MMSE	21.00(9.00)	29.00(3.00)	25.00(6.00)	17.00(8.00)	< 0.001 ^{a*}
ADAS-Cog	16.33(15.49)	6.69(2.22)	13.40(9.37)	22.85(14.69)	< 0.001 ^{a*}

Data are presented as median (interquartile range); *P*^a, Kruskal–Wallis test; *P*^b, the chi-square test. Abbreviation: *NC* normal cognition, *MCI* mild cognitive impairment, *ADD* dementia due to Alzheimer's disease, *MMSE* Mini-mental State Examination, *ADAS-Cog* Alzheimer's Disease Assessment Scale-Cognitive section

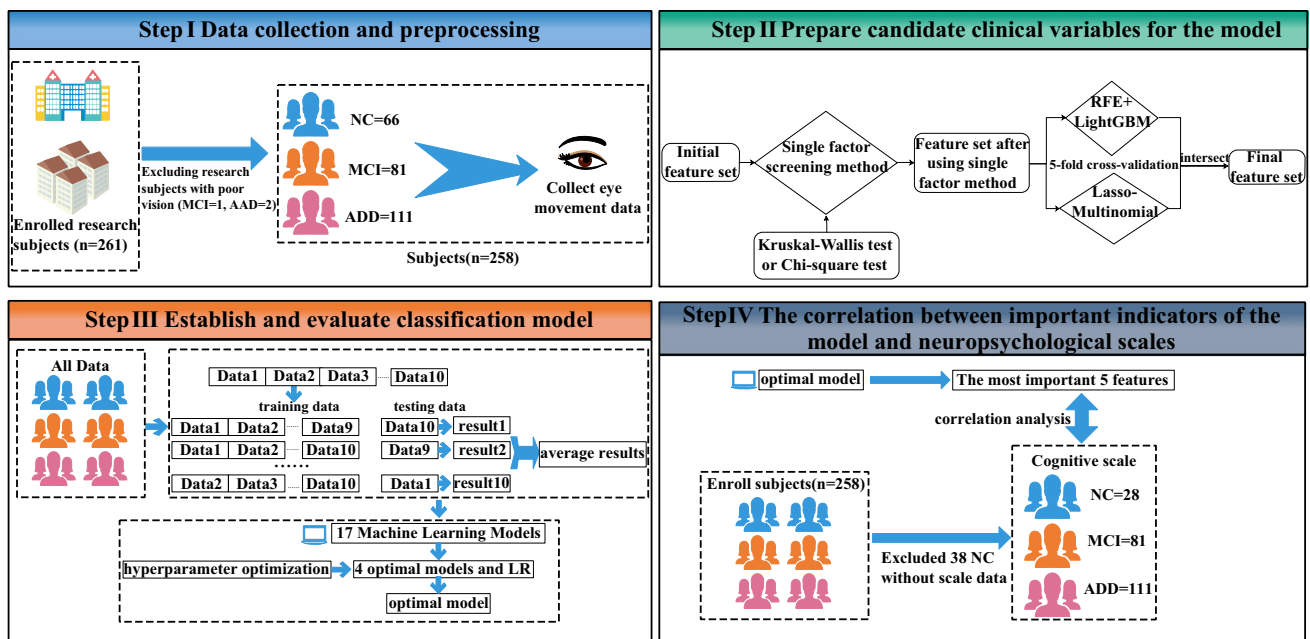


Fig. 1 The flowchart of the proposed model, which was composed of four steps: data collection and preprocessing; feature selection; establishment and evaluation of classification model; the correlation between important indicators of the model and neuropsychological scales

reflex mode. Tasks were displayed on a 15.6-inch LED monitor and a screen resolution of 1920 by 1080 pixels. All subjects were seated approximately 60 cm from the screen; their heads were properly fixed on the support frame and their eye position was horizontally aligned with the center of the screen. The sign of proper head fixation was that two bright circles, around 0.5–1 cm in size, appeared in the center of the screen. The instrument tracked eye movements in horizontal ($\pm 30^\circ$) and vertical ($\pm 30^\circ$) positions. The testing was performed in a separate, quiet room without strong light stimulation. The built-in software, Tobii Studio, was used to collect and analyze the eye-movement data.

Candidate Feature Selection

Univariate analysis was carried out for eye movement parameters and general information. Specifically, Kruskal–Wallis tests were used to compare continuous variables, and two-sided chi-squared test was used to compare categorical variables. Next, the meaningful variables detected by univariate analysis were subjected to further screening. Light gradient boosting machine-based recursive feature elimination (LightGBM-RFE) and lasso-multinomial algorithms in conjunction with a five-fold cross-validation approach were employed to obtain the two optimum feature sets. Finally, the intersection of two optimum feature sets was used to build the machine learning model. The flowchart is shown in Fig. 1, Step II.

Establishment and Validation of Machine Learning Models

The diagnostic model for classifying ADD, MCI and normal cognition was constructed based on the filtered feature set using machine learning algorithms including gradient boosting classifier (GBC) [41], light GBM [42], random forest classifier [43], extra trees classifier [44], naïve Bayes [45], logistic regression [46], ridge classifier [47], linear discriminant analysis [48], decision tree classifier [49], quadratic discriminant analysis [50], ada boost classifier [51], support vector machines-linear kernel-linear kernel [52], K neighbors classifier [53], and dummy classifier [54]. These algorithms were implemented by using PyCaret package (version 2.3.3). The leave-one-out method was used for ten-fold cross-validation of the model and the evaluation metric included accuracy, macro_Sensitivity, macro_Specificity, the AUC and macro_f1. The calculation formula is as follows: $\text{Accuracy} = (\text{TP} + \text{TN})/N$; $\text{macro_Sensitivity} = (\text{Sensitivity_class1} + \text{Sensitivity_class2} + \text{Sensitivity_class3})/N$; $\text{macro_Specificity} = (\text{Specificity_class1} + \text{Specificity_class2} + \text{Specificity_class3})/N$; $\text{macro_F1} = (\text{F1_class1} + \text{F1_class2} + \text{F1_class3})/N$. The optimal model with the highest macro_f1 value was used as the final model. The hyperparameters of the model and the tuning range for hyperparameters can be found in Supplementary Table 1. The flowchart is shown in Fig. 1, Step III.

Fully Connected Neural Network (FCNN) is a fundamental deep learning model that utilizes a hierarchical structure

to map input data to output results [55]. FCNN consists of multiple layers, including an input layer, hidden layers, and an output layer. Input layer accepts an input feature vector with a dimension of input_dim. First hidden layer contains 64 neurons. The activation function is ReLU (rectified linear unit), which introduces non-linearity. Second hidden layer contains 32 neurons. The activation function is ReLU. Output layer contains 3 neurons corresponding to 3 categories. The activation function is Softmax, used to output the probability for each category. For model compilation, the optimizer is adaptive moment estimation (Adam). The loss function is categorical cross-entropy, suitable for multi-class problems.

Feature Transformation-Transformer (FT-Transformer) is the latest advancement in deep learning, building on the success of other well-established architectures such as convolutional neural networks and recurrent neural networks [56]. FT-Transformer includes a linear embedding layer, two transformer encoder layers, a fully connected layer, and a dropout layer. The parameters for the transformer encoder layers are as follows: the input feature dimension is 16, the embedding dimension is 32, each layer employs 4 parallel attention mechanisms, and the dropout rate is 0.1, meaning that 10% of neurons are randomly “dropped out” during each forward pass. The activation function is ReLU, and the training is conducted over 100 epochs. The Adam optimizer is used for model optimization. Finally, the model’s performance is evaluated using tenfold cross-validation.

Self-attention model [57] consists of an embedding layer, self-attention layer, average pooling, dropout layer, and fully connected layer. The embedding layer elevates the dimensionality, transforming the input feature dimension from 16 to an embedding dimension of 32, thereby enhancing the model’s capacity to capture information. The self-attention layer encompasses the computation of attention scores, application of softmax, and weighted summation. Specifically, attention scores are calculated by determining the similarity between queries and keys through a dot product, normalized by dividing by the square root of the vector dimension. The softmax function is utilized to convert attention scores into attention weights, and the weighted summation involves multiplying these attention weights by the corresponding values to produce the final output. The dropout rate is set to 0.1, and the model is trained for 100 epochs. The Adam optimizer is used for model optimization.

Analysis of Key Features

The top five important features were extracted in the best-performing machine learning model. Pearson’s correlation analysis was adopted to calculate the correlation between the top five important features and cognitive scales. Group comparisons as well as correlations have been performed

for the three most relevant scales. The flowchart is shown in Fig. 1, Step IV.

Statistical Analysis Environment

All statistical analysis and computations were done using R version 4.2.2 and Python Version 3.6.2. Categorical variables were expressed as frequency (percentage). The continuous variables conforming to the normal distribution were expressed as the mean \pm standard deviation and the other not conforming to normal distribution were presented as median (interquartile range). All differences with $p < 0.05$ were considered statistically significant. Plots were generated with the packages Matplotlib and ggplot2.

Results

Eye Movement Characteristics in Different Severities of Cognitive Impairment

A total of 38 eye movement features were acquired using the built-in software of an eye tracker. Subsequently, 25 were statistically significant among ADD, MCI, and normal cognition by the univariate analysis. Detailed data is shown in Table 2.

Selection of Diagnostic Markers via Lasso and LightGBM-RFE

Two distinct algorithms (Lasso and LightGBM-RFE) were used for selecting feature sets screened from the meaningful variables identified through univariate analysis. For the Lasso algorithm, feature set was selected based on the minimum mean cross-validated error of the lambda value and a set of 21 parameters was selected (Fig. 2A, B). For LightGBM-RFE algorithm, the feature set was selected based on the highest macro_f1 value, and the classifier produced the minimum error when the feature number was 20 (Fig. 2C). Overall, 16 features shared between the Lasso and LightGBM-RFE algorithms as diagnostic markers for cognitive impairment were finally selected for further analysis (Fig. 2D). The optimal feature set encompassed features from four tasks, especially the prosaccade and anti-saccade tasks. The detailed features information is found in Table 3.

Establishing Machine Learning Model for Diagnosing ADD, MCI and Normal Cognition

After feature selection, the machine models were trained using Pycaret, and Supplementary Table 2 describes the performance of these models. The models’ performance was evaluated using AUC, precision, recall, accuracy and

Table 2 Comparison of multiple models of eye movement in different individuals with cognitive impairment

	Total	NC	MCI	ADD	<i>P</i> -value	Stats
Prosaccade task						
Gaze duration, (ms)	2230.56(185.53)	2248.78(209.37)	2231.31(147.69)	2226.39(212.51)	0.303	2.390
Number of gaze point	200.75(16.70)	202.30(18.84)	200.82(13.29)	200.38(19.04)	0.302	2.394
Saccadic latency, (ms)	280.82(88.60)	243.82(54.61)	274.20(74.64)	322.65(105.75)	<0.001*	70.9
Saccadic distance, (pixels)	1816.71(1148.48)	1633.79(680.85)	1747.13(838.94)	2225.80(1626.93)	<0.001*	29.3
Saccadic average speed, (ms/pixels)	5.70(8.66)	2.36(6.87)	6.98(9.05)	5.33(10.62)	0.009*	9.353
Fixation duration, (ms)	331.47(218.36)	300.06(261.38)	317.08(187.20)	361.51(238.26)	0.007*	9.974
First fixation duration, (ms)	69.45(38.11)	68.44(39.62)	74.43(34.68)	67.39(44.69)	0.104	4.519
Number of fixation point	3.28(3.83)	3.25(4.75)	3.08(1.72)	4.00(4.88)	0.011*	9.008
Fixation latency, (ms)	674.46(570.43)	620.64(532.95)	807.80(578.87)	582.60(568.05)	<0.001*	15.7
Gaze duration in left area/total area, %	44.71(7.50)	45.65(3.60)	45.40(6.46)	43.16(10.90)	<0.001*	18.4
Gaze duration in middle area/total area, %	10.35(7.37)	8.13(4.11)	10.06(6.68)	13.15(8.88)	<0.001*	36.7
Gaze duration in right area/total area, %	44.63(8.15)	45.93(3.73)	44.25(7.53)	43.10(11.23)	0.007*	9.939
Number of forward saccades	158.43(41.18)	169.40(26.19)	161.88(29.00)	140.86(69.54)	<0.001*	35
Number of backward saccades	76.37(157.94)	7.29(163.70)	146.78(160.85)	18.63(136.20)	0.018*	7.986
Anti-saccade task						
Gaze duration, (ms)	2238.83(200.79)	2233.84 (192.36)	2205.56(136.07)	2262.50(241.94)	0.081	5.012
Number of gaze point	201.49(18.07)	201.05(17.31)	198.50(12.25)	203.63(21.78)	0.081	5.021
Saccadic latency, (ms)	672.70(366.00)	466.71(135.37)	675.67(225.62)	866.05(642.85)	<0.001*	119
Saccadic distance, (pixels)	2949.80(1218.66)	2616.88(1017.08)	2958.81(1336.21)	3146.16(1431.18)	<0.001*	17.835
Saccadic average speed, (ms/pixels)	9.63(14.50)	7.72(12.50)	12.49(14.78)	3.03(15.74)	0.007*	9.912
Fixation duration, (ms)	382.05(210.29)	350.89(210.91)	359.85(186.04)	419.09(225.51)	0.003*	11.506
First fixation duration, (ms)	68.58(38.43)	76.87(35.27)	66.73(37.50)	63.47(39.79)	0.023*	7.535
Number of fixation point	3.63(3.57)	3.44(3.46)	3.13(2.72)	4.17(4.50)	0.106	4.492
Fixation latency, (ms)	798.61(847.86)	713.88(646.71)	1050.88(918.06)	625.42(924.35)	0.016*	8.256
Gaze duration in left area/total area, %	43.15(7.21)	43.16(4.04)	44.34(8.33)	42.14(9.95)	0.071	5.292
Gaze duration in middle area/total area, %	11.27(5.17)	10.53(3.45)	11.76(5.35)	11.46(5.68)	0.242	2.835
Gaze duration in right/total area, %	45.15(6.94)	45.41(5.68)	44.47(7.31)	45.64(9.56)	0.196	3.257
Number of forward saccades	28.50(33.42)	15.56(17.07)	27.50(17.38)	45.00(48.75)	<0.001*	93.4
Number of backward saccades	37.50(100.09)	69.60(150.49)	31.88(60.18)	44.25(67.00)	0.459	1.558
Error rate, %	0.00(0.34)	0.00(0.00)	0.00(0.04)	0.375(0.875)	<0.001*	109
Fixation task						
Gaze duration, (ms)	22,683.33(1572.22)	22,855.56(1172.22)	22,588.89(1455.56)	22,500(2133.33)	0.272	2.603
Number of gaze point	2043.00(143.00)	2057.00(106.00)	2036.00(133.00)	2025.00(192.00)	0.260	2.697
Number of fixation point	30.00(38.00)	25.50(50)	27.00(27)	37.00(53)	0.029*	7.087
Fixation jump-offs number	7.00(10.00)	4.00(6.00)	8.00(12.00)	7.00(10.00)	0.003*	11.423
Smooth pursuit tasks						
Gaze duration, (ms)	2223.61(192.01)	2248.61(198.96)	2206.48(152.08)	2217.59(224.07)	0.010*	9.153
Number of gaze point	200.13(17.28)	202.38(17.91)	198.58(13.69)	199.58(20.17)	0.010*	9.158
Number of fixation point	6.00(3.13)	6.00(22.25)	5.88(2.50)	6.25(4.46)	0.242	2.834
Pursuit number	3.58(1.17)	3.75(1.03)	3.63(1.08)	3.38(1.42)	0.017*	8.209
Average pursuit number	34.00(14.00)	36.00(16.00)	38.00(16.00)	31.00(14.00)	0.002*	12.566

Data are presented as median (interquartile range); Kruskal–Wallis test for comparisons between three groups. Abbreviations: *NC* normal cognition, *MCI* mild cognitive impairment, *ADD* dementia due to Alzheimer's disease

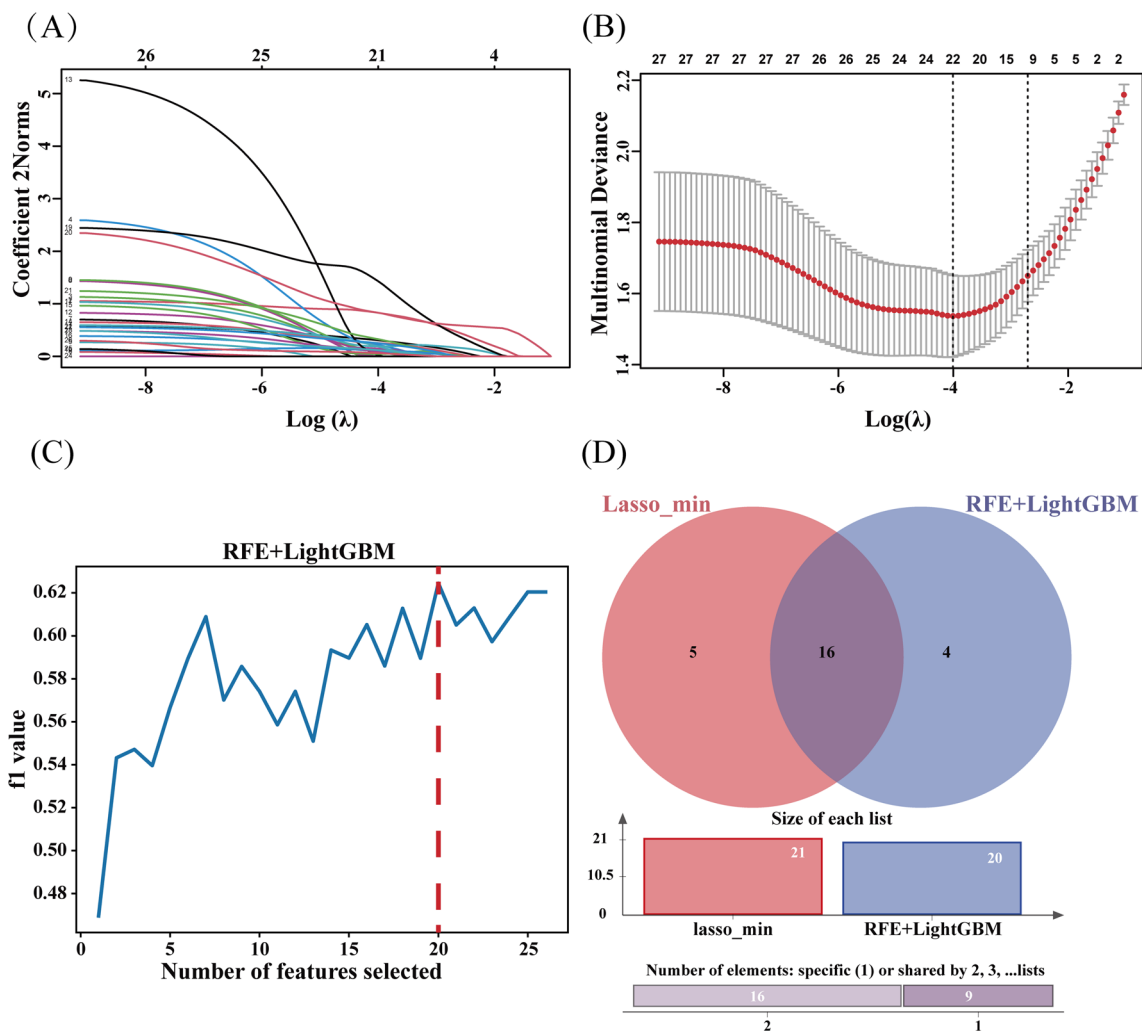


Fig. 2 **A** Lasso coefficient profiles of the twenty-five eye movement parameters. **B** Selection of the optimal lambda value through the tenfold cross-validation. **C** Line graph shows the f1 value based on

different numbers of eye movement parameters in LightGBM-RFE model. **D** Screening of sixteen eye movement parameters using Lasso and LightGBM-RFE algorithms

Table 3 The 16 eye movement parameters screened by Lasso and LightGBM-RFE algorithms

Task	Variables
Prosaccade task	Saccadic latency; gaze duration in right area/ total area; number of backward saccades; fixation point; number of forward saccades; saccadic average speed
Anti-saccade task	Saccadic latency; error rate; number of forward saccades; fixation latency; first fixation duration; saccadic distance; fixation duration
Fixation task	Fixation point; fixation jump-off number
Smooth pursuit task	Gaze duration

F1-score metrics. Results showed that GBC, LightGBM, random forest classifier, and extra trees classifier were the four best-performing models according to the F1-score.

Logistic regression was allowed to proceed to subsequent parameter optimization since it is widely used to deal with classification problems. After parameter optimization, the GBC model showed the optimal balanced prediction performance after parameter optimization and the accuracy, recall, precision, and F1-score were 0.6820, 0.6681, 0.6761, and 0.6632. Table 4 and Fig. 3 show the specific performance of the five machine learning models.

To further validate the performance of the GBC model, comparative experiments were conducted using the FCNN, FT-Transformer, and self-attention models. The accuracy, recall, precision, and F1-score of the FCNN model were 0.6508, 0.6508, 0.6870, and 0.6553, respectively (Fig. 4A). The corresponding metrics for the FT-Transformer model were 0.6091, 0.5970, 0.5969, and 0.5802 (Fig. 4B). For the self-attention model, these metrics were 0.5818, 0.5576, 0.6053, and 0.5404 (Fig. 4C). The results indicate that

Table 4 The specific performance of the five machine learning models

Model	Accuracy	Recall	Precision	F1
Gradient boosting classifier	0.6820	0.6681	0.6761	0.6632
Light gradient boosting machine	0.6503	0.6399	0.6567	0.6383
Random forest classifier	0.6196	0.6036	0.6243	0.6010
Extra trees classifier	0.6629	0.6717	0.6617	0.6452
Logistic regression	0.6120	0.6057	0.5905	0.5879

the performance of these three models is inferior to that of the GBC model. This discrepancy may be due to the one-dimensional vector nature of the eye-movement data in this study, which may not be well-suited for end-to-end machine learning model development. Additionally, the three models require a substantial volume of data for robust training to achieve effective feature representations and generalization. In contrast, the GBC model may perform better on small-scale structured datasets.

The Correlations of Top Five Eye Movement Parameters Ranked by Importance Based on GBC Model and Neuropsychological Scales

To determine the contribution of individual features to the GBC model's performance, we computed the Gini importance and identified the five most important features. The five most important features were saccadic latency in anti-saccade task, error rate in anti-saccade task, saccadic latency in prosaccade task, number of forward saccades in anti-saccade task, and gaze duration in the right area/total area (%) in prosaccade task. Detailed results are presented in Fig. 5.

Thereafter, correlation analyses were conducted between these features and the neuropsychological test scores (Fig. 6A). Saccadic latency in prosaccade task was positively correlated with IADL ($r=0.41$; $p<0.001$), negatively correlated with CDT ($r=-0.49$; $p<0.001$) and TMT-A ($r=-0.38$; $p<0.001$). Saccadic latency in anti-saccade task was negatively correlated with FCSRT-immediate total recall ($r=-0.34$; $p<0.001$), BNT-total score ($r=-0.37$; $p<0.001$) and TMT-A ($r=-0.46$; $p<0.001$). Error rate in anti-saccade task was negatively correlated with CDT ($r=-0.48$; $p<0.001$), TMT-A ($r=0.57$; $p<0.001$), and TMT-B ($r=0.49$; $p<0.001$). Number of forward saccades in anti-saccade task was negatively correlated with reverse-DST ($r=-0.44$; $p<0.001$), TMT-A ($r=-0.55$; $p<0.001$), and TMT-B ($r=-0.46$; $p<0.001$). The detailed presentation of the subgroup analysis can be observed in Fig. 6B.

Discussion

The present study establishes a three-classification diagnostic model for ADD, MCI, and normal cognition with 68.2% of accuracy and 66.32% of F1-score. Moreover, the new link of eye movements with episodic memory, language, and IADL has been detected. Together, our results indicate that eye movement biomarkers are of great significance for the diagnosis of AD in clinical practice. To the best of our knowledge, this is the first study to evaluate the potential of eye movement biomarkers in differentiating ADD, MCI and normal cognition, and to elucidate the association between eye movement features and individual cognitive functions.

The Three-Classification Model for Diagnosing AD Based on Eye Movements

In this study, the proposed three-classification diagnostic model could accurately and directly distinguish between ADD, MCI, and normal cognition. In contrast, the binary-classification models established in previous studies are only able to differentiate between AD or MCI and normal cognitive function—examples include Opwonya et al. [19], Jiang et al. [20] and Liu et al. [24]. Other studies that included normal cognition, MCI and AD did not construct the diagnostic models, instead simply comparing the eye movement parameters among the three groups, lacking the practical interest in clinical work [58, 59]. Therefore, our study addresses this gap and establishes a foundation for future eye-movement applications in the clinical aiding diagnosis of AD.

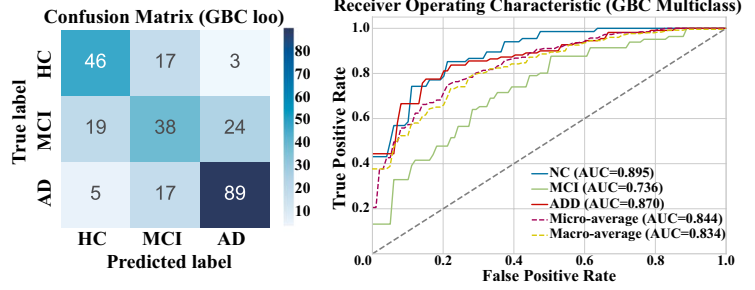
Non-invasive, Easy-to-Implement, and Cost-Effective Eye Movement Technology

In a recent study, a combination of multimodal data, encompassing MRI, PET, cerebrospinal fluid biomarkers, and genetic features, has been employed to conduct a multiclass diagnosis of AD. The study achieves accuracies of 66.7% and F1 scores of 64.9% [25]. However, these multimodal methods are complicated, invasive, and high-cost. Although the features adopted in the study are the central pathological hallmarks of AD, it is noteworthy that the clinical symptoms of AD patients do not always align with the evidence of pathological changes [58]. Hence, using pathological criteria to identify the early stage of AD, especially MCI, is problematic.

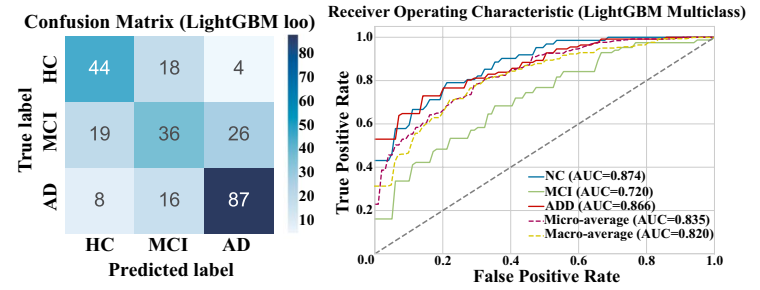
Eye movement involves a complex oculomotor control system formed by a wide range of cerebral regions, and there is evidence that the pathology associated with AD can affect the oculomotor brain areas [23, 60–62]. Therefore, by analyzing eye movement data, specific patterns in the

Fig. 3 The confusion matrix and receiver operating characteristic of the five machine learning models

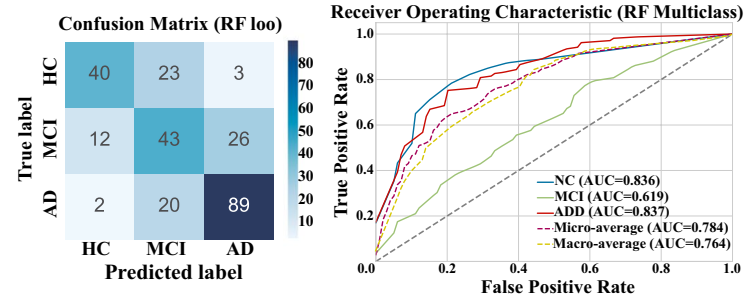
(A) Gradient Boosting Classifier



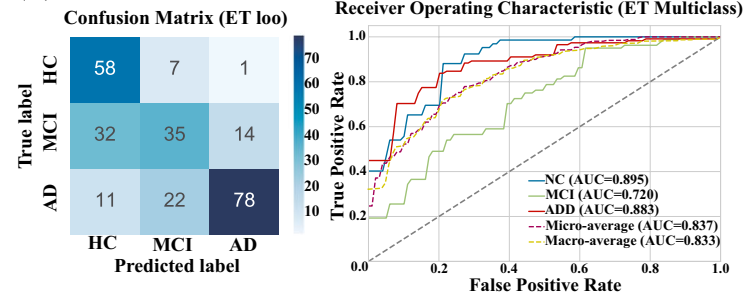
(B) Light Gradient Boosting Machine



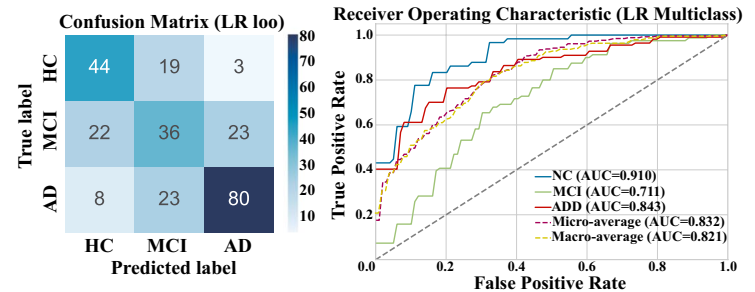
(C) Random Forest Classifier



(D) Extra Trees Classifier



(E) Logistic Regression



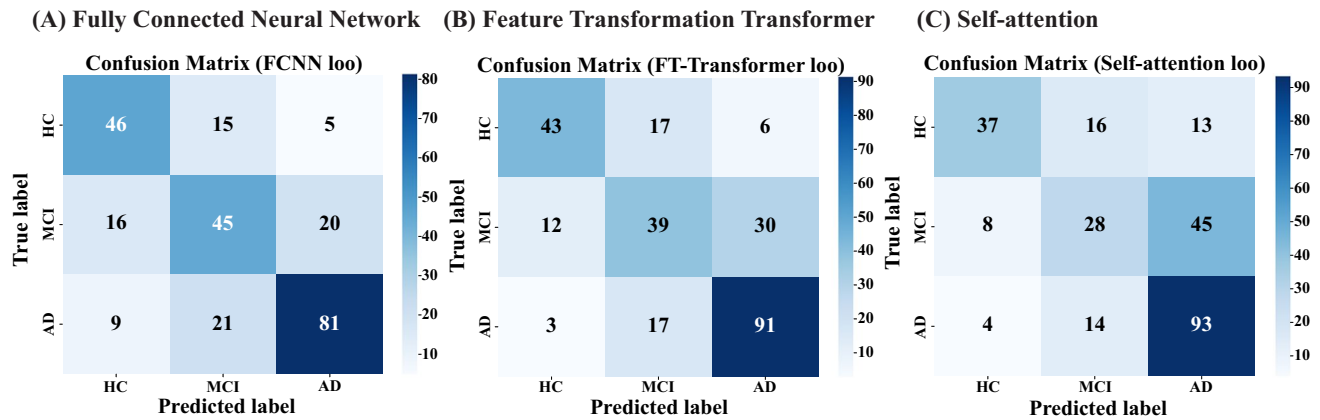


Fig. 4 The confusion matrix and receiver operating characteristic of FCNN, FT-transformer, and self-attention model

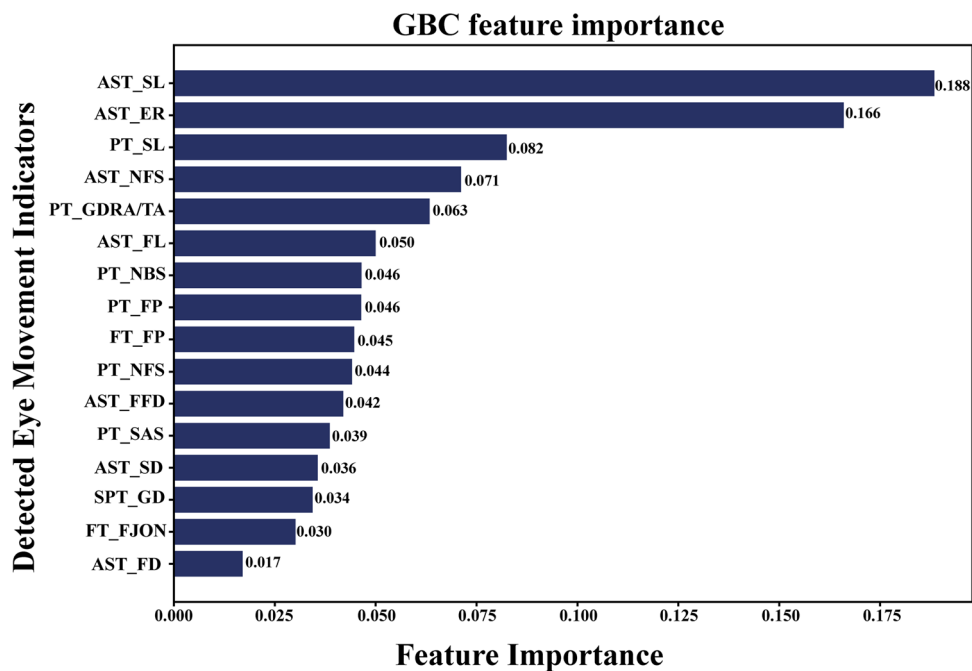


Fig. 5 The plot shows the ranking of the relative importance of features in GBC models. Abbreviation: AST_SL, saccadic latency in anti-saccade task; AST_ER, error rate in anti-saccade task; PT_SL, saccadic latency in prosaccade task; AST_NFS, number of forward saccades in anti-saccade task; PT_GDRA/TA, gaze duration in right area/total area; AST_FL, fixation latency in anti-saccade task; PT_NBS, number of backward saccades in prosaccade task; PT_FP, fixa-

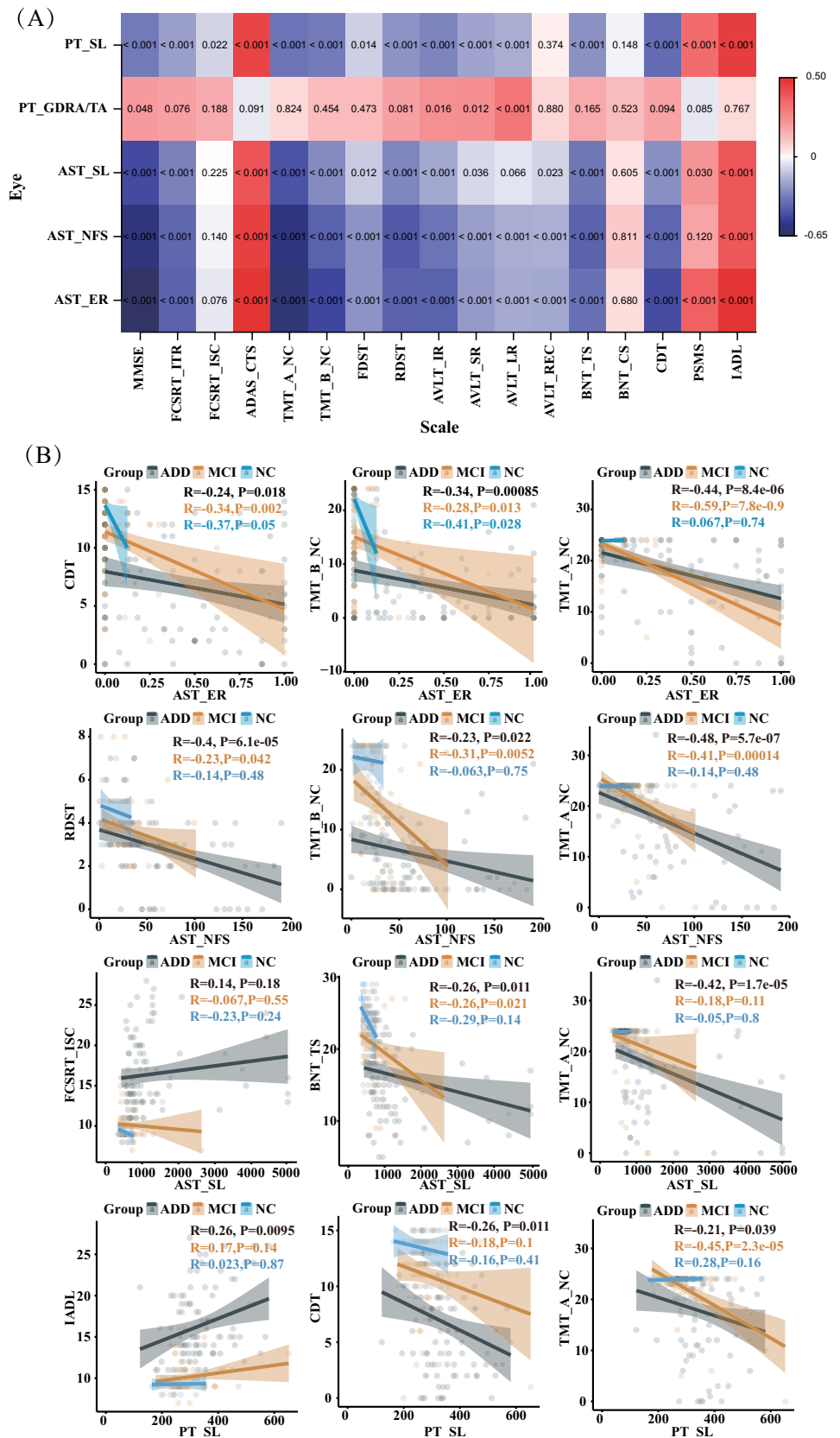
tion point in prosaccade task; FT_FP, fixation point in fixation task; PT_NFS, number of forward saccades in prosaccade task; AST_FFD, first fixation duration in anti-saccade task; PT_SAS, saccadic average speed in prosaccade task; AST_SD, saccadic distance in anti-saccade task; SPT_GD, gaze duration in smooth pursuit task; FT_FJON, fixation jump-offs number in fixation task; AST_FD, fixation duration in anti-saccade task

processing of visual stimuli in patients with AD can be identified, which may be associated with the cognitive and neurological features of early AD. Furthermore, eye movement is a non-invasive, timely, easy to perform and cost-effective technique, and is easily accepted by patients [63]. The combination between eye movement parameters and GBC model is beneficial to identifying patients with AD in a timely manner and prompt necessary treatment interventions.

Proposed Model Outperforms Current Studies in Accuracy

The accuracy of our proposed model based on eye movement parameters in the three classifications of ADD, MCI, and normal cognition is 68.20%, which is higher than majority of current studies [29]. In the study by Cicalese et al. [30], the multiclass diagnosis of AD was performed using

Fig. 6 **A** Heatmap of correlation between eye movement parameters and neuropsychological scales. **B** Scatter plot of correlation between eye movement parameters in subgroups and neuropsychological scales. Abbreviation: AST_SL, saccadic latency in anti-saccade task; AST_ER, error rate in anti-saccade task; PT_SL, saccadic latency in prosaccade task; AST_NFS, number of forward saccades in anti-saccade task; PT_GDRA/TA, gaze duration in right area/total area; MMSE, Mini-mental State Examination; FCSRT_I TR, immediate total recall in the Free and Cued Selective Reminding Test; FCSRT_ISC, index of sensitivity of cueing in the Free and Cue Selective Reminding Test; ADAS_CTS, total scores in the Alzheimer’s Disease Assessment Scale-Cognitive section; TMT_A_NC, number of correct in Trail Making Test-A; TMT_B_NC, number of correct in Trail Making Test-B; FDST, Forward Digit Span Testing; RDST, Reverse Digit Span Testing; AVLT_IR, immediate recall in Auditory Verbal Learning Test; AVLT_SR, short-term free delayed recall in Auditory Verbal Learning Test; AVLT_LR, long-term free delayed recall in Auditory Verbal Learning Test; AVLT_REC, recognition in Auditory Verbal Learning Test; BNT_TS, total scores in Boston Naming Test; BNT_CS, phonemic cue scores in Boston Naming Test; CDT, Clock Drawing Test; PSMS, physical self-maintenance scale; IADL, instrumental activity of daily life



electroencephalography (achieved accuracies of 65.52%) and fNIRS (achieved accuracies of 58.62%), both accuracies were lower than our model. In fact, the model in this study also exhibits superior diagnostic accuracy compared to current binary classification models. The DISCN model, developed by Liu et al. [24], achieved an AUC of 0.81 in differentiating between AD and normal cognition, while our model attained an AUC of 0.87.

The high accuracy of our model can be attributed to several factors: First, unlike the image data used by Liu et al. and Cicalese et al., which is easily affected by noise and variations leading to model interference and reduced accuracy, the structured data utilized in this study is less susceptible to such factors. Additionally, structured data features are highly intuitive, with clear relevance and certain advantages. Second, out of 17 machine models tested in this study, the GBC model demonstrates the best performance. The GBC operates as an ensemble method, integrating multiple weak learning models to form a powerful model, enabling it to achieve high accuracy in prediction and classification tasks [64]. Lastly, participants for this study were strictly selected according to inclusion and exclusion criteria, ensuring the quality and reliability of the data.

Anti-saccade Tasks Are More Valuable for the Diagnosis of AD than the Other Three Tasks

Through ranking the importance of the features, the five most important parameters are all derived from the prosaccade tasks and the anti-saccade tasks. This observation suggests that the saccadic tasks may possess greater diagnostic value for AD compared to the fixation and smooth pursuit task. What is more, the anti-saccade task may be more sensitive to cognitive changes than the prosaccade task.

Prosaccade latency, anti-saccade latency, and anti-saccade error rate are the three most significant biomarkers to discriminate between patients with and without cognitive impairment, which is in good agreement with previous findings [15, 22, 65]. Besides these, the number of forward saccades and gaze duration in the right area to the total area also show a good performance in class diagnosis of AD, which has not been reported in previous studies. If the saccade is in the same direction as the target's movement, the saccade is classified as a forward saccade. ADD and MCI have a greater number of forward saccades in the anti-saccade task compared with normal cognition. This indicates that ADD and MCI have an increased number of error saccades in the opposite direction of the instructions, similar to the error rate of anti-saccade task. Meanwhile, the proportion of gaze duration in both the left and right area to the total area is reduced in patients with ADD and MCI. This reduction may be attributed to the slowing of reaction speed in AD patients, resulting in prolonged gaze duration in the middle area.

The Novel Association Between Eye Movements and Episodic Memory, Language, and IADL Has Been Detected

While progressive memory loss is the primary cognitive deficit in all patients with AD, other non-memory domains are also affected at the early stage including attention, visuospatial abilities, and language functions [31]. Attention and executive functions, in particular, are often the first and common non-memory domains affected by AD [66]. In this study, saccadic latency in the prosaccade task, saccadic latency, error rate, and number of forward saccades of the anti-saccade task were negatively correlated with TMT-A and DST tests. This suggests a potential association between attention deficits and increased saccadic latency and error rate in patients with AD, aligning with prior research outcomes [17, 67]. Indeed, there exists an extensive literature demonstrating the close relationship between saccadic eye movements and attention. However, the relationship between deficit of eye movement and episodic memory impairment has received less attention [68].

Episodic memory impairment is widely recognized as a prominent cognitive characteristic in the early stages of AD [69]. It is reported that FCSRT-immediate total recall score demonstrates the best value for predicting the dementia progression in 2-year follow-up [70]. Our results discover a negative association between parameters in anti-saccade task and FCSRT-immediate total recall, indicating that eye movements might potentially serve as an indicator of episodic memory in patients with AD. Therefore, we hypothesize that eye movement parameters, particularly in the context of the anti-saccade task, may also have the potential to predict the progression of dementia. Subsequent investigations will be undertaken to address this issue.

IADL is considered to be important for maintaining an independent lifestyle in the community [67]. This study has identified a correlation between the increased error rate of anti-saccade and the prolonged latency of prosaccade performance with the decline in IADL. Another study shows a strong positive correlation between anti-saccade performance and the driving performance. Older participants at risk of driving impairment exhibit abnormal anti-saccade error rates, similar to patients with dementia [71]. Overall, the saccadic latency in the prosaccade task and the anti-saccade error rates have the potential to serve as useful screening tools for the IADL in individuals with AD.

Eye movement abnormalities can impact the language processing abilities of patients with AD [72]. Our study revealed that eye movement parameters in prosaccade and anti-saccade tasks are significantly associated with the total scores of spontaneous naming and semantic cueing naming in the BNT, while showing no correlation with phonemic cueing scores. It is established that spontaneous naming

errors and semantic cueing errors are early indicators of AD, in contrast to phonemic cueing [73]. Our study indirectly supported this conclusion and proposed that eye movement abnormalities have the potential to be a crucial indicator in the evaluation of language functional impairment in AD patients.

Future Work

Despite this study has made some progress, much work still needs to be done. For instance, the other types of dementia including frontotemporal dementia, Lewy body dementia, and dementia in Parkinsonism are not included in the study. Future research should encompass these types of dementia to determine the specificity of eye movement in the diagnosis of AD.

There are some similarities between MCI and AD or between MCI and NC, so including MCI will reduce the diagnostic accuracy. The ROC curves depicted in Fig. 3A reveal that the diagnostic accuracy of MCI ($AUC = 0.736$) is comparatively lower than that of NC ($AUC = 0.895$) and ADD ($AUC = 0.870$). As previously indicated by Weiming Lin et al., the classification of progressive MCI and stable MCI is crucial for enhancing the accuracy of three-classification diagnosis of AD [29]. Our future work will undertake the classification of MCI subtypes to attain improved levels of accuracy.

The collection of eye movement data can be conducted flexibly, allowing for completion in various comfortable environments beyond the confines of hospitals (Neuroimaging techniques are not capable of doing so.). With the advent of digital healthcare, the mobile phone camera combined with machine learning algorithms could afford mobile phone eye tracking [74]. Our results proposed a scientific basis for remote evaluation of cognitive function and efficacy detection of remote cognitive rehabilitation. Furthermore, Jessica et al. proposed that computer vision techniques can be used to analyze the eye movements of AD patients in naturalistic scenarios, with areas of interest utilized for early detection [23]. This points the way for future research on eye movement diagnosis of AD.

In a nutshell, future research could incorporate more diverse datasets and subtype the MCI patients. Moreover, there is potential in exploring the use of mobile phone eye tracking for remote assessment of cognitive function.

Conclusion

In this study, a three-classification model for distinguishing ADD, MCI, and normal cognition is established based on machine learning and eye movement features. The

classification accuracy of the model is 68.20%, which is higher than the majority of previous models. Furthermore, this study also reveals the correlations between the eye movement parameters and various subdomains of cognition, including attention, episodic memory, immediate memory, and language and visuospatial skills.

Abbreviations ADD: Dementia due to Alzheimer's disease; MCI: Mild cognitive impairment; PET: Positron emission tomography; AD: Alzheimer's disease; AUC: Area under the receiver operating characteristic curve; MMSE: Mini-mental State Examination; FCSRT: Free and Cued Selective Reminding Test; IADL: Instrumental activity of daily life; BNT: Boston Naming Test; DISCN: Depth-induce Integrated Comparison serial attention Network; MRI: Magnetic resonance imaging; ADAS-Cog: Alzheimer's Disease Assessment Scale-Cognitive section; TMT: Trail Making Test; DST: Digit Span Test; CDT: Clock Drawing Test; AVLT: Auditory Verbal Learning Test; PSMS: Physical self-maintenance scale; LightGBM-RFE: Light gradient boosting machine-based recursive feature elimination; GBC: Gradient boosting classifier; FCNN: Fully Connected Neural Network; FT-Transformer: Feature Transformation-Transformer

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12559-024-10346-5>.

Author Contribution Jiaqi Song was a major contributor in writing the manuscript and collecting data. Haodong Huang took primary responsibility for the data analysis. Jiarui Liu, Jiani Wu, Yingxi Chen, Lisong Wang, and Fuxin Zhong assisted in data collection. Zihan Lin and Xinyi Tang assisted in compiling the data. Xiaoqin Wang, Mengyu Yan, and Wenbo Zhang contributed to the design of the study and review of the manuscript. Xintong Liu was responsible for cognitive assessment. Yang Lü and Weihua Yu contributed the initial idea and critically reviewed the manuscript. All authors read and approved the final manuscript.

Funding This study was supported by grants from Chongqing Talent Plan (cstc2022ycjh-bgzxm0184), Science Innovation Programs Led by the Academicians in Chongqing under Project (2022YSZX-JSX0002C-STB), Key Project of Technological Innovation and Application Development of Chongqing Science & Technology Bureau (CSTC2021jcsx-gksb-N0020), STI2030-Major Projects (No. 2021ZD0201802), Program for Youth Innovation in Future Medicine, Chongqing Medical University (W0166), and the fund for top talent cultivation project of Chongqing Medical University (BJRC202311).

Data Availability No datasets were generated or analysed during the current study.

Declarations

Ethics Approval and Consent to Participate The protocol was approved by the Medical Ethics Committee of The First Affiliated Hospital of Chongqing Medical University (approval number: 20212901; time of ethics approval: 10 May 2021). The study conforms with the World Medical Association Declaration of Helsinki published on the website of the Journal of American Medical Association.

Consent for Publication Written informed consent was obtained from the participants.

Competing Interests The authors declare no competing interests.

References

1. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health*. 2022 Jan 6;7(2):e105–25.
2. Arvanitakis Z, Bennett DA. What is dementia? *JAMA*. 2019;322(17):1728.
3. Fratiglioni L, Launer LJ, Andersen K, Breteler MM, Copeland JR, Dartigues JF, et al. Incidence of dementia and major subtypes in Europe: a collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology*. 2000;54(11 Suppl 5):S10–15.
4. Mild cognitive impairment – beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment - Winblad - (2004) - *Journal of Internal Medicine* - Wiley Online Library [Internet]. [cited 2023 Nov 15]. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2796.2004.01380.x?sid=nlm%3Apubmed>
5. Mufson EJ, Counts SE, Perez SE, Ginsberg SD. Cholinergic system during the progression of Alzheimer's disease: therapeutic implications. *Expert Rev Neurother*. 2008;8(11):1703–18.
6. Wong W. Economic burden of Alzheimer disease and managed care considerations. *Am J Manag Care*. 2020;26(8 Suppl):S177–83.
7. Vickrey BG, Mittman BS, Connor KI, Pearson ML, Della Penna RD, Ganiats TG, et al. The effect of a disease management intervention on quality and outcomes of dementia care. *Ann Intern Med*. 2006;145(10):713–26.
8. Ossenkoppele R, van der Kant R, Hansson O. Tau biomarkers in Alzheimer's disease: towards implementation in clinical practice and trials. *Lancet Neurol*. 2022;21(8):726–34.
9. Chappelle M, Iaccarino L, Soleimani-Meigooni D, Rabino-vici GD. The role of amyloid PET in imaging neurodegenerative disorders: a review. *J Nucl Med Off Publ Soc Nucl Med*. 2022;63(Suppl 1):13S-19S.
10. Rentz DM, Parra Rodriguez MA, Amariglio R, Stern Y, Sperling R, Ferris S. Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: a selective review. *Alzheimers Res Ther*. 2013;5(6):58.
11. Duke Han S, Nguyen CP, Stricker NH, Nation DA. Detectable neuropsychological differences in early preclinical Alzheimer's disease: a meta-analysis. *Neuropsychol Rev*. 2017;27(4):305–25.
12. Jansen CE. Cognitive changes associated with cancer and cancer therapy: patient assessment and education. *Semin Oncol Nurs*. 2013;29(4):270–9.
13. Flaks MK, Yassuda MS, Regina ACB, Cid CG, Camargo CHP, Gattaz WF, et al. The short cognitive performance test (SKT): a preliminary study of its psychometric properties in Brazil. *Int Psychogeriatr*. 2006;18(1):121–33.
14. Crawford TJ, Higham S, Renvoize T, Patel J, Dale M, Suriya A, et al. Inhibitory control of saccadic eye movements and cognitive impairment in Alzheimer's disease. *Biol Psychiatry*. 2005;57(9):1052–60.
15. Kaufman LD, Pratt J, Levine B, Black SE. Executive deficits detected in mild Alzheimer's disease using the antisaccade task. *Brain Behav*. 2012;2(1):15–21.
16. Opwonya J, Doan DNT, Kim SG, Kim JI, Ku B, Kim S, et al. Saccadic eye movement in mild cognitive impairment and Alzheimer's disease: a systematic review and meta-analysis. *Neuropsychol Rev*. 2022;32(2):193–227.
17. Noiret N, Carvalho N, Laurent É, Chopard G, Binetruy M, Nico-lier M, et al. Saccadic eye movements and attentional control in Alzheimer's disease. *Arch Clin Neuropsychol*. 2018;33(1):1–13.
18. Peltsch A, Hemraj A, Garcia A, Munoz DP. Saccade deficits in amnesic mild cognitive impairment resemble mild Alzheimer's disease. *Eur J Neurosci*. 2014;39(11):2000–13.
19. Opwonya J, Ku B, Lee KH, Kim JI, Kim JU. Eye movement changes as an indicator of mild cognitive impairment. *Front Neurosci*. 2023;15(17):1171417.
20. Jiang J, Yan Z, Sheng C, Wang M, Guan Q, Yu Z, et al. A novel detection tool for mild cognitive impairment patients based on eye movement and electroencephalogram. *J Alzheimers Dis*. 2019;72(2):389–99.
21. Nie J, Qiu Q, Phillips M, Sun L, Yan F, Lin X, et al. Early diagnosis of mild cognitive impairment based on eye movement parameters in an aging Chinese population. *Front Aging Neurosci*. 2020;12: 221.
22. Yang Q, Wang T, Su N, Xiao S, Kapoula Z. Specific saccade deficits in patients with Alzheimer's disease at mild to moderate stage and in patients with amnesic mild cognitive impairment. *Age*. 2013;35(4):1287–98.
23. Beltrán J, García-Vázquez MS, Benois-Pineau J, Gutierrez-Robledo LM, Dartigues JF. Computational techniques for eye movements analysis towards supporting early diagnosis of Alzheimer's disease: a review. *Comput Math Methods Med*. 2018;2018:2676409.
24. Liu Y, Zhang W, Wang S, Zuo F, Jing P, Ji Y. Depth-induced saliency comparison network for diagnosis of Alzheimer's disease via jointly analysis of visual stimuli and eye movements. *arXiv.org* [Internet]. 2024 [cited 2024 May 25]; Available from: <https://www.semanticscholar.org/paper/Depth-induced-Saliency-Comparison-Network-for-of-of-Liu-Zhang/d4294c61c0b02c987f644f57e704e5100c31d869>.
25. Liu S, Liu S, Cai W, Che H, Pujol S, Kikinis R, et al. Multi-modal neuroimaging feature learning for multi-class diagnosis of Alzheimer's disease. *IEEE Trans Biomed Eng*. 2015;62(4):1132–40.
26. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge - PubMed [Internet]. Available from: <https://pubmed.ncbi.nlm.nih.gov/25652394/>. Accessed 22 Nov 2023.
27. Bron EE, Smits M, van der Flier WM, Vrenken H, Barkhof F, Scheltens P, et al. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: the CADDementia challenge. *Neuroimage*. 2015;1(111):562–79.
28. Tong T, Gray K, Gao Q, Chen L, Rueckert D. Multi-modal classification of Alzheimer's disease using nonlinear graph fusion. *Pattern Recognit*. 2017;1(63):171–81.
29. Lin W, Gao Q, Du M, Chen W, Tong T. Multiclass diagnosis of stages of Alzheimer's disease using linear discriminant analysis scoring for multimodal data. *Comput Biol Med*. 2021;134:104478.
30. Cicalese PA, Li R, Ahmadi MB, Wang C, Francis JT, Selvaraj S, et al. An EEG-fNIRS hybridization technique in the four-class classification of Alzheimer's disease. *J Neurosci Methods*. 2020;15(336):108618.
31. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement J Alzheimers Assoc*. 2011;7(3):263–9.
32. Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189–98.
33. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry*. 1984;141(11):1356–64.
34. Kemtes KA, Allen DN. Presentation modality influences WAIS digit span performance in younger and older adults. *J Clin Exp Neuropsychol*. 2008;30(6):661–5.

35. O'Rourke JF, Beglinger LJ, Smith MM, Mills J, Moser DJ, Rowe KC, et al. The trail making test in prodromal Huntington disease: contributions of disease progression to test performance. *J Clin Exp Neuropsychol*. 2011;33(5):567–79.
36. Williams BW, Mack W, Henderson VW. Boston naming test in Alzheimer's disease. *Neuropsychologia*. 1989;27(8):1073–9.
37. Sunderland T, Hill JL, Mellow AM, Lawlor BA, Gundersheimer J, Newhouse PA, et al. Clock drawing in Alzheimer's disease. A novel measure of dementia severity. *J Am Geriatr Soc*. 1989;37(8):725–9.
38. Query WT, Berger RA. AVLT memory scores as a function of age among general medical, neurologic and alcoholic patients. *J Clin Psychol*. 1980;36(4):1009–12.
39. Peña-Casanova J, Gramunt-Fombuena N, Quiñones-Ubeda S, Sánchez-Benavides G, Aguilar M, Badenes D, et al. Spanish Multicenter Normative Studies (NEURONORMA Project): norms for the Rey-Osterrieth complex figure (copy and memory), and free and cued selective reminding test. *Arch Clin Neuropsychol Off J Natl Acad Neuropsychol*. 2009;24(4):371–93.
40. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969;9(3):179–86.
41. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29(5):1189–232.
42. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: *Proceedings of the 31st international conference on neural information processing systems*. Red Hook: Curran Associates Inc.; 2017. pp. 3149–57. (NIPS'17).
43. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
44. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. 2006;63(1):3–42.
45. Naïve Bayes classifier - data mining algorithms - Wiley Online Library [Internet]. [cited 2024 May 28]. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118950951.ch4>.
46. McCullagh P. *Generalized linear models*. 2nd ed. New York: Routledge; 2019. p. 532.
47. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*. 1970;12(1):55–67.
48. Hastie T, Tibshirani R, Friedman JH, Friedman JH. *The elements of statistical learning: data mining, inference, and prediction*. Linear methods for classification. New York: Springer; 2009. pp. 101–37.
49. Breiman L. *Classification and regression trees*. Routledge; 2017.
50. *Pattern Recognition and machine learning* [Internet]. Available from: <https://springer.longhohoe.net/book/10.1007/978-0-387-45528-0>. Accessed 27 May 2024.
51. Drucker H. *Improving regressors using boosting techniques*. ResearchGate. 1997. Available from: https://www.researchgate.net/publication/2424244_Improving_Regressors_Using_Boosting_Techniques. Accessed 3 Jan 2024.
52. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol TIST*. 2011;2(3):1–27.
53. Hall P, Park BU, Samworth RJ. Choice of neighbor order in nearest-neighbor classification. *Ann Stat*. 2008;36(5):2135–2152.
54. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. *Scikit-learn: machine learning in Python*. *Mach Learn*. 2011;12:2825. PYTHON.
55. Nielsen MA. *Neural networks and deep learning*. 2015. Available from: <http://neuralnetworksanddeeplearning.com>. Accessed 29 May 2024.
56. Gorishniy Y, Rubachev I, Khurlov V, Babenko A. *Revisiting deep learning models for tabular data* [Internet]. arXiv; 2023. Available from: <http://arxiv.org/abs/2106.11959>. Accessed 24 Jul 2024.
57. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. *Attention is all you need* [Internet]. arXiv; 2023. Available from: <http://arxiv.org/abs/1706.03762>. Accessed 24 Jul 2024.
58. Monsell SE, Kukull WA, Roher AE, Maarouf CL, Serrano G, Beach TG, et al. APOE4 carriers and non-carriers with the clinical diagnosis of Alzheimer's dementia and minimal amyloid plaques. *JAMA Neurol*. 2015;72(10):1124–31.
59. Quaresima V, Bisconti S, Ferrari M. A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. *Brain Lang*. 2012;121(2):79–89.
60. Tzekov R, Mullan M. Vision function abnormalities in Alzheimer disease. *Surv Ophthalmol*. 2014;59(4):414–33.
61. Rüb U, Del Tredici K, Schultz C, Büttner-Ennever JA, Braak H. The premotor region essential for rapid vertical eye movements shows early involvement in Alzheimer's disease-related cytoskeletal pathology. *Vision Res*. 2001;41(16):2149–56.
62. Boxer AL, Garbutt S, Seeley WW, Jafari A, Heuer HW, Mirsky J, et al. Saccade abnormalities in autopsy-confirmed frontotemporal lobar degeneration and Alzheimer disease. *Arch Neurol*. 2012;69(4):509–17.
63. Readman MR, Polden M, Gibbs MC, Wareing L, Crawford TJ. The potential of naturalistic eye movement tasks in the diagnosis of Alzheimer's disease: a review. *Brain Sci*. 2021;11(11):1503.
64. Zhang N, Liu C, Chen Z, An L, Ren D, Yuan F, et al. Prediction of adolescent subjective well-being: a machine learning approach. *Gen Psychiatry*. 2019;32(5):e100096.
65. Chehrehnegar N, Nejati V, Shati M, Esmaeili M, Rezvani Z, Haghi 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):1591–600.
66. Perry RJ, Hodges JR. Attention and executive deficits in Alzheimer's disease. A critical review. *Brain J Neurol*. 1999;122(Pt 3):383–404.
67. Ramzaoui H, Faure S, Spotorno S. Alzheimer's Disease, visual search, and instrumental activities of daily living: a review and a new perspective on attention and eye movements. *J Alzheimers Dis*. 2018;66(3):901–25.
68. Pereira MLF, von Camargo MZA, Aprahamian I, Forlenza OV. Eye movement analysis and cognitive processing: detecting indicators of conversion to Alzheimer's disease. *Neuropsychiatr Dis Treat*. 2014;10:1273–85.
69. Gold CA, Budson AE. Memory loss in Alzheimer's disease: implications for development of therapeutics. *Expert Rev Neurother*. 2008;8(12):1879–91.
70. Mura T, Coley N, Amieva H, Berr C, Gabelle A, Ousset PJ, et al. Cognitive decline as an outcome and marker of progression toward dementia, in early preventive trials. *Alzheimers Dement J Alzheimers Assoc*. 2022;18(4):676–87.
71. Schmitt KU, Seeger R, Fischer H, Lanz C, Muser M, Walz F, et al. Saccadic eye movement performance as an indicator of driving ability in elderly drivers. *Swiss Med Wkly*. 2015;145(0708):w14098–w14098.
72. Premeti A, Bucci MP, Isel F. Evidence from ERP and eye movements as markers of language dysfunction in dyslexia. *Brain Sci*. 2022;12(1):73.
73. Lin CY, Chen TB, Lin KN, Yeh YC, Chen WT, Wang KS, et al. Confrontation naming errors in Alzheimer's disease. *Dement Geriatr Cogn Disord*. 2013;37(1–2):86–94.
74. Brousseau B, Rose J, Eizenman M. Hybrid eye-tracking on a smartphone with CNN feature extraction and an infrared 3D model. *Sensors*. 2020;20(2):543.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.