


Article

Machine Learning-Based Detection of Cognitive Impairment from Eye-Tracking in Smooth Pursuit Tasks

Vida Groznik ^{1,2,*} , Andrea De Gobbis ^{1,2} , Dejan Georgiev ^{1,3} , Aleš Semeja ²  and Aleksander Sadikov ^{1,2,*} ¹ Faculty of Computer and Information Science, University of Ljubljana, 1000 Ljubljana, Slovenia² NEUS Diagnostics d.o.o., 1000 Ljubljana, Slovenia³ Department of Neurology, University Medical Centre Ljubljana, 1000 Ljubljana, Slovenia

* Correspondence: vida.groznik@fri.uni-lj.si (V.G.); aleksander.sadikov@fri.uni-lj.si (A.S.)

Abstract

Mild cognitive impairment represents a transitional phase between healthy ageing and dementia, including Alzheimer's disease. Early detection is essential for timely clinical intervention. This study explores the viability of smooth pursuit eye movements (SPEM) as a non-invasive biomarker for cognitive impairment. A total of 115 participants—62 with cognitive impairment and 53 cognitively healthy controls—underwent comprehensive neuropsychological assessments followed by an eye-tracking task involving smooth pursuit of horizontally and vertically moving stimuli at three different speeds. Quantitative metrics such as tracking accuracy were extracted from the eye movement recordings. These features were used to train machine learning models to distinguish cognitively impaired individuals from controls. The best-performing model achieved an area under the ROC curve (AUC) of approximately 68 %, suggesting that SPEM-based assessment has potential as part of an ensemble of eye-tracking based screening methods for early cognitive decline. Of course, additional paradigms or task designs are required to enhance diagnostic performance.

Keywords: machine learning; eye-tracking; smooth pursuit; cognitive impairment

Academic Editors: Giorgio De Nunzio and Luana Conte

Received: 31 May 2025

Revised: 7 July 2025

Accepted: 7 July 2025

Published: 11 July 2025

Citation: Groznik, V.; De Gobbis, A.; Georgiev, D.; Semeja, A.; Sadikov, A. Machine Learning-Based Detection of Cognitive Impairment from Eye-Tracking in Smooth Pursuit Tasks. *Appl. Sci.* **2025**, *15*, 7785. <https://doi.org/10.3390/app15147785>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mild cognitive impairment (MCI) refers to a decline in cognitive abilities that is larger than expected for an individual's age and level of education, yet not severe enough to significantly disrupt daily functional activities [1]. Originally, the term was introduced to describe the intermediate stage between normal ageing and Alzheimer's disease (AD) [2], but it is now recognised as a potential precursor to various forms of dementia [1]. In clinical cohorts, annual conversion rates from MCI to AD are between 10 and 15 percent [3]. AD is the most widespread neurodegenerative disease and its prevalence is increasing. As new drugs for the treatment of AD have recently been approved worldwide [4], there is a need for an early diagnosis of cognitive impairment, which can potentially slow the progression of the disease and thus improve patients' quality of life.

Smooth pursuit eye movements (SPEM) are slow controlled eye movements that allow the eyes to closely follow a moving object. There is growing evidence that smooth pursuit eye movements are closely linked to cognitive processes [5]. SPEM have been linked to attention, working memory, executive function, and decision making [5]. SPEM have been functionally linked to the frontal eye field, the dorsolateral prefrontal cortex, the basal ganglia, and the cerebellum [6]. SPEM deficits have been found in different neurological disorders, including Parkinson's disease [7,8] and mild cognitive impairment [9]. Different

impairments of SPEM have been described in MCI so far, including reduced pursuit gain, increased saccadic intrusions, impaired predictive pursuit, and delayed initiation [9]. Similar findings have been described in AD as well [10]. However, the applicability of SPEM to differentiate between MCI and normal healthy subjects has not been specifically studied so far. Therefore, the main aim of this study was to assess the efficacy of horizontal and vertical SPEM with different velocities to differentiate between patients with MCI from healthy subjects by the use of machine learning algorithms.

The paper is organised as follows. First, we present the materials and methods, including the enrolment criteria, assessment procedures, the description of our cohort, and the methods used for the data analysis. Next, in Section 3, we present the results of the statistical and machine learning analyses, followed by the related work in Section 4. The discussion of our results is presented in Section 5, and the conclusion is in Section 6.

2. Materials and Methods

2.1. Subject Enrolment

Participants were drawn from residential senior care homes. Some were referred by their doctor due to their expressed concerns about possible memory or thinking issues (subjective cognitive impairment), while others joined the study voluntarily after learning about it from other participants or directly from the research team.

Participants were eligible for inclusion in the study if they were over 40 years of age, with or without a diagnosis of cognitive impairment (CI), and had provided informed consent to take part in the study, either personally or through their legal representative.

Individuals were excluded from participation if they had uncorrected visual impairments, co-existing neurological conditions, or psychiatric disorders—including those who scored above 10 on the Geriatric Depression Scale (GDS-15 [11]). Additional exclusion criteria included a history of drug or alcohol abuse, as well as an unwillingness or inability to complete all required assessments.

The data were collected as part of a clinical study approved by the National Medical Ethics Committee of the Republic of Slovenia and conducted in accordance with the principles of the Declaration of Helsinki.

2.2. Assessment Procedures

Each participant underwent a neurological and psychological assessment, followed by an eye-tracking test procedure. A detailed description of the neurological and psychological assessments can be found in Groznik et al. [12]; for clarity and completeness, a brief summary is included below.

2.2.1. Psychological Assessment

A psychologist assessed participants' higher cognitive functions, including memory and executive function, using standardised tests such as ACE-R (Addenbrooke Cognitive Examination-Revised) [13], FAB (Frontal Assessment Battery) [14], CTMT (Comprehensive Trail Making Test) [15], and GDS-15 (Geriatric Depression Scale-15 questions) [11]. The placement of the diagnostic group was based on test cut-off points adjusted for age and education, with ACE-R and MMSE (Mini Mental State Examination, part of ACE-R) [16] scores serving as key indicators.

The final assessment, including confirmation of healthy status, was determined by a neurologist after a comprehensive clinical evaluation and review of the psychological report.

2.2.2. Neurological Examination

Each participant underwent a cognitive assessment carried out by a neurologist, with further evaluation of motor and non-motor symptoms where clinically indicated. Demographic and relevant medical information was collected using a structured questionnaire.

Based on the results of the neurological and psychological evaluations, participants were classified into one of four categories: cognitively healthy (no signs of cognitive decline), borderline (some cognitive changes observed, but not sufficient to meet criteria for mild cognitive impairment), mild cognitive impairment (MCI), or possible Alzheimer's disease (AD). Cognitive diagnoses were made according to the criteria set out in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) [17].

Individuals diagnosed with MCI demonstrated noticeable deficits in at least one of the six core cognitive domains: memory and learning, complex attention, executive function, language, perceptual–motor skills, or social cognition—while still being fully independent in both basic and instrumental activities of daily living. Individuals considered to have possible Alzheimer's disease (AD) also had impairments in at least one cognitive domain, but unlike the MCI group, they exhibited difficulties with basic activities of daily living and required assistance with instrumental activities of daily living. Participants who showed some cognitive decline in one domain, but not to a degree that met the criteria for MCI, were classified as borderline.

2.2.3. Eye-Tracking

Eye-tracking recordings were carried out using a Tobii 4C eye-tracker (produced by Tobii AB, Stockholm, Sweden) operating at 90 Hz, alongside proprietary software developed by NEUS Diagnostics, d.o.o. (Ljubljana, Slovenia), and supporting hardware, including an examiner's laptop and a 23.6-inch monitor (1920 × 1080 resolution) for participant viewing. A trained technician oversaw the procedure. The participants were seated approximately 70 cm from the screen and interacted with the tasks solely through eye movements, without physical contact with the equipment. The NEUS software presented visual tasks and guided participants through the session, simultaneously capturing and storing gaze data for subsequent analysis. All data were anonymised and made accessible only to authorised researchers.

The smooth pursuit examination included six different tasks. In each task, the patient was asked to follow a computer-based moving dot. The dot moved in a sinusoidal pattern, either vertically or horizontally in three different speeds: slow (period of 4800 ms, frequency of ≈ 0.21 Hz), medium (period 2400 ms, frequency ≈ 0.42 Hz), and fast (period 1600 ms, frequency ≈ 0.625 Hz). Each task was 10 s long. Figure 1 shows our experimental flowchart as seen from the participant's point of view. There were no other graphical elements on the participant's screen during the test.

Before the test begins, participants listen to simple instructions explaining that they should carefully follow the movement of the dot using only their eyes. Each task starts with the dot appearing at the centre of the screen. It then begins to move across the screen (either vertically or horizontally), and when the dot stops moving, a blank black screen is shown. The first task is always at a slow speed, followed by medium, and finally the fast version. The entire set of tasks is repeated three times. Figure 2 shows an example of eye movements from a healthy control while performing the medium-speed smooth pursuit task.

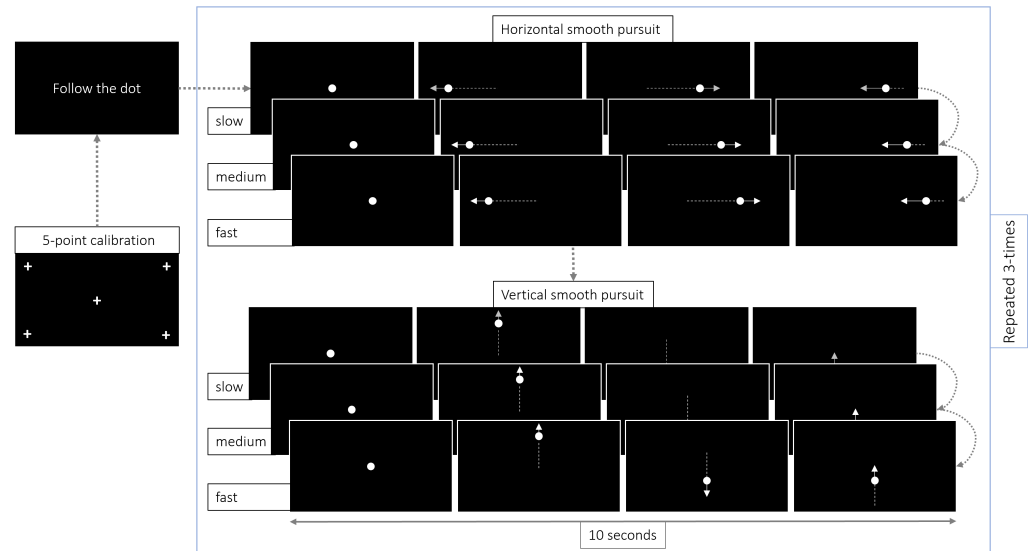


Figure 1. An illustration of our experiment as seen from the participant's point of view. The experiment starts with a 5-point calibration, followed by short instructions to the participant. From this point on, the test consists of a horizontal smooth pursuit tasks (in three speeds) followed by a vertical smooth pursuit task (in three speeds). These tasks are then repeated three times. Please note the participant sees only the dot during the smooth pursuit task. The dotted lines and arrows are included in this illustration for the purpose of presenting the direction of the movement.

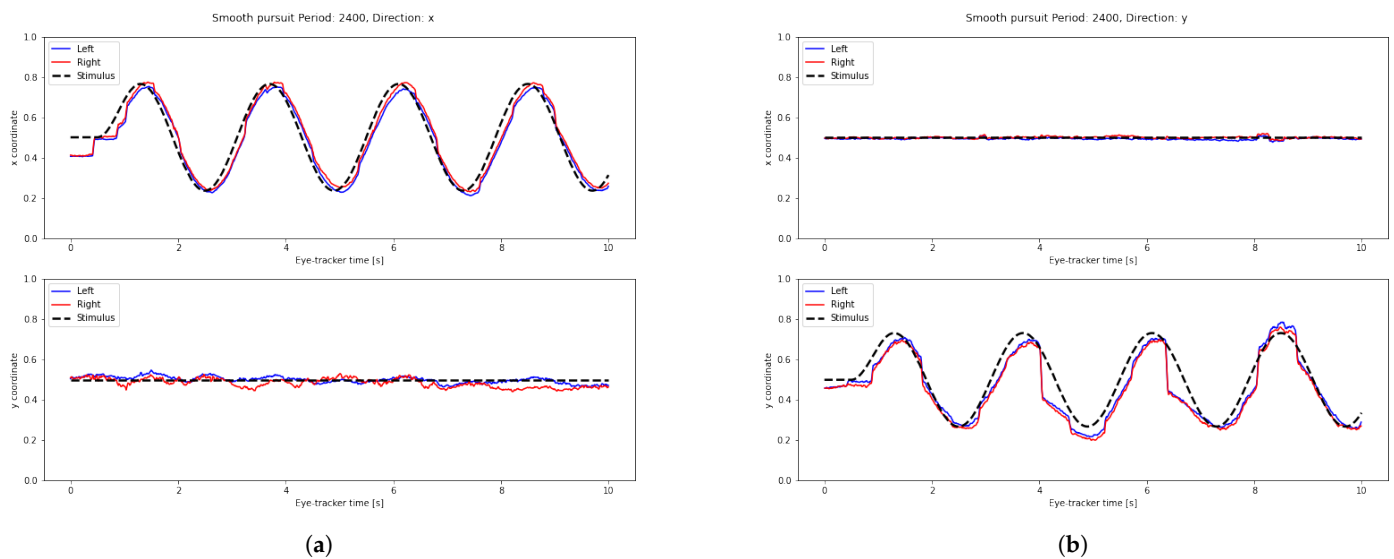


Figure 2. Trace from a smooth pursuit task (period 2400 ms, frequency ≈ 0.42 Hz) performed by a healthy control. The black dotted lines represent the movement of the stimulus on the screen, while the red and blue lines show the eye movement traces of the right and left eyes, respectively. The graphs in the top row show movement along the x-axis, while those in the bottom row show movement along the y-axis. (a) Trace for the horizontal smooth pursuit task. (b) Trace for the vertical smooth pursuit task.

2.3. Cohort Description

Data from 115 consecutively recruited participants were included in the analysis. Of these, 53 were diagnosed as cognitively healthy, 32 as borderline, 19 with mild cognitive impairment (MCI), and 11 with possible Alzheimer's disease (AD). Full details on the age and gender distribution by diagnostic group are provided in Table 1.

For the purpose of machine learning, we defined a binary classification task with two groups: healthy controls (HC), consisting of participants without signs of cognitive decline,

and cognitively impaired (CI), which included those participants classified as borderline, MCI, or possible AD. Borderline cases were included in the CI group to reflect the presence of detectable cognitive changes and to support the model's sensitivity to early, potentially progressive stages of impairment.

Table 1. Gender, age, and cognitive scores distribution per diagnosis/group.

	Healthy	Borderline	MCI	Possible AD	CI
N	53	32	19	11	62
<i>Gender</i>					
Female	40	24	12	9	45
Male	13	8	7	2	17
<i>Age</i>					
Median	63	68.5	72	83	72
Range	48–83	60–87	43–91	72–94	43–94
<i>MMSE</i>					
Mean	28.98 ± 1.03	28.38 ± 1.21	26.79 ± 1.65	23.64 ± 2.46	27.05 ± 2.36
<i>ACE-R</i>					
Mean	94.26 ± 2.91	87.41 ± 4.21	81.53 ± 6.83	62.36 ± 7.98	81.16 ± 10.84
<i>GDS-15</i>					
Mean	1.25 ± 1.93	1.56 ± 1.81	2.16 ± 2.54	3.18 ± 3.22	2.03 ± 2.37

2.4. Data Analysis

The primary goal of the machine learning analysis was to distinguish between individuals with cognitive impairment (CI) and healthy controls (HC) by examining their eye movements during smooth pursuit tasks. Therefore, we ran two experiments:

1. Investigating the statistical difference in the distribution of SPEM features extracted during eye-tracking between healthy and CI participants;
2. Assessing the performance of machine learning algorithms on the task of predicting CI based on SPEM features.

2.4.1. Data Preprocessing

Eye-tracking data were recorded as two sequences of (x, y) screen coordinates, one for each eye, and their corresponding timestamps t , with a sampling rate of 90 Hz. The dot position on the screen was recorded as a separate time series of positions (x, y) on the screen with its own timestamps t^* at a frequency of 60 Hz.

To begin with, we removed events containing invalid gaze coordinates, which typically occur when the eye-tracker is unable to register the eyes, such as during blinks or when participants look away from the screen. We also excluded five participants due to the high prevalence of invalid gaze coordinates in their recordings. Specifically, ignoring gaps of less than 150 ms caused by blinking, we excluded the participants whose ratio of invalid samples to the total was greater than 10% for all three repetitions of a specific type of task. Finally, since dot and gaze positions are sampled according to different time grids, we undersampled gaze positions at 60 Hz. We also excluded an additional six participants with the incompleting recording that might be due to the unstable connection to the database at the time of recording.

To gain clearer insights into the differences between the two groups (CI vs. HC), we transformed the raw data into features relevant to the medical domain. These features were as follows: ME (mean squared error between the dot and the gaze), ME of the gaze speed (computed with finite differences), ME when initializing smooth pursuit (first 0.7 s), gaze gain coefficient (computed as the amplitude at the oscillation frequency of the Fourier

transformed data), range spanned by gaze, number of fixations [18], percentage of invalid samples, number of predictive saccades [18], and the average distance between the gaze position of the two eyes. Among these features, all of them besides the ME were filtered by their high p-values of a Mann–Whitney U test between healthy controls and cognitively impaired subjects. Thus, we decided to only include ME for further analyses.

Following data cleaning, we computed the ME values as follows. Please note that this feature was calculated for each of the six tasks—one for each combination of stimulus speed (slow, medium, fast) and direction (horizontal, vertical); the same is true also for the other features above.

$$ME = \begin{cases} \sum_i (x_d - x_g)^2, & \text{horizontal smooth pursuit} \\ \sum_i (y_d - y_g)^2, & \text{vertical smooth pursuit} \end{cases}, \quad (1)$$

where (x_d, y_d) are the dot positions, and (x_g, y_g) are the gaze positions. Once the ME was computed for both eyes, the minimum value was kept as the final feature. The choice of only keeping the best performing eye was made due to the existence of a dominant eye. Moreover, due to calibration errors, one of the two eyes might have been recorded incorrectly, and choosing the minimum error helps alleviate this problem. Finally, since we had a value for each one of the three repetitions, we took the average over the repetitions.

As a result of this preprocessing step, we obtained six features (three speeds and two directions), for 104 participants (47 healthy and 57 CI—10 AD, 17 MCI, 30 borderline).

2.4.2. Statistical Analysis

To determine whether healthy and CI participants displayed different characteristics during smooth pursuit, we compared the distribution of ME for the two groups. For each feature, we employed a two-tailed Mann–Whitney U test to compare the groups. We selected a base confidence level of $\alpha = 0.05$, which corresponds to $\hat{\alpha} = 0.0083$ after Bonferroni correction.

2.4.3. Machine Learning Algorithms

We explored five machine learning algorithms: logistic regression (LR), decision tree (DT), random forest (RF), naïve Bayes with Gaussian likelihood (G-NB), and support vector classifier (SVC).

2.4.4. Machine Learning Pipeline

The main pipeline was a two-level nested cross-validation (CV) procedure. The outer CV layer was reserved for computing statistics on the two metrics used to evaluate the algorithms: accuracy (CA) and area under the ROC curve (AUC). The inner CV layer was used to tune the algorithms' hyperparameters. All data splits were obtained through randomized splitting and were stratified with respect to the two classes. We used 10 folds for the outer layer and 5 folds for the inner layer. Specifically, for each of the 10 outer folds, we

1. Considered the current folder as test set;
2. Used the remaining 9 folds as a training set to optimise the algorithms' hyperparameters with random search 5-fold cross validation;
3. Selected the best values of the hyperparameters using CA as criterion;
4. Evaluated the model by computing CA and AUC on the test fold.

Through this pipeline, we obtained 10 values of the two metrics for each of the five algorithms and calculated their mean values and the standard deviation (see Table 2).

The complete machine learning pipeline was implemented using the scikit-learn v1.3.2 Python library.

3. Results

3.1. Statistical Analysis

We checked for the difference in distributions of all the observed features. The distributions between healthy and cognitively impaired subjects consistently differed the most for all features that assessed how well the subjects followed the location of the stimulus (mean squared error of the difference in location across all time points). These features, along with the associated p -values, are presented in Figure 3. From the plots, it can be seen that all six of the presented features (differing in speed and direction of movement) unanimously showed that the errors (higher values of the features) increased with cognitive impairment. On the other hand, from the distribution density, we can see that the two groups were not clearly separated.

For the two most discriminative features—fast horizontal and slow vertical—we further analysed and plotted their respective distributions across different levels of cognitive impairment. In Figure 4, we observe the gradual shift towards increasing errors (higher values on the x-axis), starting with the healthy subjects, then the borderline group, MCI, and finally subjects with suspected dementia. This supports the notion that the precision of smooth pursuit gradually declines with an increasing level of cognitive impairment.

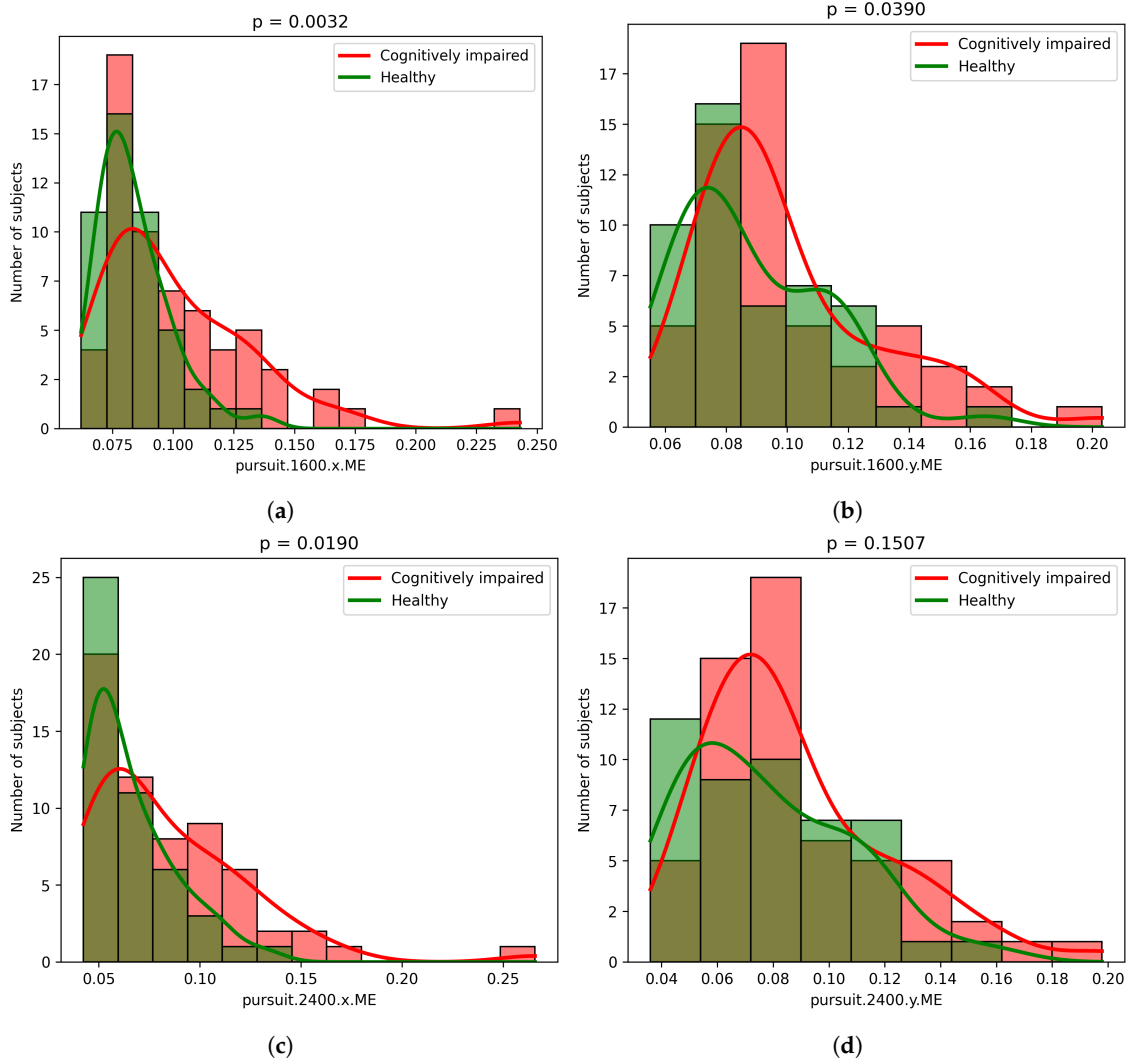


Figure 3. Cont.

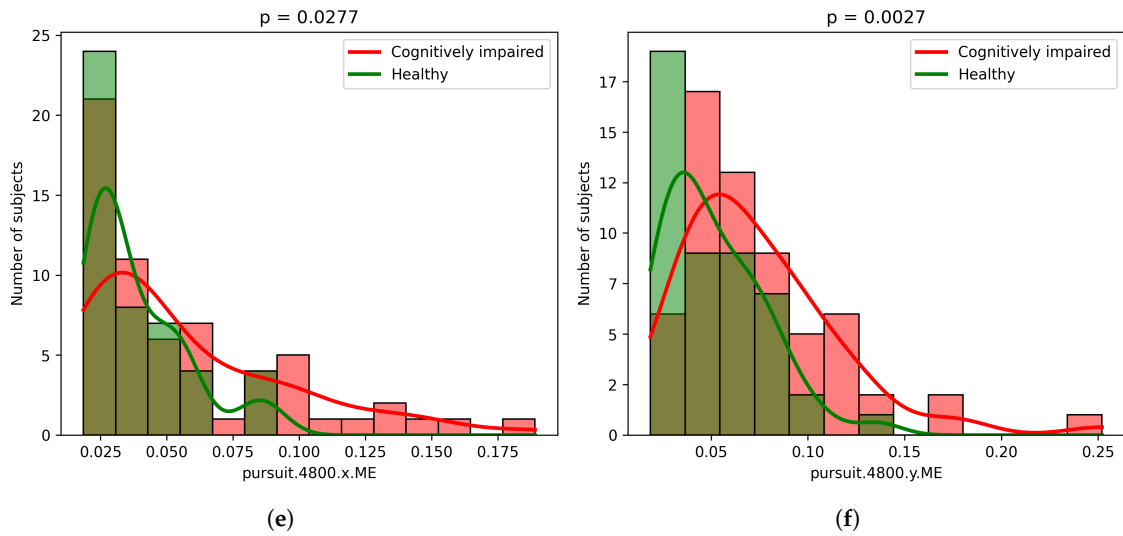


Figure 3. Distribution of smooth pursuit features over the two groups—red colour showing cognitively impaired subjects, and green showing healthy controls. (a) Fast (period 1600 ms, frequency ≈ 0.625 Hz) horizontal smooth pursuit task. (b) Fast (period 1600 ms, frequency ≈ 0.625 Hz) vertical smooth pursuit task. (c) Medium speed (period 2400 ms, frequency ≈ 0.42 Hz) horizontal smooth pursuit task. (d) Medium speed (period 2400 ms, frequency ≈ 0.42 Hz) vertical smooth pursuit task. (e) Slow (period 4800 ms, frequency ≈ 0.21 Hz) horizontal smooth pursuit task. (f) Slow (period 4800 ms, frequency ≈ 0.21 Hz) vertical smooth pursuit task.

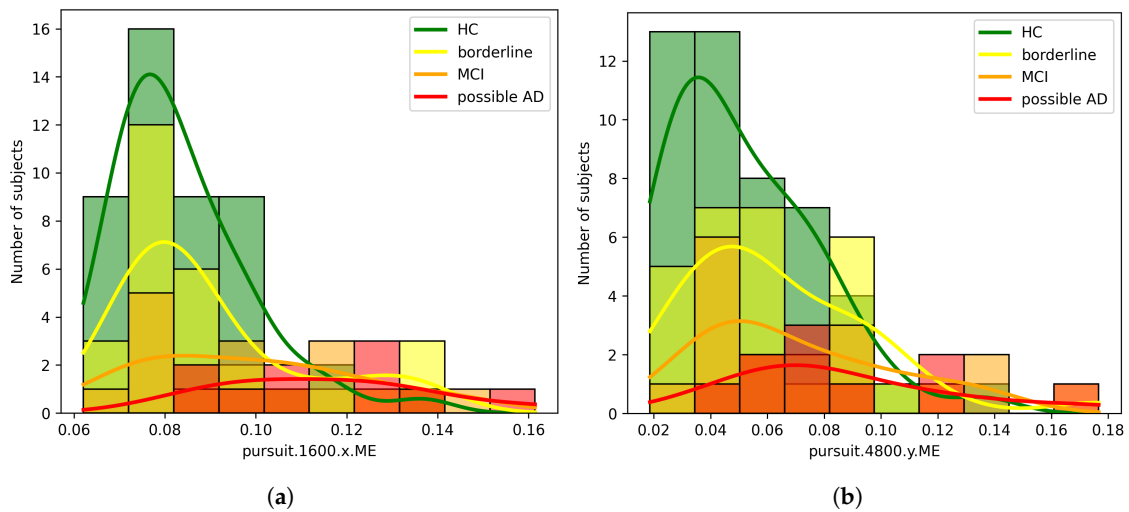


Figure 4. Distribution of smooth pursuit features over the four groups for the two most important tasks. (a) Fast (period 1600 ms, frequency ≈ 0.625 Hz) horizontal smooth pursuit task. (b) Slow (period 4800 ms, frequency ≈ 0.21 Hz) vertical smooth pursuit task.

3.2. ML Model Results

The results in Table 2 show that the best performing algorithm in terms of the AUC is logistic regression with an AUC of 0.675 and a CA of 0.619. Although the performance of the random forest based on CA is slightly better (0.628), it slightly underperforms based on the AUC (0.625). Compared to the majority classifier, all models perform better in terms of the AUC. Additionally, all models but one (SVC) perform better also in terms of the CA.

Table 2. Performance evaluation of classification algorithms. Reported values include the mean classification accuracy (CA) and area under the ROC curve (AUC), along with their respective standard deviations.

Algorithm	CA	AUC
Logistic regression	0.619 ± 0.174	0.675 ± 0.236
Decision tree	0.557 ± 0.150	0.561 ± 0.149
Random forest	0.628 ± 0.152	0.625 ± 0.219
Naïve Bayes with Gaussian likelihood	0.597 ± 0.136	0.672 ± 0.204
Support vector classifier	0.530 ± 0.142	0.603 ± 0.182
Majority (dummy)	0.548 ± 0.041	0.500 ± 0.000

Based on these results, we decided to further analyse the performance of the LR model with the AUC of 0.675. Figure 5a shows the ROC curve of the LR model (green line) compared to the naive majority classifier (grey dotted line). We can see that the model captures almost 25 % of positive cases with very few false positives at the very beginning, and if we allow for 20 % of false positives, the model captures almost 50 % of the positive cases. This can be further seen from the Figure 5b, where each bar presents a subject included in the study. The class of the participant is shown by using different colours—red for AD, orange for MCI, yellow for borderline, and green for HC. The subjects are ordered by the probability of a positive (cognitively impaired) class.

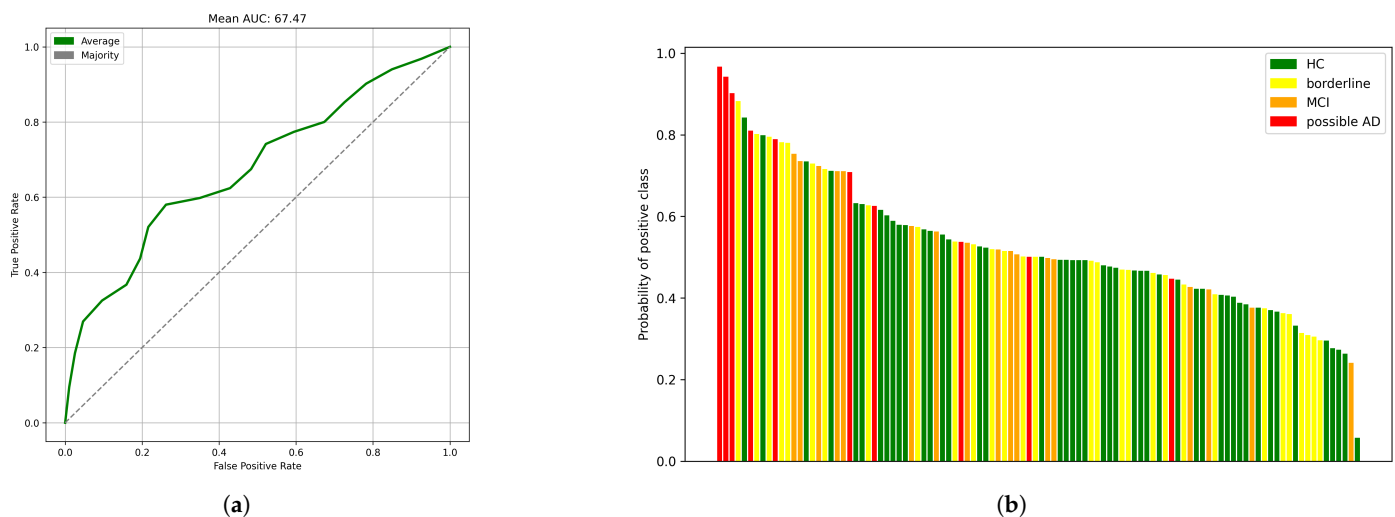


Figure 5. (a) ROC curve for the LR model (green line) and for the majority classifier (grey line). (b) A plot presenting each participant sorted by the probability of being classified as impaired. The colours represent their actual classes.

4. Related Work

Several recent studies have explored the use of eye-tracking for the early detection of cognitive impairment, often relying on broader task batteries and higher-end equipment. For instance, Oyama et al. [19] introduced a 178-second protocol using high-performance near-infrared eye-tracking to present participants with ten short movie-based tasks assessing different cognitive domains such as memory, attention, and reasoning. The system achieved a classification AUC of 0.845 for distinguishing between HC and MCI and 0.888 for general cognitive impairment (MCI+AD vs HC), with performance comparable to the MMSE. However, the test setup included fixed head positioning and video-based stimuli requiring advanced equipment and controlled testing environments.

Li et al. [20] proposed a similarly comprehensive system combining multiple visual stimuli and feature types—including saccades, fixations, smooth pursuit, and pupil size—to

develop a multi-feature classification model for MCI. Their approach applied advanced feature selection and fusion techniques to maximise discrimination between impaired and healthy participants. Although powerful, such systems are less suited to use outside of research or specialised clinical contexts due to their complexity and resource requirements.

By contrast, our method was designed to test the lower bound of complexity by using only the smooth pursuit task and consumer-grade hardware. While the resulting classification accuracy is lower than those reported in more complex systems, our approach requires no verbal output, no additional cognitive instructions, and minimal participant burden. These characteristics make it well-suited for scalable and cost-effective screening applications, particularly in non-specialist or telemedicine contexts. Another likely reason why our results show lower accuracy is the presence of borderline subjects in our dataset. They account for almost a third of the dataset and are the most difficult to accurately detect being between HC and MCI groups in the CI continuum. This can also be seen in Figure 4. Of course, detecting individuals in the borderline group would be the most rewarding as this is the earliest possible detection.

Our findings also align with smaller studies targeting similar goals. For example, Aloran et al. [21] found significant differences in smooth pursuit amplitude and fixation variability between AD patients and controls using a simplified pursuit task. Although based on a small sample, their results suggest that even basic oculomotor parameters can offer discriminative value when measured in controlled conditions.

5. Discussion

In the present study, we investigated the potential of using smooth pursuit eye movements (SPEM) to distinguish individuals with varying levels of cognitive impairment from healthy controls. Our goal was to design a simple language independent task that encouraged participants to move their eyes as naturally as possible while following a moving object.

Our findings support previous research suggesting that SPEM performance is sensitive to cognitive decline [5,9]. In particular, we found that the mean squared error between the stimulus and the subject's gaze was significantly higher in individuals with cognitive impairment, especially during fast horizontal and slow vertical motion. These findings are consistent with prior studies that have demonstrated impairments in pursuit gain, increased saccadic intrusions, and delayed initiation in individuals with MCI [9] and AD [10].

To better illustrate these differences, we present examples of eye movement traces from two participants during these tasks. Figure 6a,b show the trace of a healthy control, while Figure 6c,d show the trace of an individual with possible Alzheimer's disease (AD).

Comparing the two traces, it is clear that the individual with AD had more difficulties tracking the stimulus accurately. This was especially noticeable in the fast horizontal task (Figure 6c), where the subject struggled to follow the object across the screen. In the slow vertical task (Figure 6d), the trace suggests that the participant tried to predict the stimulus's position and "jumped" with their gaze ahead, rather than smoothly following its path. On the other hand, the healthy subject was mostly able to follow the stimulus smoothly in both tasks (Figure 6a,b). Of course, these are just two typical examples of the eye movement traces we recorded.

Although we found consistent statistical differences between the groups in most of the pursuit tasks (see Figure 3), the overlap in how the results were spread out (see Figure 5b) suggests that SPEM alone might not be enough to clearly tell individuals apart for clinical use. Still, our model built using the logistic regression algorithm shows some promise. The model reached an AUC of 0.675, which is not high enough for diagnosis on its own, but it is a good starting point given how simple and non-invasive the SPEM task is. The ROC

curve showed that the model could pick up some cases of cognitive impairment with few false positives, which means it could be useful as part of a wider screening approach.

Another interesting finding is that the pursuit accuracy worsened as the level of cognitive impairment increased. We saw this gradual change from HC, to borderline cases, to MCI, and to possible AD (see Figure 4). This suggests that SPEM might not only help with identifying cognitive problems, but could also be useful for tracking how the condition progresses or for spotting people at higher risk.

The groups in our study were not age matched. However, the main goal here was to see whether eye-tracking during smooth pursuit tasks could pick up eye movement changes in people with cognitive decline compared to those without it, regardless of age. The study participants were recruited randomly and consecutively, which means we ended up with some age differences—something that was expected, since cognitive decline tends to occur more with age.

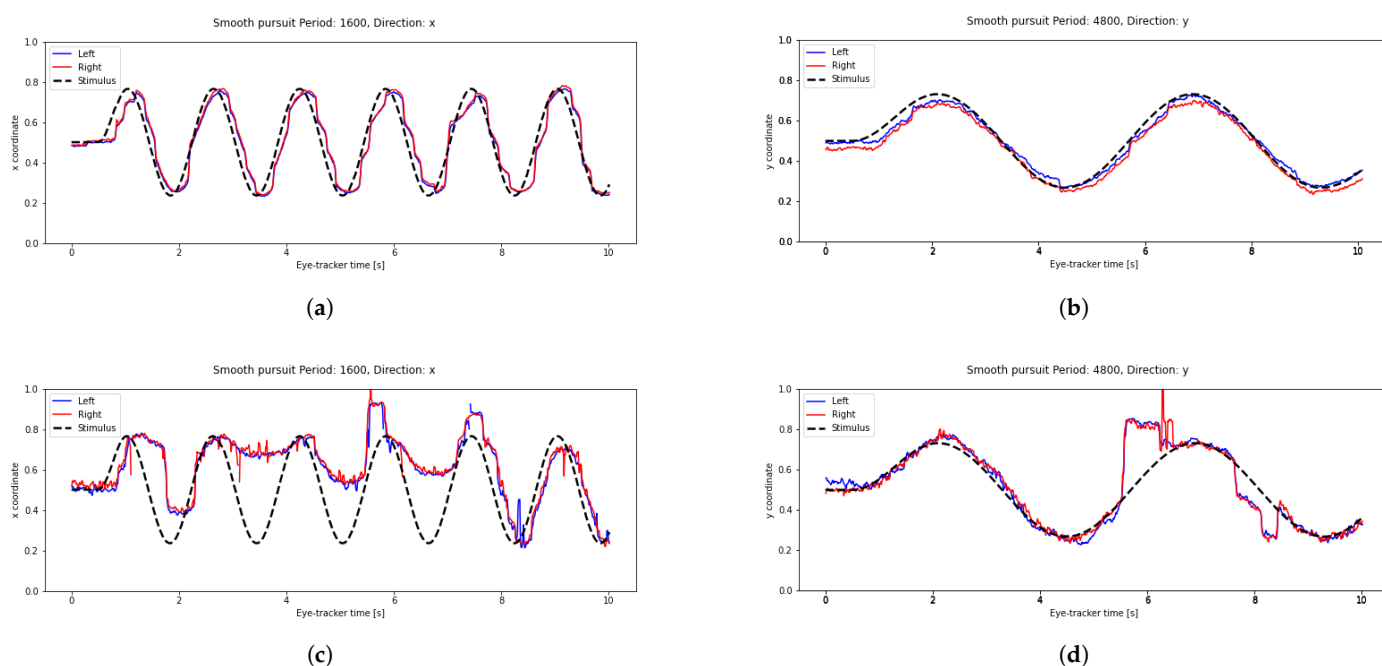


Figure 6. Traces from the most important smooth pursuit tasks performed by a subject without cognitive impairment (Figures (a,b)) and a subject with possible AD (Figures (c,d)). The black dotted lines represent the movement of the stimulus on the screen, while the red and blue lines show the eye movement traces of the right and left eyes, respectively. (a) Trace for the fast horizontal smooth pursuit task (period 1600 ms, frequency ≈ 0.625 Hz) performed by HC subject. (b) Trace for the slow vertical smooth pursuit task (period 4800 ms, frequency ≈ 0.21 Hz) performed by HC subject. (c) Trace for the fast horizontal smooth pursuit task performed by AD subject. (d) Trace for the slow vertical smooth pursuit task performed by AD subject.

The machine learning results were modest, which indicates that analysing only the SPEM would not be enough to effectively differentiate CI subjects from HC. However, it shows promise in being one of the possible tasks in a battery of different neuropsychological tasks that could be performed using the eye gaze, like reading [12,22]. In addition, our sample size was fairly small, and because the study was only conducted at one point in time, we cannot say much about how SPEM performance might change over time.

In future iterations of this work, we plan to explore methods for increasing the cognitive load of the smooth pursuit task while retaining the minimalistic design that underpins its feasibility and accessibility. One promising direction is inspired by the dual-task study

presented by Kaye et al. [5], in which participants performed simple arithmetic tasks (solving visually or auditorily presented simple equations) while tracking a moving target. Their results suggest that adding such cognitive demands can increase the task's sensitivity to early-stage impairment, particularly in domains related to working memory and attention. Therefore, incorporating additional cognitive load into the task may improve its ability to detect cognitive impairment at earlier stages.

Inspired by this, we are considering augmenting the current paradigm with non-verbal cognitive load conditions—such as mental arithmetic or pattern recognition tasks—delivered visually and without requiring verbal or manual responses. This would preserve the low-burden contact-free nature of our method while potentially increasing its sensitivity. Importantly, such modifications would be designed to avoid additional stimulus modalities (e.g., auditory input), which may not be feasible in all settings.

Additionally, while our current work focused solely on smooth pursuit, future implementations may incorporate additional simple tasks that could be completed in a single session without excessively increasing the burden. Prior work by Oyama et al. [19] showed that including multiple gaze-based tests targeting various cognitive domains yielded higher classification performance.

Longitudinal data collection and multimodal integration (e.g., combining smooth pursuit metrics with reading- or fixation-based features [12,22]) are also being considered to improve the classification performance. Overall, our goal is to maintain the simplicity and efficiency of the current protocol while gradually enhancing its diagnostic value through minimally intrusive cognitive extensions.

Furthermore, a recent review by Sekar et al. [23] outlines the broader diagnostic potential of oculomotor assessments across neurodegenerative diseases, noting that abnormalities in saccades, fixations, and pursuit movements may each carry disease-specific signatures. Importantly, they emphasise the need to bridge experimental protocols and clinically feasible tools—a gap our current method helps address by using an accessible, scalable design.

6. Conclusions

This study confirmed the hypothesis that eye-movement behaviour during smooth pursuit tasks differs in cognitively impaired subjects and healthy controls. The features we used helped highlight these differences and gave us more insight into how gaze behaviour changes with cognitive decline. We also saw that performance on these tasks tends to worsen as the level of impairment increases—from borderline cases to MCI and suspected AD. This supports the use of eye-tracking in this context and suggests that detecting cognitive issues through eye movements becomes more likely as the condition progresses.

Our machine learning results show that the SPEM-based features introduced here could be a useful way to detect early signs of cognitive decline through a simple, affordable, and non-invasive test. The somewhat modest results suggest that the short task used in our study might not be sufficient as a stand-alone test, but it has the potential to be part of a set of eye-tracking-based tasks for detecting cognitive decline.

Author Contributions: Conceptualization, V.G. and A.S. (Aleksander Sadikov); methodology, V.G., D.G. and A.S. (Aleksander Sadikov); validation, D.G. and A.S. (Aleš Semeja); formal analysis, V.G. and A.D.G.; investigation, V.G., D.G. and A.S. (Aleksander Sadikov); resources, V.G., A.S. (Aleš Semeja) and A.S. (Aleksander Sadikov); data curation, A.D.G. and A.S. (Aleksander Sadikov); writing—original draft preparation, V.G. and A.D.G.; writing—review and editing, A.S. (Aleksander Sadikov), D.G. and A.S. (Aleš Semeja); visualization, V.G. and A.D.G.; supervision, A.S. (Aleksander Sadikov), A.S. (Aleš Semeja) and D.G.; project administration, V.G.; funding acqui-

sition, V.G., A.S. (Aleš Semeja) and A.S. (Aleksander Sadikov). All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding under project NEUS from the European Institute of Innovation and Technology (EIT) Health KIC. This body of the European Union receives support from the European Union’s Horizon 2020 research and innovation programme. The research was also partially supported by the Slovenian Research and Innovation Agency under the research programme Artificial intelligence and intelligent systems, grant no. P2-0209.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the National Medical Ethics Committee of the Republic of Slovenia (approval numbers: 0120-400/2015-5 dated 2 April 2016 and 0120-400/2015/9 dated 22 May 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available from NEUS Diagnostics d.o.o., but restrictions apply to the availability of these data, which were used under license for the current study and are not publicly available. Processed data will however be available from the corresponding author upon reasonable request and with permission of NEUS Diagnostics d.o.o., following an embargo from the date of publication to allow for commercialisation of the research findings.

Acknowledgments: The authors would like to thank all the neurologists, psychologists, technicians, and administrative support staff who were responsible for patient onboarding and data collection. Parts of the text were edited for clarity and grammar with the assistance of ChatGPT-4o-2025-04-30, an AI language model developed by OpenAI. The authors reviewed and approved all content.

Conflicts of Interest: V.G., A.S. (Aleš Semeja), and A.S. (Aleksander Sadikov) are co-owners of NEUS Diagnostics d.o.o., A.D.G. was employed by NEUS Diagnostics d.o.o. as a researcher, and D.G. declares no potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AD	Alzheimer’s Dementia
CI	Cognitive Impairment
MCI	Mild Cognitive Impairment

References

- Jongsiriyanyong, S.; Limpawattana, P. Mild cognitive impairment in clinical practice: A review article. *Am. J. Alzheimer’s Dis. Other Dement.* **2018**, *33*, 500–507. [[CrossRef](#)] [[PubMed](#)]
- Flicker, C.; Ferris, S.H.; Reisberg, B. Mild cognitive impairment in the elderly. *Neurology* **1991**, *41*, 1006. [[CrossRef](#)] [[PubMed](#)]
- Farias, S.T.; Mungas, D.; Reed, B.R.; Harvey, D.; DeCarli, C. Progression of mild cognitive impairment to dementia in clinic-vs community-based cohorts. *Arch. Neurol.* **2009**, *66*, 1151–1157. [[CrossRef](#)] [[PubMed](#)]
- Mostafa, A.; Tiu, S.; Khan, F.; Baig, N.A. The Efficacy of Anti-amyloid Monoclonal Antibodies in Early Alzheimer’s Dementia: A Systematic Review. *Ann. Indian Acad. Neurol.* **2025**, 10–4103. [[CrossRef](#)] [[PubMed](#)]
- Kaye, G.; Johnston, E.; Burke, J.; Gasson, N.; Marinovic, W. Differential Effects of Visual and Auditory Cognitive Tasks on Smooth Pursuit Eye Movements. *Psychophysiology* **2025**, *62*, e70069. [[CrossRef](#)] [[PubMed](#)]
- Jin, Z.; Gou, R.; Zhang, J.; Li, L. The role of frontal pursuit area in interaction between smooth pursuit eye movements and attention: A TMS study. *J. Vis.* **2021**, *21*, 11. [[CrossRef](#)] [[PubMed](#)]
- Diotaiuti, P.; Marotta, G.; Di Siena, F.; Vitiello, S.; Di Prinzio, F.; Rodio, A.; Di Libero, T.; Falese, L.; Mancone, S. Eye Tracking in Parkinson’s Disease: A Review of Oculomotor Markers and Clinical Applications. *Brain Sci.* **2025**, *15*, 362. [[CrossRef](#)] [[PubMed](#)]
- Popovic, Z.; Gilman Kuric, T.; Rajkovic Latic, I.; Matosa, S.; Sadikov, A.; Groznic, V.; Georgiev, D.; Tomic, S. Correlation between non-motor symptoms and eye movements in Parkinson’s disease patients. *Neurol. Sci.* **2025**, *46*, 1–9. [[CrossRef](#)] [[PubMed](#)]
- Lin, J.; Xu, T.; Yang, X.; Yang, Q.; Zhu, Y.; Wan, M.; Xiao, X.; Zhang, S.; Ouyang, Z.; Fan, X.; et al. A detection model of cognitive impairment via the integrated gait and eye movement analysis from a large Chinese community cohort. *Alzheimer Dement.* **2024**, *20*, 1089–1101. [[CrossRef](#)]

10. Tao, M.; Cui, L.; Du, Y.; Liu, X.; Wang, C.; Zhao, J.; Qiao, H.; Li, Z.; Dong, M. Analysis of eye movement features in patients with Alzheimer’s disease based on intelligent eye movement analysis and evaluation system. *J. Alzheimer Dis.* **2024**, *102*, 1249–1259. [[CrossRef](#)]
11. Conradsson, M.; Rosendahl, E.; Littbrand, H.; Gustafson, Y.; Olofsson, B.; Lövhelm, H. Usefulness of the Geriatric Depression Scale 15-item version among very old people with and without cognitive impairment. *Aging Ment. Health* **2013**, *17*, 638–645. [[CrossRef](#)] [[PubMed](#)]
12. Groznic, V.; Možina, M.; Lazar, T.; Georgiev, D.; Semeja, A.; Sadikov, A. Validation of reading as a predictor of mild cognitive impairment. *Sci. Rep.* **2025**, *15*, 12834. [[CrossRef](#)] [[PubMed](#)]
13. Mioshi, E.; Dawson, K.; Mitchell, J.; Arnold, R.; Hodges, J.R. The Addenbrooke’s Cognitive Examination Revised (ACE-R): A brief cognitive test battery for dementia screening. *Int. J. Geriatr. Psychiatry J. Psychiatry Late Life Allied Sci.* **2006**, *21*, 1078–1085. [[CrossRef](#)] [[PubMed](#)]
14. Slachevsky, A.; Villalpando, J.M.; Sarazin, M.; Hahn-Barma, V.; Pillon, B.; Dubois, B. Frontal assessment battery and differential diagnosis of frontotemporal dementia and Alzheimer disease. *Arch. Neurol.* **2004**, *61*, 1104–1107. [[CrossRef](#)] [[PubMed](#)]
15. Bowie, C.R.; Harvey, P.D. Administration and interpretation of the Trail Making Test. *Nat. Protoc.* **2006**, *1*, 2277–2281. [[CrossRef](#)] [[PubMed](#)]
16. Folstein, M.F.; Folstein, S.E.; McHugh, P.R. “Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **1975**, *12*, 189–198. [[CrossRef](#)] [[PubMed](#)]
17. Association, A.P.; Force, D.T. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5™*, 5th ed.; American Psychiatric Publishing, Inc.: Washington, DC, USA, 2013. [[CrossRef](#)]
18. Holmqvist, K.; Andersson, R. *Eye-Tracking: A Comprehensive Guide to Methods, Paradigms and Measures*; Oxford University Press: Oxford, UK, 2017.
19. Oyama, A.; Takeda, S.; Ito, Y.; Nakajima, T.; Takami, Y.; Takeya, Y.; Yamamoto, K.; Sugimoto, K.; Shimizu, H.; Shimamura, M.; et al. Novel Method for Rapid Assessment of Cognitive Impairment Using High-Performance Eye-Tracking Technology. *Sci. Rep.* **2019**, *9*, 12932. [[CrossRef](#)] [[PubMed](#)]
20. Li, N.; Wang, Z.; Ren, W.; Zheng, H.; Liu, S.; Zhou, Y.; Ju, K.; Chen, Z. Enhancing Mild Cognitive Impairment Auxiliary Identification Through Multimodal Cognitive Assessment with Eye Tracking and Convolutional Neural Network Analysis. *Biomedicine* **2025**, *13*, 738. [[CrossRef](#)] [[PubMed](#)]
21. Aloran, S.; Nakhleh, S.; Alrhman, I.A. Developing a Non-Invasive Eye Tracking Screening Tool for early Detection of Alzheimer’s Disease. In Proceedings of the 2023 2nd International Engineering Conference on Electrical, Energy, and Artificial Intelligence (EICEEAI), Sirnak, Turkey, 25–26 December 2023 ; pp. 1–5. [[CrossRef](#)]
22. Groznic, V.; Možina, M.; Lazar, T.; Georgiev, D.; Sadikov, A. Gaze behaviour during reading as a predictor of mild cognitive impairment. In Proceedings of the 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Viartual, 27–30 July 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–4. [[CrossRef](#)]
23. Sekar, A.; Panouillères, M.T.; Kaski, D. Detecting Abnormal Eye Movements in Patients with Neurodegenerative Diseases—Current Insights. *Eye Brain* **2024**, *16*, 3–16. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.