Investigating the Impact of Parkinson's Disease on Brain Computations: An Online Study of Healthy Controls and PD Patients

Artur Chudzik[1(✉)], Aldona Drabik[1], and Andrzej W. Przybyszewski[1,2]

[1] Faculty of Computer Science, Polish-Japanese Academy of Information Technology, Warsaw, Poland
{artur.chudzik,adrabik,przy}@pjwstk.edu.pl
[2] Department of Neurology, University of Massachusetts Medical School, 65 Lake Avenue, Worcester, MA 01655, USA
andrzej.przybyszewski@umassmed.edu
https://pja.edu.pl , https://nd.pja.edu.pl

Abstract. Early detection of Parkinson's disease (PD) is critical for effective management and treatment. In our recent study, we collected data on brain computations in individuals with PD and healthy controls using an online platform and multiple neuropsychological tests. Using logistic regression, we achieved an accuracy rate of 91.1% in differentiating PD patients and healthy controls. However, two PD patients were classified as healthy subjects, and two healthy individuals were misclassified as PD patients. We also utilized multinomial logistic regression to predict the UPDRS3 group of patients and healthy individuals, achieving the same high accuracy. Our findings suggest that cognitive and behavioral tests can detect early changes in brain computations, potentially indicating the onset of PD before clinical symptoms appear. This has significant implications for early detection and intervention of neurological disorders, improving outcomes and quality of life for affected individuals. Overall, our study provides new insights into the utility of neuropsychological tests and statistical methods for detecting and monitoring PD.

Keywords: Parkinson's disease · Brain computations · Online testing

1 Introduction

This study focuses on the test that detects early neurological symptoms of major public health problems related to neurodegenerative diseases (ND). NDs are incurable and debilitating conditions that result in progressive degeneration and death of nerve cells. The process starts with an asymptomatic stage when the person feels fine and shows no signs of neurodegenerative disease, and clinical examination also will show no abnormalities. During the asymptomatic stage of neurodegenerative disease, individuals may not exhibit any symptoms and may not seek medical attention, leading to a lack of abnormalities detected during

N. T. Nguyen et al. (Eds.): ACIIDS 2023, LNAI 13996, pp. 235–246, 2023.
https://doi.org/10.1007/978-981-99-5837-5_20

clinical examination. All NDs are progressing relentlessly over the years and have proved to be stubbornly incurable. Thus, we believe that it is crucial to find a way to detect the early onset of NDs.

1.1 Alzheimer's and Parkinson's Disease

The most common neurodegenerative disorders are Alzheimer's disease (AD) and Parkinson's disease (PD) [4], and they are in our area of interest because of their partial similarities in symptoms. First, it is necessary to provide a comprehensive explanation of Parkinson's disease as our patients are afflicted with this condition. Parkinson's disease is a progressive neurodegenerative disorder that begins to develop approximately 20 years prior to the appearance of symptoms, during which a significant portion of the brain is already affected. It is characterized by the progressive loss of dopaminergic neurons in the brain, leading to a range of motor and non-motor symptoms, including tremors, rigidity, and cognitive impairment. Early detection of Parkinson's disease is crucial for the effective management and treatment of the condition. PD affects three fundamental systems: motor, cognitive, and emotional. The disease typically starts with motor impairments such as bradykinesia. In contrast, Alzheimer's disease initially affects cognitive abilities like mild cognitive impairment (MCI), which not all PD patients have. Although MCI is not a typical symptom of Parkinson's disease, it can occur during the disease progression as an early stage where symptoms are not severe but are detectable. As the disease advances, individuals may experience late-stage complications, including advanced cognitive and motor symptoms. It is important to note that preclinical symptoms **may vary** between Alzheimer's disease and Parkinson's disease, the presentation and progression of symptoms **can vary widely** between individuals, and **no two cases of PD are exactly alike**. In summary, both AD and PD characteristics **may include** the following common manifestation [3,14,18].

- Mild cognitive impairment (MCI), such as language or visuospatial perception and memory impairment (mainly in AD, not always in PD);
- affected rote memory and frontotemporal executive functions;
- depression;
- sleep problems;
- automatic response inhibition decay;
- difficulty with emotion recognition;
- motor slowness symptoms (predominantly with PD, but also associated with preclinical AD).

1.2 Digital Biomarkers

In this study, we place emphasis on the use of digital biomarkers, which are measurable and quantifiable medical signs collected through digital devices or platforms, to gain insights into brain computations. Reaction time (RT) is a widely studied digital biomarker that plays a crucial role in measuring neurological function, particularly in individuals with PD. Previous research, including our own studies [14],

has demonstrated that measures such as saccadic delay and movement-related potentials can serve as reliable indicators of the state of PD. Others [10] have employed movement-related potentials in a choice reaction time task to explore the underlying causes of reaction time delay in Parkinson's disease. Movement-related potentials showed that motor processes required more time for Parkinson's disease patients making complex responses. The study also found that one or more premotor processes were slowed in Parkinson's disease patients based on delayed onset of movement-related potentials. These findings suggest that reaction time may be a valuable measure for tracking the progression of Parkinson's disease and the effectiveness of treatment. It is worth noting that reaction time is just one measure of neurological function that can be used with other measures. It may also help evaluate the impact of Parkinson's disease on the nervous system.

1.3 Digital Screening

We utilized an online platform to administer neuropsychological tests, which are widely used as the gold standard for assessing cognitive function. Our aim was to detect early changes in brain computations in individuals with PD, which could indicate the onset of the disease before the appearance of clinical symptoms. In addition to the participants' responses, we also collected additional temporal measures, including Instrumental Reaction Time (IRT) and Time-to-Submit - TTS. IRT measures the time between the screen appearing and the participant's first option selection, while TTS measures the time it takes for the participant to click the submit button. In the following text, TTS is also referred to as "response time". It is well recognized that neuropsychological testing has great diagnostic and screening power, but it requires proper training, tools, and time. Our goal is to evaluate if a single online tool can support these operations, and thus provide a cross-sectional set of neuropsychological examinations that will contribute to the overall understanding of the patient's psychophysical state. To ensure the validity of our results, we recruited both a group of individuals diagnosed with PD and a group of healthy controls for evaluation. While it is well-established that reaction time generally decreases with age, with previous studies estimating an average decrease of 4-10 ms per year [2,17], our observations of instrumental reaction times in our study were higher than expected based on age-related decline alone. Despite an average age difference of approximately 47 years between the groups, as detailed in the "Results" section, our findings suggest that factors beyond age-related decline may have contributed to the observed differences in reaction time between the PD and healthy control groups.

2 Methods

We intended to create an online method of neuropsychological assessment. The implementation in the form of a computer test started with the requirements gathering and prototype. First, we asked trained psychologists and neurologists to create a general overview of the battery of tests used in their practice, being a gold standard. We decided to use a multi-tiered approach to assessment, including the tests described below.

2.1 GDS-15

Geriatric Depression Scale - is a short version (15 questions) of the test developed in '86 by Sheikh and Yesavage [16]. The short version contains 15 of the 30 questions from the extended version that showed the most significant correlation with signs of depression. Out of 15 items, 10 indicate the presence of depression when given a positive answer, while the remaining items (questions 1, 5, 7, 11, 13) indicate depression when given a negative answer. Scores 0–4 are considered "normal" depending on age and education; 5–8 indicate mild depression; 9–11 indicate moderate depression and 12–15 indicate severe depression. The GDS has 92% sensitivity and 89% specificity as assessed by diagnostic criteria [8]. A validation study comparing long and short GDS forms to self-assessment of depressive symptoms successfully distinguished adults with depression from non-depressed people with a high correlation ($r = 0.84$, $p < 0.001$) [16]. The online implementation of the study in our version consists of 15 questions, displayed individually, with a single choice option between "yes" or "no". The sample question from this set is: "Have you dropped many of your activities and interests?" Every question is consistent with the official translation of the test in a selected language version. The test in its standard form consists of questions and answers printed on a single A4 sheet, and therefore it is possible to resolve it non-linearly. Our version shows each question separately, which allows us to measure both reaction (instrumental reaction time - IRT) and response (time-to-submit - TTS) time.

2.2 TMT A&B

Trail Making Test - is a neuropsychological test for visual attention and task switching, developed in '55 by Reitan [15]. It consists of two parts. In both, instruction to the subject is to connect a set of dots as quickly as possible while maintaining accuracy. The test can provide information on visual search speed, scanning, processing speed, mental flexibility, and executive functioning. The TMT A and B results are as high as the number of seconds to complete the task; higher scores follow the level of impairment. In part A with 25 dots - a healthy person can finish it on average in 29 s, and a patient with deficiencies in more than 78 s. In part B with 25 dots - a healthy person can finish it on average in 75 s, and a patient with deficiencies in more than 273 s. The standard form test asks to combine tracks 1-2-3- (version A) or 1-A-2-B- (version B) on the paper with a pen on the paper tray. We ask patients to select circles in a given order three times in the online version. First test: version A relies on 15 circles, and this part focuses mainly on examining cognitive processing speed. Second: version B (short) consists of 10 circles (5 with letters and 5 with numbers), and version B (long) is 20 circles (10 for both letters and numbers). These versions assess executive functioning. Each time we allocate circles randomly with uniform distribution on the screen. It is worth mentioning that there is no record of the error rate in the pen and paper version of the test. Because the online version is self-assessed, we had to implement this feature and notify the

user of making a mistake by marking the circle in red and the correct connection displayed as a green circle. This mechanism gives the users feedback to get on the right path themselves. The completion, however, might be longer than in the standard version since there is no supervisor. Here, we record the error rate for each part of that task, IRT and TTS.

2.3 CDT and CCT

The Clock Drawing Test - is used to screen cognitive disorders in clinical practice. The origins of this test are not clear, but the probable precursor was Sir Henry Head [5,6]. There are many ways to conduct this test, but a common task is to draw a clock with a face, all numbers, and hands showing a given time. One way is to draw two lines perpendicular to each other, obtaining four quadrants of the clock's face. Then, we can count the number of digits in each quadrant, and if the quadrant is correct while it contains three numbers (error score is between 0 and 3 for each quadrant). The standard score is below 4 points. In the original study, a score over 4 revealed a sensitivity of 87% and a specificity of 82% for identifying dementia. Another test related to drawing tasks might be CCT - the Cube Copying test, valid (yet limited) for routine clinical dementia screening. As presented in Maeshima et al. [11], quantitatively scored cube copying can estimate cognitive dysfunction in dementia patients. The execution of both tasks in the digital form relied on the area of the screen divided by opaque lines, mimicking a standard notebook. Participants drew a figure with a cursor or a finger on mobile devices. We were concerned about the performance of older patients who were not fluent with computer technology because drawing on the computer screen introduces a novelty factor. Also, this interface lacks the naturalness of the pen-and-paper method. However, most users completed both assignments. Those tasks were not time-restricted. Nevertheless, we recorded IRT and TTS as well as the paintings.

2.4 MoCA

Montreal Cognitive Assessment - is a screening test developed in '05 by Nasreddine et al. [12] is a cognitive test to detect MCI. The test checks language, memory, visual and spatial thinking, reasoning, and orientation. MoCA scores range from 0 to 30, and 26 or more is considered normal. In the original study, people without cognitive impairment scored 27.4 (average); subjects with mild cognitive impairment (MCI) scored 22.1 (average); people with Alzheimer's disease scored 16.2 (average). The test has a 90.0–93.0% sensitivity and a specificity of 87.0% in the MCI assessment. MoCA implements three earlier tasks: TMT B, CCT, and CDT. The following tasks are related to language fluency. First: "Name this animal". We depict a cow, horse, and a lion, and we ask the participant to type the name of the presented animal into the text field. Additionally, our task depicts the lion with the incorrect number of legs because this disturbance seems to be a response time delay factor in a patient who suffers from AD. The second task from this series is a repetition of two syntactically complex sentences: "I only know that John is the one to help today". and "The cat always hid under

the couch when dogs were in the room". We asked patients to replicate both sentences in a written form, disabling the copy-paste option. The third language fluency task was to write as many English words as possible that start with the letter F. The patient had 60 s for execution. Considering that older participants are less fluent in typing, we introduced two mechanisms that could align their chances. First, we delayed the countdown by the number of seconds calculated as (number of words in the task * 60/average reading speed per minute), assuming that the lower boundary of the average reading speed is 200 words per minute. Next, each keystroke stopped the countdown, allowing writing the whole word even at a slow pace. Lack of the keystroke during the next consequent 3 s was starting the countdown again. The next group of tasks focuses on attention and concentration. First, we display one letter per second, and a person has to click the button each time the letter "A" shows on the screen; next, we ask about the serial subtraction starting at 100. Likewise, we present two sets of numbers; each time, the subject must repeat them by writing in the forward or backward order to evaluate the working memory. We measure the error rate and average response time for all tasks. Also, we assess the abstract reasoning by a describe-the-similarity task. We ask about what two pairs of words have in common (in a single word): watch + ruler and train + bike, and we evaluate answers alongside limited dictionaries of means of transportation, traveling, measuring, and instruments. Here also we measure the error rate, IRT, and TTS. The next part focuses on short-term memory. We involved two learning trials of five nouns and delayed recall after approximately five minutes. For the first trial, the patient must write words and receive visual cues if they are correct. If not, it is possible to rewind and see them again. If this operation fails more than twice, we save this fact into the database, skipping into the next question. We display this task again at the end of the MoCA part. Each time, we count the error rate, IRT, and TTS. Finally, we evaluate the spatio-temporal orientation by asking the subject for the date and where the test occurs. We validate the provided year, month, exact date, and day of the week with the system clock, counting the number of errors. The place is scored manually after the test. For each part, we measure the instrumental reaction time and Time-to-Submit. We are aware of an essential, fundamental difference in switching from a verbal task (hears -> speaks) to a written form (sees -> writes), especially when taking into consideration motor problems (writing), leading to a field of uncertainty that we must treat with utmost meticulousness.

2.5 Epworth

Epworth Sleepiness Scale - is an eight questions test that focuses on daytime sleepiness, created in '91 by Johns [7]. On a 4-point scale (0-3), subjects are asked to rate the likelihood of falling asleep during eight different activities throughout the day. The Epworth score (sum of 8 responses, each scored on a 0-3 scale) can range from 0 to 24. The higher the Epworth score, the higher the average tendency to "daytime sleepiness". The test showed 93.5% sensitivity and 100% specificity in the narcolepsy diagnosis study [9]. In our online version

of this test, we ask participants to determine the likelihood of falling asleep in multiple situations, such as "Sitting and reading" or "Watching TV". Possible answers are: "zero probability of falling asleep", "unlikely to fall asleep", "average probability of falling asleep", and "high probability of falling asleep". Each answer has 0-3 points accordingly. Here, we display each question with possible answers separately, each time measuring IRT and TTS.

2.6 FER

Facial Emotion Recognition - is a set of tests dedicated to recognizing emotions conveyed through different channels, where one of them is to match a label with a given emotional expression. Multiple studies suggest that the results of patients with PD are performing significantly worse than that of healthy controls [1]. The link between facial expression and FER impairment reveals since the earliest studies on FER in recall embodied simulation theory, suggesting that disturbed motor processing can lead to deficiency in emotion recognition. We decided to implement this task with six faces expressing particular emotions, alongside six radio buttons with emotions' names. We presented each face separately and obtained all of them from the "Warsaw set of emotional facial expression pictures" (WSEFEP) [13]. Each face presented anger, disgust, fear, happiness, sadness, or surprise. We selected those pictures with the highest recognition marks (e.g., accuracy with intended display) from independent judges. The test evaluates the correctness of the answer, IRT, and TTS for each displayed expression.

2.7 Online Study

To conclude, we distinguished 66 questions requiring various forms of responses, and we implemented them as web application components. Computer assessment allowed us to extend classical metrics: each question could hold a precise IRT and TTS along with the answer. Measuring time on the client-side is crucial for assessing the performance of participants. One widely used method for measuring these metrics is the JavaScript method `performance.now()`, which provides a high-resolution timestamp in milliseconds. Unlike other methods that rely on the system clock, `performance.now()` is not affected by changes to the system clock and provides a more accurate representation of the time it takes for code to execute. `performance.now()` returns a DOMHighResTimeStamp value that represents the number of milliseconds elapsed since the performance timing origin, which is typically the time the page was loaded or refreshed. This method is often used in conjunction with other JavaScript functions, such as `setTimeout()` and `requestAnimationFrame()`, to measure the time it takes for code to execute and to optimize performance. In our study, we utilized this method to measure the time to first selection (IRT), and Time-to-Submit (TTS). The user interface of the application was implemented using the React JavaScript library, which is widely used for building modern, scalable, and interactive web applications.

3 Results

The present study aimed to investigate the effects of Parkinson's Disease (PD) on brain computations using an online platform. Temporal values (IRT and TTS) were recorded in milliseconds, but for improved legibility and comprehension, the results are presented in seconds. Both IRT and TTS are averages (calculated without outliers) based on partial measurements of single questions. To determine the statistical significance of group differences, p-values were calculated and comparisons were considered statistically significant if the p-value was less than 0.05. Statistical analyses were conducted using SPSS 29.

3.1 Comparison of Cognitive and Sleep-Related Measures

A total of 45 participants were recruited for this study, with 15 PD patients (8 females, 7 males) with a mean age of 70.8 years (standard deviation [SD] = 5.93) and 30 healthy controls (3 females, 27 males) with a mean age of 24 years. The selection of participants was based on the availability of individuals who met the criteria for each group. While the age difference between the PD and healthy control groups was noticeable, it is important to note that age was not utilized as a variable in the machine learning analysis, making it less relevant to the study objectives. The severity of motor symptoms in patients with Parkinson's disease was assessed using the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (UPDRS) Part III. Patients were grouped into five categories based on their UPDRS3 scores: Group 0 (score 0–9) n = 0, Group 1 (10–19) n = 5, Group 2 (20–29) n = 3, Group 3 (30–39) n = 3, Group 4 (40+) n = 4. All healthy controls were classified into Group 0 (n = 30). First, we found that the PD patients had a slightly lower mean MOCA score (24.67, SD = 3.519) than healthy controls (26.27, SD = 1.202), but the difference was not significant (p = 0.107).

Similarly, the PD patients had a slightly higher mean Epworth score (5.13, SD = 1.685) than healthy controls (4.13, SD = 1.776), but the difference was still not significant (p = 0.077). Also, we found that the mean GDS15 score for PD patients (5.60, SD = 1.056) was slightly lower than that of healthy controls (6.43, SD = 1.305), but again the difference was not significant (p = 0.38). Finally, the mean FER score for both groups was also not significantly different (PD patients: 6.87, SD = 0.352; healthy controls: 6.77, SD = 0.626) (p = 0.57). These findings suggest that there was no significant impairment in facial expression recognition in the PD group compared to the healthy control group. We found, however, that there are significant differences between the scores of TMT B (PD patients: 4.54, SD = 7.70; healthy controls: 0.67, SD = 1.20) (p = 0.043).

3.2 Temporal Results in Cognitive Tests

We also measured the participants' IRT and TTS for each cognitive tests' question. In the MoCA test, the mean instrumental reaction time for the healthy group was significantly faster (3.62 s) than the clinical group (5.90 s) ($p < 0.001$). Similarly, the healthy group also had a significantly faster Time-to-Submit (8.00 s) compared to the clinical group (13.67 s) ($p < 0.001$). The same

pattern was observed in the Epworth Sleepiness Scale test, where the healthy group had a significantly faster mean instrumental reaction time (4.70 s) and TTS (6.45 s) compared to the clinical group (8.57 s and 10.45 s, respectively) ($p < 0.001$).

In contrast, there was no significant difference between the healthy and clinical groups in instrumental reaction time and response time in the Geriatric Depression Scale (GDS-15) test. The mean instrumental reaction time for the healthy group was 4.57 s, and for the clinical group, it was 5.82 s. The Time-to-Submit for the healthy group was 6.58 s, and for the clinical group, it was 7.18 s. We also measured the participants' IRT and TTS in the Facial Expression Recognition (FER) task. The mean instrumental reaction time for the healthy group was significantly faster (3.49 s) than the clinical group (5.23 s) ($p < 0.001$). However, there was no significant difference between the healthy and clinical groups in the TTS (6.06 s for the healthy group and 6.74 s for the clinical group).

3.3 Predicting Health Status with Cognitive and Emotional Measures

In our study, logistic regression was employed as the statistical method to predict the binary outcome of a patient's health status based on cognitive and emotional measures. Logistic regression models the probability of the binary outcome by applying a logistic function, which transforms a linear combination of the predictor variables. It is a widely used method in machine learning and particularly suitable when the dependent variable is categorical. The logistic regression model in our study utilized default parameter values, including the probabilities of inclusion (PIN = 0.05) and exclusion (POUT = 0.10), as well as a tolerance value (TOLERANCE = 0.0001) to assess multicollinearity. The PIN represents the probability that a variable will be included in the model, while the POUT represents the probability of excluding a variable. The tolerance value indicates the degree of multicollinearity, with a lower value indicating a higher degree of correlation among predictor variables, which can affect the interpretation of regression coefficients.

The results of our experiment showed promising findings in terms of differentiating PD patients and healthy controls based on cognitive and behavioral tests. Our initial attempt to detect healthy controls using only the MOCA score resulted in a 77.8% accuracy rate. However, when we included additional tests such as the Epworth Sleepiness Scale, and Geriatric Depression Scale, the accuracy dropped to 73.3%. Moreover, adding FER score parameter had no impact on this value. It is noteworthy that the inclusion of instrumental reaction time measurements in the MOCA test resulted in a significant increase in accuracy rate to 84.4%, indicating their potential in PD detection. Additionally, combining the results of all tests with IRT for MoCA resulted in a high accuracy rate of 91.1% with a sensitivity of 86.67% and a specificity of 93.33%, underscoring the significance of employing a combination of cognitive and behavioral tests in conjunction with IRT to enhance accuracy and establish a possible digital biomarker for early detection of the disease.

Table 1. Classification Results of Multinomial Logistic Regression using TMT B, IRT, and TTS Measures.

Observed	Predicted					
	G0	G1	G2	G3	G4	% Correct
G0	30	0	0	0	0	100.0
G1	1	4	0	0	0	80.0
G2	1	1	1	0	0	33.3
G3	0	0	0	3	0	100.0
G4	1	0	0	0	3	75.0
Overall Percentage	**73.3**	**11.1**	**2.2**	**6.7**	**6.7**	**91.1**

3.4 Predicting Parkinson's Disease Severity with TMT B and Temporal Measures

We utilized multinomial logistic regression to predict the UPDRS3 group of both PD patients and healthy controls based on their TMT B scores, IRT, and TTS measures. Multinomial logistic regression is a statistical method used to predict categorical outcomes with more than two categories. In our case, patients were grouped into five categories based on their UPDRS3 scores, with healthy controls classified as Group 0. The model was implemented with maximum iterations set to 100, maximum step halving set to 5, and log-likelihood and parameter convergence set to 0. Our analysis showed that using only TMT B score and IRT, we achieved an accuracy of 82.2% in predicting the UPDRS3 group. However, when TTS was added to the model, the overall accuracy increased to 91.1% (Table 1). These results suggest that TMT B error rate, IRT, and TTS might be reliable measures for predicting the UPDRS3 group of patients with PD.

4 Discussion

The primary goal of our research group is to investigate new and innovative ways to detect and diagnose neurodegenerative diseases, such as Parkinson's Disease and Alzheimer's Disease, as early as possible. Early detection is essential because it allows for timely interventions, potentially leading to improved outcomes and quality of life for affected individuals.

In our latest study, we investigated the effects of PD on brain computations using an online platform. We collected cognitive and behavioral data from PD patients and healthy controls, measuring IRT and TTSs, as well as performance on a battery of cognitive tests. Our findings suggest that cognitive and behavioral tests can be used to detect early changes in brain computations, potentially indicating the onset of PD before clinical symptoms appear. There was no significant difference in the mean Montreal Cognitive Assessment score between the PD patients and the healthy controls. The mean Epworth Sleepiness Scale score was slightly higher in the PD group than in the healthy group, although the

difference was not significant. Our study also revealed that the mean Geriatric Depression Scale (GDS-15) score in the PD group was only marginally lower than in the healthy group, and the difference was not significant. Moreover, we measured the participants' IRT and TTS for each cognitive tests' question. It is worth noting that the PD patients in our study were undergoing treatment with medications which have positive impact on brain computations. Our findings suggest that IRT and TTSs were significantly slower in the PD group compared to the healthy group, particularly in the MoCA and Epworth tests. Interestingly, we found no significant difference between the groups in IRT and TTSs in the GDS-15 test. In the next step we performed a logistic regression analysis to evaluate the effectiveness of our cognitive and behavioral tests in differentiating PD patients and healthy controls, being a first step for early disease detection based on online testing approach. The initial attempt to detect healthy controls using only the MoCA score resulted in a 77.8% accuracy rate. However, when additional tests such as the Epworth Sleepiness Scale and Facial Expression Recognition task were included, together only with MoCA IRT, the accuracy rate increased to 91.1%. This result suggests that a combination of cognitive and behavioral tests may be more effective in identifying early changes in brain computations associated with PD. As a next part of analysis, we performed a multinomial logistic regression analysis to evaluate the effectiveness of our cognitive and behavioral tests in differentiating PD patients and healthy controls. Our decision to focus on the TMT B test was based on its widespread use as a neuropsychological test that has demonstrated high sensitivity in detecting cognitive impairments in PD patients, particularly in attention and executive function domains. The first experiment included only the TMT B score and IRT, resulting in an accuracy rate of 82.2%. We then added TTS to the model, resulting in an increased accuracy rate of 91.1%. These results suggest that adding temporal measures such as IRT and TTS to cognitive tests such as TMT B can improve the accuracy of predicting UPDRS3 group classification. Of course, as with any study, there are limitations to our research. One limitation is the small sample size, which could impact the generalizability of our findings. Furthermore, we only included a limited set of cognitive and behavioral tests in our study. Future research should explore the use of additional tests to improve the accuracy of early detection of PD. Despite these limitations, our study provides evidence that cognitive and behavioral tests can be used to detect early changes in brain computations associated with PD. In extrapolating the results of our study, it is plausible to apply the findings to other neurodegenerative diseases, such as Alzheimer's disease. Similar to PD, early detection of AD is crucial for timely interventions and improved outcomes. Cognitive and behavioral tests, along with measures such as IRT and TTS, can potentially serve as digital biomarkers to detect early changes in brain computations associated with AD. However, further research is necessary to validate the effectiveness of these tests specifically for AD and explore the potential integration of cognitive and behavioral tests with innovative technologies like chatbots to enhance the assessment process. By leveraging digital biomarkers and innovative approaches, we can advance

early detection and diagnostic strategies for various neurodegenerative diseases, ultimately improving patient outcomes and quality of life.

References

1. Argaud, S., Vérin, M., Sauleau, P., Grandjean, D.: Facial emotion recognition in Parkinson's disease: a review and new hypotheses. Mov. Disord. **33**(4), 554–567 (2018)
2. Deary, I.J., Der, G.: Reaction time, age, and cognitive ability: longitudinal findings from age 16 to 63 years in representative population samples. Aging Neuropsychol. Cogn. **12**(2), 187–215 (2005)
3. Goldman, J.G., Aggarwal, N.T., Schroeder, C.D.: Mild cognitive impairment: an update in Parkinson's disease and lessons learned from Alzheimer's disease. Neurodegener. Dis. Manag. **5**(5), 425–443 (2015)
4. Hansson, O.: Biomarkers for neurodegenerative diseases. Nat. Med. **27**(6), 954–963 (2021)
5. Hazan, E., Frankenburg, F., Brenkel, M., Shulman, K.: The test of time: a history of clock drawing. Int. J. Geriatr. Psychiatry **33**(1), e22–e30 (2018)
6. Head, H.: Aphasia and Kindred Disorders of Speech. Cambridge University Press, Cambridge (2014)
7. Johns, M.W.: A new method for measuring daytime sleepiness: the epworth sleepiness scale. Sleep **14**(6), 540–545 (1991)
8. Koenig, H.G., Meador, K.G., Cohen, H.J., Blazer, D.G.: Self-rated depression scales and screening for major depression in the older hospitalized patient with medical illness. J. Am. Geriatr. Soc. **36**(8), 699–706 (1988)
9. Kumar, S., Bhatia, M., Behari, M.: Excessive daytime sleepiness in Parkinson's disease as assessed by Epworth sleepiness scale (ESS). Sleep Med. **4**(4), 339–342 (2003)
10. Low, K.A., Miller, J., Vierck, E.: Response slowing in Parkinson's disease: a psychophysiological analysis of premotor and motor processes. Brain **125**(9), 1980–1994 (2002). https://doi.org/10.1093/brain/awf206
11. Maeshima, S., et al.: Usefulness of a cube-copying test in outpatients with dementia. Brain Inj. **18**(9), 889–898 (2004)
12. Nasreddine, Z.S., et al.: The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. J. Am. Geriatr. Soc. **53**(4), 695–699 (2005)
13. Olszanowski, M., Pochwatko, G., Kuklinski, K., Scibor-Rylski, M., Lewinski, P., Ohme, R.K.: Warsaw set of emotional facial expression pictures: a validation study of facial display photographs. Front. Psychol. **5**, 1516 (2015)
14. Przybyszewski, A.W., Śledzianowski, A., Chudzik, A., Szlufik, S., Koziorowski, D.: Machine learning and eye movements give insights into neurodegenerative disease mechanisms. Sensors **23**(4), 2145 (2023)
15. Reitan, R.M.: The relation of the trail making test to organic brain damage. J. Consult. Psychol. **19**(5), 393 (1955)
16. Sheikh, J.I., Yesavage, J.A.: Geriatric depression scale (GDS): recent evidence and development of a shorter version. Clin. Gerontol. J. Aging Mental Health (1986)
17. Thompson, J.J., Blair, M.R., Henrey, A.J.: Over the hill at 24: persistent age-related cognitive-motor decline in reaction times in an ecologically valid video game task begins in early adulthood. PLoS ONE **9**(4), e94215 (2014)
18. Zokaei, N., Husain, M.: Working memory in Alzheimer's disease and Parkinson's disease. In: Processes of Visuospatial Attention and Working Memory, pp. 325–344 (2019)