Recognizing Patterns of Parkinson's Disease Using Online Trail Making Test and Response Dynamics – Preliminary Study

Artur Chudzik[1], Jerzy P. Nowacki[1], and Andrzej W. Przybyszewski[1,2(✉)]

[1] Polish-Japanese Academy of Information Technology, Faculty of Computer Science, 86 Koszykowa Street, 02-008 Warsaw, Poland
przy@pja.edu.pl

[2] Department of Neurology, UMass Chan Medical School, 65 Lake Avenue, Worcester, MA 01655, USA

Abstract. Neurodegenerative diseases (NDs), including Parkinson's (PD) and Alzheimer's (AD) disease are devastating conditions that affect millions worldwide, with the number of cases expected to rise significantly in the coming years. Despite considerable advancements in understanding their pathophysiology, etiology, and treatment, there is still a lack of effective disease-modifying interventions. Currently, no cure exists and there is an urgent need for modern tools that allow precise detection and objective severity scoring for the development of new therapeutic targets and approaches. Therefore, this study evaluates the effectiveness of an online version of the Trail Making Test Part A and B (TMT A and TMT B), incorporating time-based measures, to recognize cognitive and motor manifestations of Parkinson's disease severity. For validation, this research was conducted with 15 Parkinson's patients under care at UMass Chan Medical School. This study applied the TMT sensitivity to executive function impairments by measuring response and reaction times, to correlate these with stages of PD severity. Machine learning models (Naïve Bayes, Logistic Regression, Support Vector Machine, and Random Forest) were used to predict the disease severity based on TMT performance. Among these, Random Forest was the most effective, achieving scores with an Area Under the Curve (AUC) of 0.92 (80% accuracy), indicating good performance in distinguishing between mild and advanced stages of PD. Although limited by a small sample size, this preliminary study highlights the role of digital tools in enhancing PD diagnostics and monitoring. Future research with larger cohorts and longitudinal designs is essential to validate these preliminary findings and further develop digital diagnostics as crucial in the fight against neurodegenerative diseases.

Keywords: Machine Learning (ML) · Trail Making Test (TMT) · Diagnostic Accuracy · Parkinson's Disease (PD) · Time-Based Measurements · Web-Based Cognitive Testing · IRT · TTS

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
A. Antonacopoulos et al. (Eds.): ICPR 2024, LNCS 15302, pp. 193–207, 2025.
https://doi.org/10.1007/978-3-031-78166-7_13

1 Introduction

Parkinson's Disease (PD) is a neurodegenerative condition that affects millions globally and it poses challenges in early detection and severity assessment. Traditional methods for detection and monitoring PD, despite their strengths, hold many limitations, such as invasiveness, high cost, or lack of accessibility [1, 2]. Recent advancements in digital biomarkers and machine learning present promising potential in bridging these gaps [3, 4].

However, research on online cognitive tests for Parkinson's disease (PD) has shown that current tools are underdeveloped and there is still an underexplored gap in using online cognitive tests and analyzing response dynamics for PD.

For example, Sousa (2022) emphasizes the lack of a common cognitive assessment battery for late-stage PD, and the need for tests that are quick, easy to use, and cover all relevant cognitive domains [5]. Here, authors recommend a cognitive assessment toolkit that considers the complex characteristics of PD, including being quick and easy to use, with minimized motor demands, and covering all relevant cognitive domains. Among the recommended instruments, the Trail Making Test corresponds well with these requirements [5].

The Trail Making Test Part A (TMT A) is a neuropsychological assessment tool primarily sensitive to processing speed and visual attention. It measures an individual's ability to rapidly connect a sequence of numbered circles in ascending order. TMT A focuses on basic scanning, attention, and motor speed.

The Trail Making Test Part B (TMT B) is a cognitive test known for its sensitivity to executive functions (involving attention, memory, visual search, motor function, and sequencing abilities). Trail B is generally more sensitive to executive functioning than TMT A since it requires multiple abilities to complete it.

But as Park et al. (2022) noted, the equivalency of paper-based and computerized tests cannot be assumed [6]. However, their findings support the comparability of TMT in computerized assessments and suggest this tool as a starting point for an early diagnostic tool. This conclusion is further supported by Mishra et al. (2022), who demonstrated that the digitized version of the TMT can determine cognitive-motor abilities and distinguish individuals with mild cognitive impairment and PD from healthy controls [7]. In their study, they observed correlations between TMT completion time and gait speed measured by a wearable accelerometer ($r = -0.4$, $p = 0.011$) and the Montreal Cognitive Assessment (MoCA) score ($r = -0.56$, $p < 0.01$).

Moreover, a study by Templeton et al. (2022) used fourteen tablet-based neurocognitive functional tests and machine learning model (decision trees) [7]. It allowed for the discrimination of PD from healthy controls (92.6% accuracy), and early and advanced stages of PD (73.7% accuracy). These results compare with current gold standard tools, such as standardized health questionnaires like the Unified Parkinson's Disease Rating Scale (UPDRS) with 78.3% accuracy and functional movement assessments with 70% accuracy.

Collectively, these findings underscore that a digital approach is feasible and allows for a comprehensive view of these conditions and their progression.

Therefore, this study explores the potential of the online version of Trail Making and machine learning models in identifying patterns indicative of Parkinson's Disease severity. By doing so, this research provides insights into quick, non-invasive, cost-effective, and accessible means of PD assessment and monitoring, potentially facilitating adapted interventions and earlier diagnosis.

2 Methods

We created an online version of Trail Making Test A and B [Fig. 1] [Fig. 2]. Then, we invited people with Parkinson's disease to solve this test. Our goal was to score their disease severity level (MILD or ADVANCED). The test took participants around two minutes to solve.

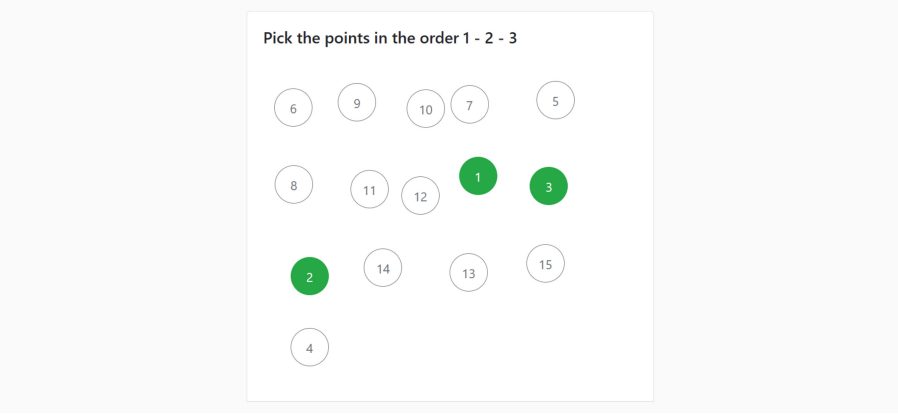


Fig. 1. The image presents an interactive cognitive task screen where the user is asked to pick points in a specific sequence. In TMT A, there are circles labeled with numbers.

In addition to the (a) number of mistakes, and (b) total time taken [TTS- time to submit, a standard measure in this test], we also recorded (c) instrumental reaction time [IRT- time to first selection] [8, 9]. That allowed us to assess the psychomotor speed and executive functions in two approaches.

In both approaches, reaction time was subtracted from the response time, to extract a pure psychometric component. In the first approach, we subtracted the time to complete TMT A from TMT B (TMT B − TMT A), that is a standard method in the clinical setup [10]. However, in the clinical context, the distinction between reaction and response times is not traditionally emphasized, and it's introduced as a novel method for this test.

Furthermore, in the second approach, we added the pure response times together (TMT A + TMT B) in order to evaluate machine learning models' performance using this variable and compare the results. In both approaches, longer times represented worse cognitive function.

All participants had confirmed Parkinson's disease diagnosis, and they were receiving treatment and advice from neurologists at UMass Chan Medical School. Eight participants had UPDRS III scores between 10 and 29 (indicative of mild symptoms of PD), and

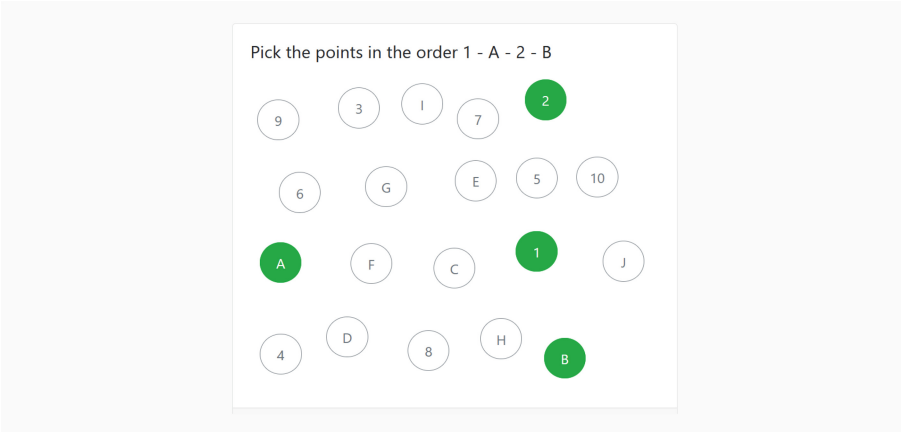


Fig. 2. The image presents an interactive cognitive task screen where the user is asked to pick points in a specific sequence. In TMT B, there are circles labeled with numbers together with letters.

seven participants with UPDRS III scores above 30 (indicative of advanced symptoms of PD). The sample size (n = 15) size makes available only preliminary comparisons and insights. However, we plan a study with a larger group to confirm these findings and explore other variables that may influence the results.

We analyzed aggregated data using IBM SPSS 29 software. We compared variables between patients with mild and advanced Parkinson's disease, with p-value significant below 0.05.

Finally, we evaluated multiple ML models in a task of disease severity prediction. We implemented four machine learning models in Python using the library scikit-learn for modeling and metrics, together with pandas for data processing, and seaborn with matplotlib for visualizations [11–14].

The size of a dataset (n = 15) produced a risk of overfitting. To mitigate this risk, we considered simpler models rather than deep neural networks. Specifically, we selected Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), and Random Forest. Moreover, before training, we extracted features importance to apply only the most relevant ones using Random Forest classifier.

Furthermore, we note that models' performance can vary based on the specific characteristics of the dataset. Therefore, for a better overview of the performance, we analyzed and compared them alongside.

To address the consideration of hyperparameter tuning, we used GridSearchCV with cross-validation, and we limited the range and number of hyperparameters. We used Stratified K-Fold cross-validation method that ensured the same proportion of classes in each fold. The dataset was balanced. Models were trained on random 10 samples from the dataset and validated on 5 independent samples (train_test_split, allocating 1/3 of the data for testing before modeling). The test data was separated before applying k-fold CV.

The flowchart of applied methodology is presented in Fig. 3. This report presents outcomes, statistics, and ML models of this preliminary research.

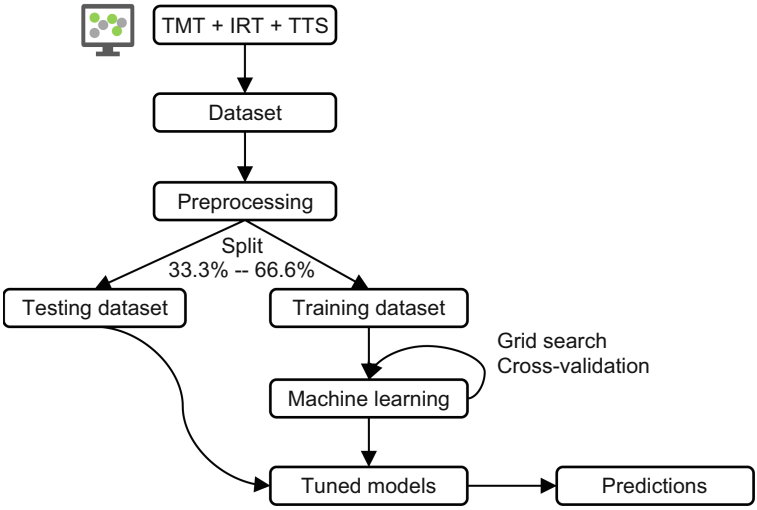


Fig. 3. Flowchart of applied methodology.

2.1 Features selection

In this study, the goal was to predict whether a participant belongs to MILD or ADVANCED group. However, with a small dataset, having too many features can lead to overfitting, where the model learns the noise in the training data instead of the actual signal.

Here, the introduction of a new composite variable can provide additional insights, especially when dealing with medical data sets and when optimizing for machine learning models. Therefore we experimented with two new composite variables (tmt_ba_medical_response_ms and tmt_ba_artificial_response_ms), that are designed to capture different aspects of TMT performance and narrow down the number of parameters.

Variable tmt_ba_medical_response_ms is calculated as the difference between the TMT B response time and reaction time, minus the difference between the TMT A response time and reaction time (Eq. 1).

$$\begin{aligned}\text{TMT_BA_MEDICAL_RESPONSE_MS} = {} & (\text{TMT_B_RESPONSE_MS} - \text{TMT_B_REACTION_MS}) \\ & - (\text{TMT_A_RESPONSE_MS} - \text{TMT_A_REACTION_MS})\end{aligned} \tag{1}$$

Conceptually, it isolates the pure response component of the TMT B task from the TMT A task, attempting to adjust for basic reaction time to highlight more specific

cognitive processing or motor execution times involved in the more complex TMT B task compared to TMT A.

On the other hand, variable tmt_ba_artificial_response_ms sums the differences between the response and reaction times for both TMT B and TMT A (Eq. 2).

$$\begin{aligned} &\text{TMT_BA_ARTIFICIAL_RESPONSE_MS} = \\ &(\text{TMT_B_RESPONSE_MS} - \text{TMT_B_REACTION_MS}) \\ &+ (\text{TMT_A_RESPONSE_MS} - \text{TMT_A_REACTION_MS}) \end{aligned} \tag{2}$$

This approach combines the total time that is required to complete both tasks, potentially serving as a single measure of the cognitive and motor demands placed on the individual by both tests.

Both approaches allowed us to conduct modeling using clinically relevant parameters, such as error count for both tests, and singular time measurement in each experiment (either tmt_ba_medical_response_ms or tmt_ba_artificial_response_ms, accordingly).

It's important to note that in clinical settings, practitioners typically measure only response time in seconds using a pen-and-paper method for the TMT, without registering reaction time with high precision. This traditional approach does not capture the nuanced differences between reaction and response times that our composite variables do. Hence, our methodology offers a more accurate and insightful analysis of TMT performance, surpassing the conventional clinical setup's capabilities.

3 Results

The study involved fifteen participants. All of them had a confirmed Parkinson's disease diagnosis, and they were receiving treatment and advice from neurologists at UMass Chan Medical School. Eight participants had UPDRS III scores between 10 and 29 (indicative of mild symptoms of PD), and seven participants with UPDRS III scores above 30 (indicative of advanced symptoms of PD). Accordingly, patients were divided into two categories: MILD and ADVANCED.

3.1 Statistical Analysis

MILD had an average age of 70.75 years (Std. Error Mean [SE] = 1.306), with four females and four males. ADVANCED had an average age of 70.86 (SE = 3.074), with four females and three males [Table 1, Fig. 4].

Both groups have a similar age profile, with MILD at an average of 70.75 years and ADVANCED at 70.86 years. The p-value of 0.974 suggests there is no significant difference in age between the two groups. The gender distribution (represented as a proportion, with 0 for males and 1 for females) is slightly higher for females in the ADVANCED group (0.57) compared to the MILD group (0.50). However, the p-value of 0.800 shows this difference is not statistically significant.

On average, MILD patients made more errors (1.50) on the TMT A than ADVANCED patients (0.86). Despite this, the p-value of 0.599 indicates that the difference is not statistically significant.

Table 1. Comparison of the characteristics of patients in two groups – MILD and ADVANCED (ADV).

Variable	MILD (n = 8)	Std. Err.	ADV (n = 7)	Std. Err.	p-value
Age(y)	70.75	1.306	70.86	3.074	0.974
Gender(0 = M, 1 = F)	0.50	0.189	0.57	0.202	0.800
TMT A Errors	1.50	0.926	0.86	0.705	0.599
TMT A IRT(s)	5.408	1.956	8.199	4.951	0.591
TMT A TTS(s)	30.637	4.953	34.152	6.690	0.675
TMT B Errors	3.50	2.478	5.71	2.523	0.544
TMT B IRT(s)	3.399	0.377	3.282	0.452	0.844
TMT B TTS(s)	60.818	11.42	81.856	11.50	0.219

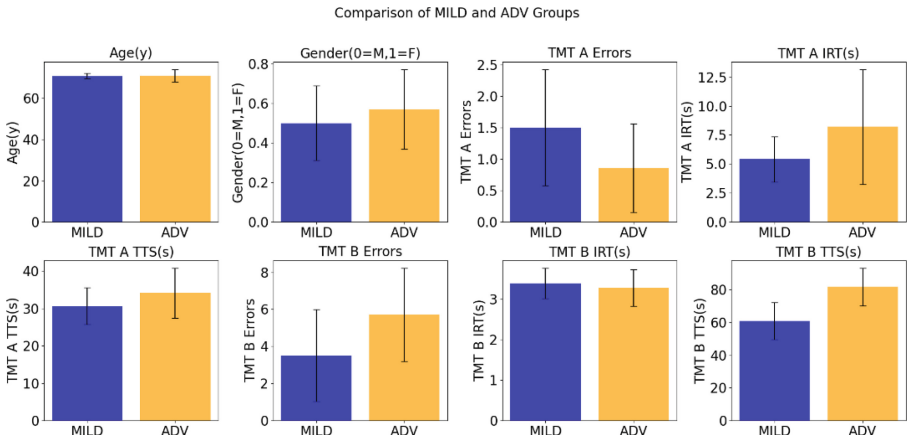


Fig. 4. Comparison of the characteristics of patients in two groups – MILD and ADVANCED (ADV).

The reaction and response times for TMT A are higher for the ADVANCED group compared to the MILD group, indicating slower performances. Specifically, reaction times average 5.408 s for MILD and 8.199 s for ADVANCED, while response times are 30.637 s for MILD and 34.152 s for ADVANCED. Neither difference is statistically significant, with p-values of 0.591 and 0.675, respectively. Interestingly, the reaction times in both groups were higher for TMT A than for TMT B, potentially due to the novelty aspect. This is because TMT A was presented as the first test, making users less familiar with it. As a result, they became more adjusted to TMT B, which may have shortened their reaction times.

Furthermore, ADVANCED patients tend to make more errors (5.71) on the TMT B than MILD patients (3.50), though this difference is not statistically significant (p-value = 0.544). Finally, for TMT B, the reaction and response times do not significantly

differ between groups, with MILD patients slightly faster in reaction times and slower in response times compared to ADVANCED. The p-values of 0.844 for reaction times and 0.219 for response times suggest these differences are not statistically significant.

3.2 Patterns in Variables

The connections between variables were explored using a heatmap form of Pearson correlation coefficients [Fig. 5]. This matrix presented a strong positive correlation of 0.69 between the age during the test and the reaction time in TMT A, suggesting that as age increases, the reaction time tends to increase as well.

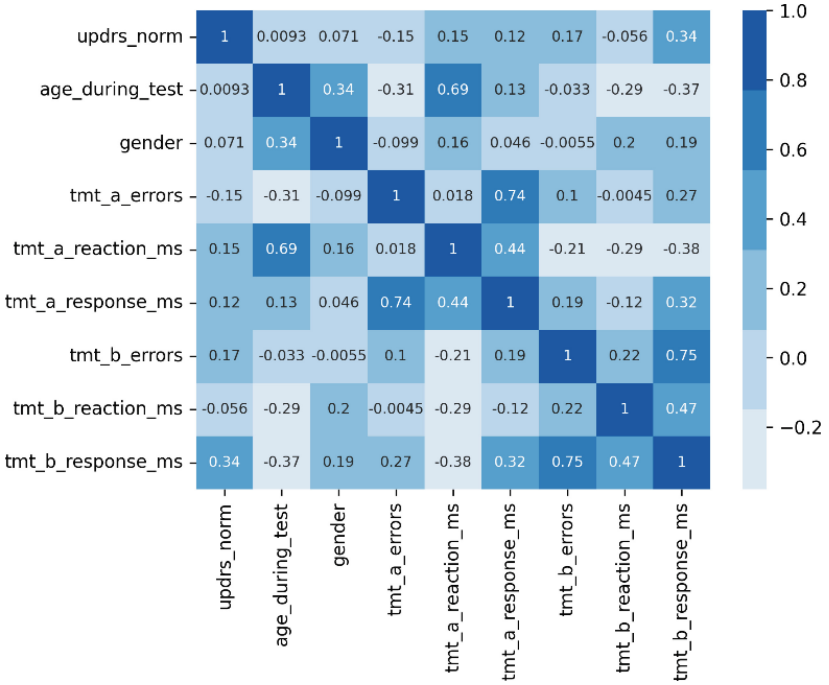


Fig. 5. A heatmap representing Pearson correlation coefficients. The darker the blue, the stronger the positive correlation; the closer to white or the presence of lighter blue indicates a weaker correlation. It's important to note that a correlation does not imply causation. These values simply indicate the strength and direction of the linear relationship between the pairs of variables.

Moreover, there is a very strong positive correlation of 0.74 between the number of errors made in TMT A and the TMT A response time, indicating that a higher number of errors is associated with a longer response time to complete TMT A.

Likewise, there is a very strong positive correlation of 0.75 between TMT B errors and TMT B response time, implying that as the number of errors increases, the total time to respond in TMT B also increases.

Additionally, there is a significant positive correlation of 0.47 between TMT B reaction time and TMT B response time, suggesting that longer reaction times are somewhat associated with longer total times to respond in TMT B.

Finally, a strong correlation exists between age during the test and TMT A reaction time, with a coefficient of 0.69, indicating that older participants may have slower reaction times in TMT A. The other correlations presented in the heatmap are weaker, meaning they show less of a linear relationship between variables. For instance, there's a weak negative correlation between TMT A Errors and Age during Test (−0.31), suggesting a slight tendency for older participants to make fewer errors on TMT A, although this relationship is not strong.

3.3 Machine Learning Models

To analyze how models learn, we used the Area Under the Curve (AUC). This is a measure of the overall performance of a classification model.

TMT Errors and Medical Response Dynamics

We incorporated the number of errors from TMT A and TMT B (tmt_a_errors, tmt_b_errors), together with medical response time (tmt_ba_medical_response_time) in order to predict UPDRS group (MILD or ADVANCED).

The Random Forest has the highest AUC of 0.92, indicating it has the best performance among the four classifiers in terms of ROC-AUC [Fig. 6]. This presents that it maximizes the true positive rate while minimizing the false positive rate better than the other classifiers in this set of ML algorithms.

Random Forest model performed noticeably better than the Logistic Regression, Support Vector Machine and Naïve Bayes. Random Forest correctly predicted 80% of the outcomes (accuracy: 0.8, precision: 0.75, sensitivity: 1.0, specificity: 0.5), which is significant when compared to other models [Table 2]. Random Forest predicted all positive cases correctly (1 true negatives, 3 true positives, and 1 false positive, with no false negatives).

The Logistic Regression and Support Vector Machine models both show a low accuracy of 0.4 (precision of 1.0 and sensitivity of 0.0), which means they were unable to correctly identify positive cases in this scenario. However, their specificity is at 1.0, indicating they could correctly identify all negative cases.

Furthermore, Naïve Bayes demonstrates the lowest accuracy among the compared models (accuracy: 0.2, precision: 0.33, sensitivity: 0.33, specificity: 0.0).

TMT Errors and Artificial Response Dynamics

In the second experiment, we incorporated the number of errors from TMT A and TMT B (tmt_a_errors, tmt_b_errors), together with artificial response time (tmt_ba_artificial_response_time) in order to predict UPDRS group (MILD or ADVANCED).

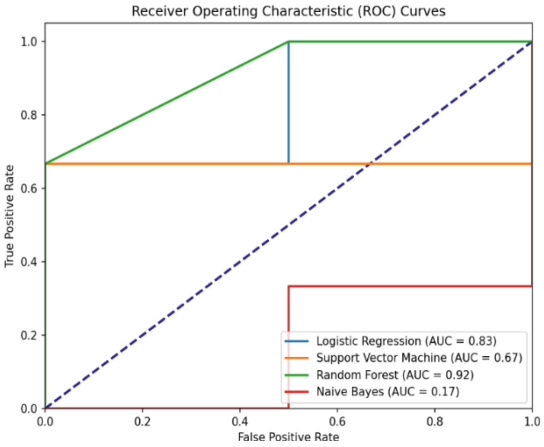


Fig. 6. The Area Under the Curve (AUC) is a measure of the overall performance of a classification model. Using Medical Response Time, Random Forest has the highest AUC of 0.92, indicating it has the best performance among the four classifiers. The Logistic Regression curve (blue line) is covered by the Support Version Machine (orange line) and Random Forest (green line) curve. The random change line represents the baseline performance of a random classifier (purple dashed line). (Color figure online)

Table 2. Comparison of models' performance using tmt_a_errors, tmt_b_errors and tmt_ba_medical_response_ms. The table presents Accuracy (Acc.), Precision (Prec.), Sensitivity (Sens.), and Specificity (Spec.).

Model Name	Hyper-tuned Params	Acc.	Prec.	Sens.	Spec.
Logistic Regression	{'C': 1.0, 'max_iter': 100000, 'penalty': 'l2', 'solver': 'lbfgs'}	0.40	1.00	0.00	1.00
Support Vector Machine	{'C': 1.0, 'kernel': 'linear', 'max_iter': 100000}	0.40	1.00	0.00	1.00
Random Forest	{ bootstrap': True, 'criterion': 'gini', 'max_features': 'sqrt', 'n_estimators': 100}	0.80	0.75	1.00	0.50
Naive Bayes	{'priors': None, 'var_smoothing': 1e-09}	0.20	0.33	0.33	0.00

The Random Forest has the highest AUC of 0.92, achieving results comparable to the first experiment Fig. 7. After the analysis of feature importance in both models it seems that Random Forest scored tmt_ba_artificial_response_ms and tmt_ba_medical_response_ms similarly (0.65), placing tmt_b_errors on second (0.23), and tmt_a_errors on third place (0.11).

Repeatedly, Random Forest model performed better than the Logistic Regression, Support Vector Machine and Naïve Bayes. Random Forest correctly predicted 80% of the outcomes (accuracy: 0.8, precision: 0.75, sensitivity: 1.0, specificity: 0.5), which is significant when compared to other models Table 3. Random Forest predicted all positive cases correctly (1 true negatives, 3 true positives, and 1 false positive, with no false negatives).

The Logistic Regression and Support Vector Machine models both show a low accuracy of 0.4 (precision of 1.0 and sensitivity of 0.0), which means they were unable to correctly identify positive cases in this scenario. However, their specificity is at 1.0, indicating they could correctly identify all negative cases.

Furthermore, Naïve Bayes demonstrates the lowest accuracy among the compared models (accuracy: 0.2, precision: 0.33, sensitivity: 0.33, specificity: 0.0).

In conclusion, Random Forest stands out with the highest accuracy (0.8), good precision (0.75), perfect sensitivity (1.0), and a specificity of 0.5. Despite the lower specificity compared to Logistic Regression and Support Vector Machine, the high sensitivity and accuracy rates highlight its overall superior performance in predicting outcomes correctly.

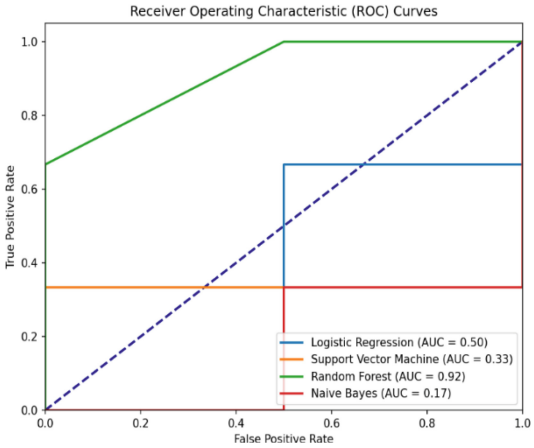


Fig. 7. The Area Under the Curve (AUC) is a measure of the overall performance of a classification model. Using Artificial Response Time, Random Forest has the highest AUC of 0.92, indicating it has the best performance among the four classifiers. The random change line represents the baseline performance of a random classifier (purple dashed line). (Color figure online)

Table 3. Comparison of models' performance using tmt_a_errors, tmt_b_errors and tmt_ba_artificial_response_ms. The table presents Accuracy (Acc.), Precision (Prec.), Sensitivity (Sens.), and Specificity (Spec.).

Model Name	Hyper-tuned Params	Acc.	Prec.	Sens.	Spec.
Logistic Regression	{'C': 1.0, 'max_iter': 100000, 'penalty': 'l2', 'solver': 'lbfgs'}	0.40	1.00	0.00	1.00
Support Vector Machine	{'C': 1.0, 'kernel': 'linear', 'max_iter': 100000}	0.40	1.00	0.00	1.00
Random Forest	{ bootstrap': True, 'criterion': 'gini', 'max_features': 'sqrt', 'n_estimators': 100}	0.80	0.75	1.00	0.50
Naive Bayes	{'priors': None, 'var_smoothing': 1e-09}	0.20	0.33	0.33	0.00

4 Discussion

This study preliminarily demonstrates the utility of the Trail Making Test with temporal measures in capturing the cognitive and motor impacts of bradykinesia in Parkinson's Disease.

Importantly, this study evaluates both TMT A (processing speed and visual attention), and TMT B (cognitive functions) together with response dynamics (initial cognitive processing and decision-making speed). This is because in Parkinson's disease, cognitive changes are independent from the motor symptoms development and thus have to be assessed independently [15]. This separation allows for the nuanced detection of PD's impact. While TMT A focuses on motor speed and visual search abilities, requiring participants to connect numbered dots in sequence, TMT B adds a cognitive layer by alternating between numbers and letters. Incorporation of the reaction and response time measurements refines this approach, making this test more sensitive to slowed voluntary movement detection.

Therefore, there is the potential to approximate the real-time effects of dopamine through TMT performance. Impaired patterns observed in the test could be indicative of underlying disruptions in dopaminergic pathways, which are central to PD's pathophysiology. This is particularly valuable given the challenge of directly assessing neurochemical changes in a clinical setting. By correlating TMT performance with known dopaminergic deficits, clinicians gain insights into the disease's neurobiological patterns.

Interestingly, integration of insights from computational models of the brain and detailed studies on neuronal oscillations in PD patients provides a deeper understanding of the disease's neural basis [16]. Mathiopoulou et al. presented that subthalamic beta oscillations are directly affected by both motor activity and therapeutic interventions such as dopamine replacement and deep brain stimulation (DBS) [17]. These beta oscillations, which are known to correlate with motor symptom severity, suggest a mechanism similar to the asynchronous process integration in the retina [16].

Just as the retina synchronizes processes to produce a coherent output from disparate sensory inputs, the subthalamic nucleus is crucial in coordinating motor commands disrupted by dopaminergic degeneration in PD. The TMT, by measuring reaction and response times, essentially assesses the efficiency of these neural synchronization patterns. Prolonged times might reflect the brain's struggle to integrate and synchronize neural processes efficiently, alike to the difficulties in processing and output synchronization.

To recognize these disrupted patterns, this study presents that Random Forest can be a helpful and accurate tool for this task. Random Forest are used because of their simplicity, ease of implementation, and their ability to perform well on a wide range of tasks with minimal hyperparameter tuning. Random Forest can capture interactions and nonlinear relationships between features, giving a possibility to model dopaminergic pathways disruptions through motor and cognitive data. It is worth noting that findings of this study align with other research that presents good performance of Random Forest in the disease severity classification task [9, 18, 19].

Here, it is important to note that this study has several limitations. In such a small group, even individual variability could account for these findings. These include the small size of the research groups, potentially limiting the generalizability. Therefore further, more detailed research (larger sample size, longitudinal study) is crucial for more detailed insights that help with the prevention of neurodegenerative diseases.

In light of this, the low performance of the Naïve Bayes model can be attributed to the small sample size of our dataset. Naïve Bayes relies on the assumption of a normal distribution of data and requires a sufficiently large dataset to accurately estimate the priors and likelihoods. Given the limited number of participants ($n = 15$), the data may not adequately capture the underlying distributions, leading to lower performance of the Naïve Bayes model. This issue might also relate to the fact that this dataset may not be representative of the broader population, further impacting the model's ability to generalize well.

Despite its limitations, this research shows the potential of a digital approach and contributes to the understanding and management of PD, particularly in customized interventions and early detection. TMT with temporal measures could be integrated into clinical practice or remote monitoring systems to better navigate and mitigate the impacts of PD.

Implementation of web version of TMT tests in clinical practice could offer a quick, non-invasive, and accessible method of assessing disease impact on cognitive and motor functions. Clinicians could use these tests for regular monitoring, enabling well-timed adjustments to treatment plans based on subtle changes in cognitive or motor performance.

In practice, implementing this approach requires creating a website that presents a TMT test with points labeled with letters and numbers, which participants must click in the correct order. This application needs to count the number of mistakes and record the start time of the first selection (IRT, e.g., calling performance.now()) and the time of the last selection (TTS, with the same method call). These metrics, along with TMT results, can provide valuable insights into delayed patterns in motor and cognitive responses.

Importantly, this technique can be enhanced by incorporating trajectory analysis of mouse movement. Moreover, there is an opportunity for further integration data going from wearable devices using sensors such as accelerometers and gyroscopes [7].

Integrating these assessments into telehealth platforms could facilitate remote monitoring, making it easier to track patient progress and intervene promptly. Such advancements could significantly enhance personalized care strategies, improving outcomes for PD patients.

To conclude, the results of this paper call upon the research community to explore these tools further and clinicians to consider their practical applications, given the significant diagnostic benefits that both they and their patients can gain.

Acknowledgments. The authors would like to thank all the participants who were involved in this study, including patients and doctors from UMass Chan Medical School.

Funding. This research received no external funding.

Institutional Review Board Statement
The study was reviewed and approved by the Institutional Review Board at UMass Chan Medical School (protocol code: IRB H0008962) to ensure compliance with the Declaration of Helsinki.

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

Data Availability Statement. The data presented in this study are available on request from the corresponding author due to privacy and legal reasons.

Conflicts of Interest. The authors declare no conflicts of interest.

References

1. Talitckii, A., et al.: Comparative study of wearable sensors, video, and handwriting to detect Parkinson's disease. IEEE Trans Instrum Meas. **71**, 1–10 (2022). https://doi.org/10.1109/TIM.2022.3176898
2. Khare, S.K., Bajaj, V., Acharya, U.R.: PDCNNet: an automatic framework for the detection of Parkinson's disease using EEG signals. IEEE Sens J. **21**, 17017–17024 (2021). https://doi.org/10.1109/JSEN.2021.3080135
3. Chudzik, A., Śledzianowski, A., Przybyszewski, A.W.: Machine learning and digital biomarkers can detect early stages of neurodegenerative diseases. Sensors. **24**, 1572 (2024). https://doi.org/10.3390/s24051572
4. Przybyszewski, A.W., Śledzianowski, A., Chudzik, A., Szlufik, S., Koziorowski, D.: Machine learning and eye movements give insights into neurodegenerative disease mechanisms. Sensors. **23**, 2145 (2023). https://doi.org/10.3390/s23042145
5. Severiano e Sousa, C., Alarcão, J., Pavão Martins, I., Ferreira, J.J.: Cognitive testing in late-stage Parkinson's disease: A critical appraisal of available instruments. Appl. Neuropsychol. Adult. **31**, 191–202 (2024). https://doi.org/10.1080/23279095.2022.2114355.
6. Park, S.-Y., Schott, N.: The trail-making-test: Comparison between paper-and-pencil and computerized versions in young and healthy older adults. Appl. Neuropsychol. Adult. **29**, 1208–1220 (2022). https://doi.org/10.1080/23279095.2020.1864374

7. Mishra, R.K., Park, C., Zhou, H., Najafi, B., Thrasher, T.A.: Evaluation of motor and cognitive performance in people with parkinson's disease using instrumented trail-making test. Gerontology. **68**, 234–240 (2022). https://doi.org/10.1159/000515940
8. Chudzik, A., Drabik, A., Przybyszewski, A.W.: Investigating the Impact of Parkinson's Disease on Brain Computations: An Online Study of Healthy Controls and PD Patients. Intelligent Information and Database Systems: 15th Asian Conference, ACIIDS 2023, Phuket, Thailand, July 24–26, 2023, Proceedings, Part II. 235–246 (2023). https://doi.org/10.1007/978-981-99-5837-5_20.
9. Chudzik, A., Przybyszewski, A.W.: Classification of Parkinson's disease using machine learning with MoCA response dynamics. Appl. Sci. **14**, 2979 (2024). https://doi.org/10.3390/app14072979
10. Foki, T., Hitzl, D., Pirker, W., Novak, K., Pusswald, G., Lehrner, J.: Individual cognitive change after DBS-surgery in Parkinson's disease patients using Reliable Change Index Methodology. Neuropsychiatrie. 32 (2018). https://doi.org/10.1007/s40211-018-0271-4.
11. Pedregosa, F., et al.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 12 (2011)
12. Waskom, M.: Seaborn: statistical data visualization. J Open Source Softw. 6, 3021 (2021). https://doi.org/10.21105/joss.03021.
13. Hunter, J.D.: Matplotlib: A 2D graphics environment. Comput. Sci. Eng. **9**, 90–95 (2007). https://doi.org/10.1109/MCSE.2007.55
14. McKinney, W.: Data Structures for Statistical Computing in Python. 56–61 (2010). https://doi.org/10.25080/Majora-92bf1922-00a.
15. Przybyszewski, A.W., Nowacki, J.P., Drabik, A., Szlufik, S., Koziorowski, D.M.: IGrC: cognitive and motor changes during symptoms development in Parkinson's Disease Patients. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (2020). https://doi.org/10.1007/978-3-030-42058-1_46.
16. Przybyszewski, A.W., Linsay, P.S., Gaudiano, P., Wilson, C.M.: Basic difference between brain and computer: Integration of asynchronous processes implemented as hardware model of the retina. IEEE Trans Neural Netw. **18**, (2007). https://doi.org/10.1109/TNN.2006.882814.
17. Mathiopoulou, V., et al.: Modulation of subthalamic beta oscillations by movement, dopamine, and deep brain stimulation in Parkinson's disease. NPJ. Parkinsons Dis. **10**, 77 (2024). https://doi.org/10.1038/s41531-024-00693-3
18. Przybyszewski, A.W., Chudzik, A., Szlufik, S., Habela, P., Koziorowski, D.M.: Comparison of different data mining methods to determine disease progression in dissimilar groups of Parkinson's patients. Fundam Inform. **176**, 167–181 (2020). https://doi.org/10.3233/FI-2020-1969
19. Chudzik, A., Szymański, A., Nowacki, J.P., Przybyszewski, A.W.: Eye-Tracking and Machine Learning Significance in Parkinson's Disease Symptoms Prediction. In: Nguyen Ngoc Thanh and Jearanaitanakij, K. and S.A. and T.B. C.S. (ed.) Intelligent Information and Database Systems, pp. 537–547. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-42058-1_45.