POLISH-JAPANESE ACADEMY OF INFORMATION TECHNOLOGY

Machine Learning Methods for Parkinson's Disease Datasets

PhD thesis in Computer Science

by Artur Chudzik, MSc, Eng

Polish-Japanese Academy of Information Technology

Supervised by prof. Andrzej Przybyszewski, PhD, DSc,
Professor of Informatics, Polish-Japanese Academy of Information Technology,
and Department of Neurology, UMass Chan Medical School

Warsaw, 2024

Title: **Machine Learning Methods for Parkinson's Disease Datasets**

Author: Artur Chudzik

Abstract: Intelligent, dependable, and explainable computational methods for brain disease datasets remain underdeveloped and insufficiently validated, despite significant progress in machine learning (ML). Addressing this gap is crucial, as neurodegenerative diseases (NDs), such as Parkinson's disease (PD), continue to affect an increasing number of individuals. Given the challenges in managing and accurately detecting NDs, this thesis explores three approaches for data collection and analysis. Specifically, it examines diffusion tensor imaging (DTI), eye tracking, and online cognitive testing, along with selected ML algorithms designed to model PD patterns within these datasets. In the first approach, ML identified brain abnormalities through DTI analysis, detecting issues such as handwriting distortion after deep brain stimulation treatment with 82% accuracy. In the second approach, ML analyzed rapid eye movements (saccades) to identify patterns that differentiate PD patients. By combining eye-tracking data with ML algorithms, models predicted symptom development using Unified Parkinson's Disease Rating Scale (UPDRS) scores with 57-79% accuracy. In the third approach, data from web-based cognitive and behavioral tests was used to identify motor and cognitive signs of PD. Using this dataset, ML distinguished healthy controls (HC) from PD patients with 93% accuracy and mild PD from advanced PD with 80% accuracy (based on their UPDRS score). Although limited by sample size, each method provided computational insights into disease progression and helped determine PD severity based on digital data and machine learning models. Thus, the findings of this thesis contribute to the field of computer science by designing, developing, and describing digital platforms and explainable ML workflows, which are applicable to complex and limited datasets.

Tytuł: **Metody uczenia maszynowego dla zbiorów danych o chorobie Parkinsona**

Autor: Artur Chudzik

Streszczenie: Inteligentne, niezawodne i wyjaśnialne metody obliczeniowe dla zbiorów danych dotyczących chorób mózgu pozostają niedostatecznie rozwinięte i niewystarczająco zweryfikowane, pomimo znaczącego postępu w dziedzinie uczenia maszynowego (ML). Zaadresowanie tego problemu jest kluczowe, ponieważ choroby neurodegeneracyjne (NDs), takie jak choroba Parkinsona (PD), dotykają systematycznie rosnącą liczbę osób. Biorąc pod uwagę wyzwania związane z zarządzaniem i precyzyjnym wykrywaniem NDs, niniejsza praca bada trzy podejścia do zbierania i analizy danych. W szczególności, praca opisuje obrazowanie tensora dyfuzji (DTI), śledzenie ruchu gałek ocznych oraz testy kognitywne online, wraz z wybranymi algorytmami ML zaprojektowanymi do modelowania wzorców PD w tych danych. W pierwszym podejściu, model ML zidentyfikował nieprawidłowości w mózgu poprzez analizę DTI, wykrywając problemy takie jak zniekształcenie pisma po leczeniu za pomocą głębokiej stymulacji mózgu z dokładnością 82%. W drugim podejściu, ML analizował szybkie ruchy gałek ocznych (sakady), aby zidentyfikować wzorce różnicujące pacjentów z PD. W tym podejściu, dzięki połączeniu danych ze śledzenia ruchu gałek ocznych z algorytmami ML, modele przewidywały rozwój objawów za pomocą wyników skali Unified Parkinson's Disease Rating Scale (UPDRS) z dokładnością 57-79%. W trzecim podejściu wykorzystano dane z internetowych testów kognitywno-behawioralnych do identyfikacji objawów motorycznych i kognitywnych PD. Wykorzystując ten zbiór danych, ML rozróżnił zdrowe osoby (HC) od pacjentów z PD z dokładnością 93% oraz łagodną postać PD od zaawansowanej PD z dokładnością 80% (na podstawie wyników UPDRS). Pomimo ograniczeń wynikających z wielkości próby, każda z metod dostarczyła wglądów w progresję choroby oraz pomogła określić stopień zaawansowania PD na podstawie danych cyfrowych i modeli uczenia maszynowego. Dzięki temu, wyniki tej pracy wnoszą wkład w dziedzinę informatyki poprzez zaprojektowanie, stworzenie i opisanie platform cyfrowych wraz z wyjaśnialnymi przepływami pracy ML, mających zastosowanie do złożonych i ograniczonych zbiorów danych.

Table of Contents

Articles Constituting the Thesis

- **[A1]** - A. Chudzik, A. Szymański, J. P. Nowacki, and A. W. Przybyszewski, "DTI Helps to Predict Parkinson's Patient's Symptoms Using Data Mining Techniques," in *Intelligent Information and Database Systems*, F. L. and H. T.-P. and T. B. Nguyen Ngoc Thanh and Gaol, Ed., Cham: Springer International Publishing, pp. 615–623, 2019. doi: 10.1007/978-3-030-14802-7_53.
- **[A2]** - A. W. Przybyszewski, A. Śledzianowski, A. Chudzik, S. Szlufik, and D. Koziorowski, "Machine Learning and Eye Movements Give Insights into Neurodegenerative Disease Mechanisms," Sensors, vol. 23, no. 4, p. 2145, Feb. 2023, doi: 10.3390/s23042145.
- **[A3]** - A. W. Przybyszewski, A. Chudzik, S. Szlufik, P. Habela, and D. M. Koziorowski, "Comparison of Different Data Mining Methods to Determine Disease Progression in Dissimilar Groups of Parkinson's Patients," Fundamenta Informaticae, vol. 176, no. 2, pp. 167–181, Dec. 2020, doi: 10.3233/FI-2020-1969.
- **[A4]** - A. Chudzik, A. Szymański, J. P. Nowacki, and A. W. Przybyszewski, "Eye-Tracking and Machine Learning Significance in Parkinson's Disease Symptoms Prediction," in *Intelligent Information and Database Systems*, K. and S. A. and T. B. and C. S. Nguyen Ngoc Thanh and Jearanaitanakij, Ed., Cham: Springer International Publishing, pp. 537–547, 2020. doi: 10.1007/978-3-030-42058-1_45.
- **[A5]** - A. Chudzik, A. Śledzianowski, and A. W. Przybyszewski, "Machine Learning and Digital Biomarkers Can Detect Early Stages of Neurodegenerative Diseases," Sensors, vol. 24, no. 5, p. 1572, Feb. 2024, doi: 10.3390/s24051572.
- **[A6]** - A. Chudzik, A. Drabik, and A. W. Przybyszewski, "Investigating the Impact of Parkinson's Disease on Brain Computations: An Online Study of Healthy Controls and PD Patients," in *Intelligent Information and Database Systems: 15th Asian Conference, ACIIDS 2023, Phuket, Thailand, July 24–26, 2023, Proceedings, Part II*, Berlin, Heidelberg: Springer-Verlag, pp. 235–246, 2023. doi: 10.1007/978-981-99-5837-5_20.
- **[A7]** - A. Chudzik and A. W. Przybyszewski, "Classification of Parkinson's Disease Using Machine Learning with MoCA Response Dynamics," Applied Sciences, vol. 14, no. 7, p. 2979, Apr. 2024, doi: 10.3390/app14072979.
- **[A8]** - A. Chudzik, J. P. Nowacki, and A. W. Przybyszewski, "Recognizing Patterns of Parkinson's Disease using Online Trail Making Test and Response Dynamics – Preliminary Study," ICPR 2024. PREPRINT.

Additional Contributions

- **[C1]** - S. Szlufik, A. Chudzik, A. Przybyszewski, and D. Koziorowski, “Amantadine treatment in Parkinson’s disease patients as a modulatory factor of SARS-Cov-2 infection,” *Parkinsonism Relat Disord*, vol. 113, Aug. 2023, doi: 10.1016/j.parkreldis.2023.105606.
- **[C2]** - N. Kowalczyk, M. Myśliwiec, M. Skorko, P. Dobrowolski, A. Chudzik, and A. Brzezicka, “Structural neuroplasticity induced by training in the form of a first-person shooter video game,” in *9th Annual Conference Aspects of Neuroscience Abstract Book*, 1st ed., Paź Marta and Karpińska Magdalena, Eds., Warsaw: Warsaw University, pp. 149–149, 2019, ISBN: 978-83-954288-1-4.
- **[C3]** - A. W. Przybyszewski and A. Chudzik, “How to Cure Alzheimer’s Disease,” Journal of Alzheimer’s Disease, vol. 99, no. 4, pp. 1221–1223, May 2024, doi: 10.3233/JAD-240231. PMID: 38788078.

I. Introduction

Neurodegenerative diseases (NDs) are devastating conditions affecting millions of people around the world (Prince et al. 2015). This makes them the leading cause of cognitive and physical disability globally (Step et al. 2024). The prevalence of these diseases increases with age. Thus, with the global trend toward increased life expectancy, the burden of age-related NDs is expected to rise accordingly (Step et al. 2024).

The most common ND is Alzheimer's disease (AD), often described as a disease that impairs memory and cognitive functioning. The second most common ND is Parkinson's disease (PD), primarily associated with movement impairments.

Despite intensive efforts and considerable advances in recent decades, biomarkers research in Parkinson's disease has yielded only limited success (Yilmaz et al. 2019). As described in this thesis, the major reason for the difficulty in biomarker detection is the heterogeneity of PD, observable in every aspect of the disease, from pathology to clinical phenotype and disease progression, but also underdevelopment of appropriate digital screening tools (Yilmaz et al. 2019). This situation negatively affects the possibility of early detection of the disease.

Because of this complexity, machine learning (ML) is increasingly being applied to Parkinson's disease diagnosis and prognosis, showing potential improvements over traditional methods (Mei et al. 2021). ML techniques, including classical models and deep learning, can analyze various data types such as brain imaging, voice recordings, facial expressions, eye movements, and clinical assessments to detect and differentiate PD. These approaches show promise in enhancing diagnostic accuracy and enabling earlier detection of this disease.

Selected ML methods demonstrate significant potential for adaptation in clinical decision-making, offering a more systematic and informed approach to PD diagnosis. However, challenges remain in translating ML models into clinical practice, including disease-specific, task-specific, and technology-specific issues (Garcia Santa Cruz et al. 2023).

Thus, ongoing research focuses on refining ML techniques and addressing current limitations to improve their effectiveness in PD diagnosis and management. This increased effectiveness is necessary because neurodegenerative diseases are not fully understood and tools to address this problem are lacking.

NDs are complex, as they are characterized by the progressive loss of neurons in specific regions of the brain (Agnello and Ciaccio 2022). Neurodegeneration often starts in localized areas of the

brain and progressively spreads through interconnected regions, worsening as the disease advances (Iturria-Medina and Evans 2015). This interconnection means that dysfunction in one area can lead to disruptions in others. However, individual variations in compensatory mechanisms make early detection of neurodegenerative diseases challenging.

Disturbingly, neurodegeneration can start approximately 25-30 years before any evident symptoms appear (Liss et al. 2021). This is because compensatory processes, including changes in brain activation within function-specific networks, help maintain normal performance despite neuronal loss (Sheng et al. 2021). This compensation works until the damage exceeds a critical threshold, after which symptoms, such as tremor in PD, become apparent. Unfortunately, by this stage, many of the affected brain structures are already severely damaged.

No cure currently exists for both most common neurodegenerative diseases, namely AD and PD (Sequeira et al. 2024). Existing drugs can either slow the progression of the disease or artificially compensate for neuronal loss.

Research Gap

Because NDs are extremally challenging in management, there is a gap in validation of methods that will be sensitive enough to detect and define objectively different stages of the disease (Lima et al. 2022; Janssen Daalen et al. 2024). Intelligent, dependable, and explainable methods for disease management remain underdeveloped and insufficiently verified.

Addressing this gap is crucial, as the application of machine learning methods to Parkinson's disease datasets can enable the development of sensitive measures for screening individuals vulnerable to PD. Effectively, this screening can lead to the introduction of disease–modifying therapies, which can slow down the disease progression.

Hypothesis

One of possible directions for diagnostic methods is an approach called digital biomarkers (Janssen Daalen et al. 2024). This approach uses data collected from interactions with technology like web, mobile and wearable devices to understand human behavior or cognitive function. These data points can be aggregated and analyzed, especially using machine learning, to provide deeper insights into the cognitive state of a patient. This integration of data and ML is important because models can detect subtle patterns in variables that help to define the hidden rules of the disease.

Hypothetically, digital biomarkers with machine learning are tools that can be applied to detect PD symptoms. However, knowledge about feasibility and validity of individual digital biomarkers remains extremely limited (Fröhlich et al. 2022; Janssen Daalen et al. 2024). Therefore, this thesis evaluates selected machine learning methods applied to Parkinson's disease datasets and presents a summary of findings.

Structure of the Thesis

A digital diagnostic approach requires understanding data through both machine learning and neuroscience terminology, which is challenging due to the specific terms and methodologies from these rather diverse fields.

To address this, the structure of the thesis is organized as follows. The "I. Introduction" provides the rationale and motivation for the study. "II. Parkinson's Disease Datasets" offers an overview of the data used to build Parkinson's disease datasets. Specific methods for processing these datasets are described in "III. Models for Classification of Parkinson's Disease." The "IV. Contributions" section discusses digital biomarkers and machine learning techniques developed and used in the studies, along with the presentation of experimental findings and analyses. These are then discussed in the "V. Discussion" and concluded in the "VI. Conclusions" section. Because of the variety of terms, there is also a "Glossary" of terms available at the end of this thesis. This structure hopefully should provide a reader with an overview of classical diagnosis of neurodegenerative diseases and modern digital methods for their detection.

Given the variety of sources and the need to clearly distinguish between the articles constituting the thesis and other references, a specific citation format has been adopted. Thus, "Articles Constituting the Thesis" are cited using square brackets with labels **[A1-A8]**, and "Additional Contributions" are labeled **[C1-C3]**. Other references follow the basic parentheses *(author-year)* citation format and are listed in the "References" section.

II. Parkinson's Disease Datasets

A proper understanding of the dataset is crucial for training effective machine learning models. This section provides a brief technical background on the disease, highlighting its complexity due to overlapping symptoms with other conditions, and discusses current treatment approaches.

Parkinson's disease does not have a strict biological definition and diagnostic criteria (Farrow et al. 2022). This means that there is no single, specific biomarker (e.g., detectable in the blood or

cerebral-spinal fluid) which amount of can be connected to the precise and objective score of PD severity. Thus, it is important to notice that to tell if someone has the beginning of PD or not, can be based only on clinical, hence subjective, observation. As a result, clinical diagnostic accuracy remains suboptimal, even when the disease is clinically fully manifested (Tolosa et al. 2021).

Disrupted Signal in Neural Networks

Parkinson's disease is primarily characterized by the loss of dopamine-producing neurons in the brain, particularly in the region called substantia nigra (SN), a critical structure in the midbrain that is part of the basal ganglia, which plays a fundamental role in movement control. The substantia nigra produces dopamine (DA), a neurotransmitter essential for smooth, coordinated muscle activity. Dopamine deficiency disrupts the basal ganglia's ability to control and coordinate movement, which is a crucial issue in PD (Latif et al. 2021).

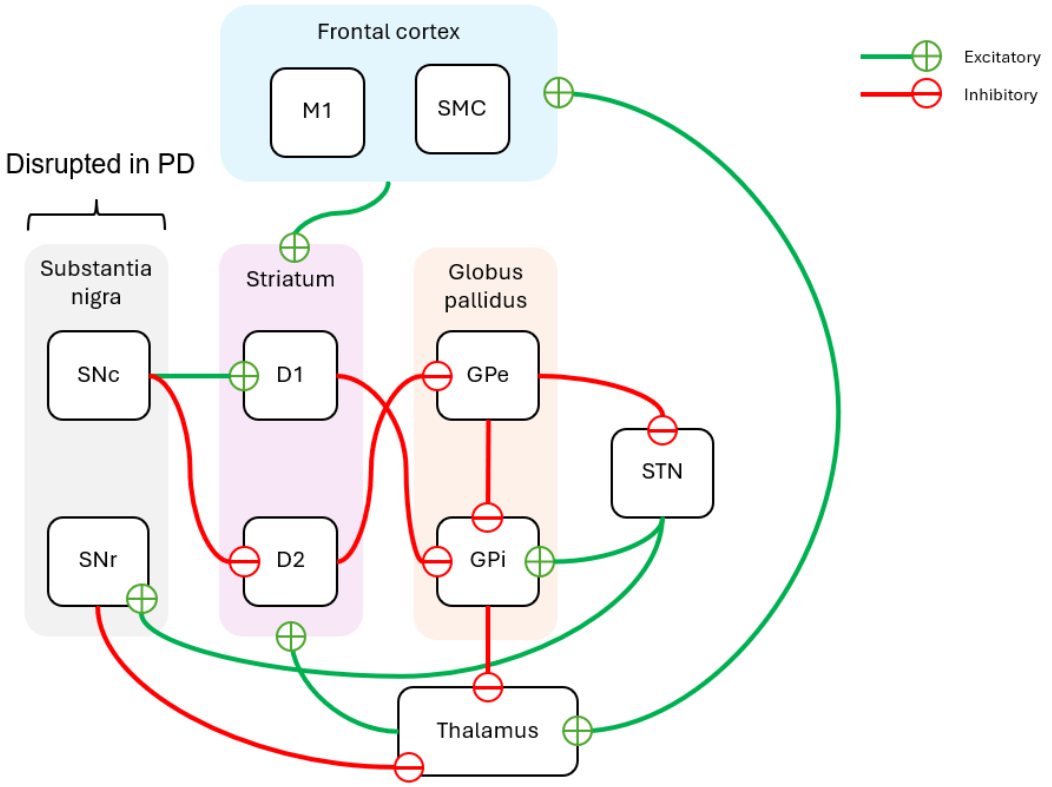

Figure 1. Schematic representation of neural pathways. The primary disruption in Parkinson's disease is the degeneration of dopamine-producing neurons in the substantia nigra pars compacta (SNc). This reduction in dopamine decreases activity in the GO pathway while increasing activity in the noGO pathway. The overactive noGO pathway (D2) leads to excessive inhibition of the globus pallidus externus (GPe) and heightened activity of the subthalamic nucleus (STN). The overactive STN then sends more excitatory signals to the globus pallidus internus (GPi) and the substantia nigra pars reticulata (SNr), which enhances their inhibitory effect on the thalamus. This excessive inhibition of the thalamus reduces excitatory feedback to the motor cortex, resulting in motor symptoms such as bradykinesia and rigidity, characteristic in Parkinson's disease.

In PD, the degeneration of dopamine-producing neurons in the substantia nigra leads to a breakdown in dopamine signaling, particularly affecting D1 and D2 receptors (Figure 1). Different neurons have different dopamine receptors, allowing signals to travel through specific pathways, labelled D1-D5 (Latif et al. 2021). Among these, D1 and D2 receptors are particularly important in the pathways that regulate movement. D1 receptors are primarily involved in the "GO" (direct) pathway, while D2 receptors participate in the "noGO" (indirect) pathway.

When dopamine levels drop, the "GO" pathway is less active, and the "noGO" pathway becomes more active (Przybyszewski et al. 2021). This shift results in increased activity in the globus pallidus externus (GPe), which in turn increases its inhibitory effect on the subthalamic nucleus (STN). The STN, which normally excites the globus pallidus internus (GPi), is now less active, leading to reduced excitation of the GPi. Moreover, the GPi becomes less inhibited overall due to the reduced activity of the "GO" pathway. As a result, the GPi applies more inhibition to the thalamus, which weakens the signals sent back to the frontal motor cortex (FC), including areas like the primary motor cortex (M1) and the supplementary motor area (SMA). This weaker signal slows down movements.

Initially, degeneration occurs in the substantia nigra and later affects other brain structures involved in movement control and adaptability. This progression leads to the characteristic symptoms of PD, such as tremors, rigidity, bradykinesia, and postural instability. The exact cause of neuronal loss in PD is not fully understood and can vary among individuals, making it difficult to identify a single biomarker for the disease.

Neurodegenerative changes of dopaminergic pathways are not always present in the same exact region of the brain of each patient (Chougar et al. 2022). For example, neurodegeneration may begin near the SN in some cases, while in others, it may start directly within the central region of the SN. This regional variability is particularly evident across different forms of Parkinsonism, including both typical PD and its atypical variants (Chougar et al. 2022).

Disrupted dopaminergic responses can be modeled artificially in computational simulations and measured using advanced imaging techniques like PET scans or fMRI (Langdon et al. 2018; Mollick et al. 2020; Reneman et al. 2021). Importantly, disrupted dopaminergic responses can be detected through digital tools like eye trackers and online cognitive testing platforms.

Overlapping Sets of Symptoms in Clinical Data

This lack of strict localization of the disease in the brain means that Parkinson's patients can have overlapping manifestations of other NDs, even including these, which are characteristic of Alzheimer's disease.

To describe these manifestations briefly, **[A5]** distinguishes three sets of symptoms, which are related to different structures in the brain being affected by the disease. These are:

- The set of motor symptoms (observed primarily in PD, less evident in AD).
- The set of cognitive symptoms (observed primarily in AD, secondarily in PD).
- The set of emotional symptoms (observed in both, but characteristic for late-onset AD).

Cognitive impairment in Parkinson's disease is diverse and varies in its impact on different cognitive functions, and it can manifest in several ways. While there are overlapping sets of symptoms between PD and AD, they are distinct diseases with different primary manifestations and treatment approaches.

However, the overlap of these sets complicates the diagnostic process, potentially leading to misdiagnoses or delays. Therefore, neurologists must use clinical evaluations and tests to determine precisely whether and how a person is affected by these diseases. These methods will be described in detail in the next chapter.

Importantly, there is an opportunity for digital tools that can measure all these disrupted dimensions together. By simultaneously measuring the sets of cognitive, motor, and emotional symptoms, digital applications can capture the complex structure of neurodegenerative diseases through the analysis of digital traces (Brem et al. 2023). However, in the context of machine learning, it is crucial to note that accurate prediction of symptoms brings challenges (Alfalahi et al. 2023).

Specifically, **[A6]** presents that when analyzing datasets, the symptoms observed in patients might be influenced by other factors, including comorbidities and individual variability in disease progression. This complexity necessitates models that can differentiate between overlapping symptoms and provide accurate predictions under these conditions.

Impact of Symptomatic Treatment on Diagnostic Data

After evaluations that led to clinical diagnosis, various treatments can support those affected by the disease. However, there is no cure for PD and AD, and only symptomatic treatment can be

offered. This includes both medication and non-pharmacological approaches (such as exercise and physical, occupational, and speech therapies) (Armstrong and Okun 2020). One of the primary medications used is levodopa, an "early" form of dopamine.

Levodopa can be ingested and transported through the blood-brain barrier, where it is converted into dopamine in the brain, helping to replenish the depleted dopamine levels. However, dopamine-based therapies typically help with initial motor symptoms. Nonmotor symptoms require nondopaminergic approaches (Armstrong and Okun 2020).

Other medications, such as dopamine agonists, MAO-B inhibitors, COMT inhibitors, and amantadine, can also be used to enhance or mimic dopamine's effects. Amantadine, originally used as an antiviral drug, seems to improve all cardinal motor features of PD (Crosby et al. 2003; Rascol et al. 2021). Interestingly, study **[C1]** used a digital questionnaire for data collection and statistical analysis indicated that amantadine reduces the severity of SARS-Cov-2.

Notably, in addition to medication, treatment plans may include therapies, and lifestyle changes to help manage symptoms and improve quality of life (Armstrong and Okun 2020).

Furthermore, advanced treatments, such as deep brain stimulation (DBS), can also be considered for some patients (Armstrong and Okun 2020). DBS involves surgically implanting electrodes in specific areas of the brain, such as the subthalamic nucleus (STN), which is highly connected to regions responsible for movement control, as described in "Disrupted Signal in Neural Networks" chapter. The electrodes deliver electrical impulses that modulate neural activity, emulating the effect of dopamine on the STN, and this mechanism can lead to reduction of motor symptoms. Therefore, this treatment can considerably enhance movement control in patients with Parkinson's disease.

Medical procedures like DBS require advanced digital imaging techniques (Butson et al. 2007). Machine learning models use 3D data from MRI scans to analyze and quantify neural structures by volume and size (Watts et al. 2020). This data provides valuable, objective input for ML models.

Lastly, in the context of data, it is important to note that patients with medicated PD may exhibit test results similar to those of healthy controls (Rosenblum et al. 2020). However, this similarity does not indicate the absence of PD, but rather that the therapy is effective. Consequently, this lack of clear differences makes it difficult for traditional diagnostic methods to accurately detect PD, requiring additional approaches for assessment.

Classification of Disease Progression

When measuring the progression of the disease it has to be noted that the assessment of cognitive function in PD is complex and requires careful consideration of both the disease-specific and general aging-related factors. However, regular cognitive evaluations, both baseline and serial, are necessary to capture the wide spectrum of cognitive impairment in PD (Goldman et al. 2018). Therefore, this section serves as a summary of different rating scales, used in the context of PD.

Standard neurological examination performed by neurologists in movement disorders includes mainly two tests, namely Unified Parkinson Disease Rating Scale (UPDRS; especially Part III) and the Hoehn and Yahr scale (Hoehn and Yahr 1967; Goetz et al. 2008; Baschi et al. 2020).

Scales of the Disease

UPDRS is a comprehensive tool used to measure the severity and progression of Parkinson's disease (Goetz et al. 2008). Part III of this test assesses motor function, including tremor, rigidity, bradykinesia (slowness of movement), and postural instability. It involves a series of tasks that the patient performs, which are then rated by the clinician. The UPDRS is sensitive to changes over time, making it useful for tracking disease progression and evaluating treatment effects.

The Hoehn and Yahr (H&Y) is the second commonly used scale to describe the progression of Parkinson's disease in stages, ranging from Stage 1 (mild symptoms) to Stage 5 (severe disability) (Hoehn and Yahr 1967). The scale provides a coarse categorization of disease severity.

In studies described in this thesis, neurologists provided measures of PD severity using UPDRS scale. This is because the higher granularity of the UPDRS (a wider scale of points versus H&Y) allows for more detailed understanding of the disease progression.

Cognitive Data

Moreover, neuropsychological testing is necessary for assessment of cognitive changes in Parkinson's disease (Goldman et al. 2018). The selection and performance of these tests is not strictly standardized and can be influenced by various factors including the patient's age, educational background, language proficiency, and physical capabilities. Particularly in PD, motor impairments can complicate the assessment. Thus, it is necessary to incorporate tests that do not heavily depend on manual skill or motor speed (Goldman et al. 2018).

Additionally, understanding cognitive function in PD requires insight into what constitutes normal versus pathological aging. This has to be done considering the specific motor and non-motor

symptoms of PD, the effects of medical or surgical treatments, and any co-morbid neuropsychiatric complications. Therefore, cognitive function assessments in PD are conducted. This approach evaluates a range of potentially affected cognitive areas (Goldman et al. 2018).

Thus, evaluations often include a combination of tests that assess global cognitive abilities and specific neuropsychological functions. These typically measure orientation, attention, executive functions, abstract reasoning, memory, language, perception, visuospatial abilities, praxis (the ability to plan and execute coordinated, purposeful movements), and motor skills (Goldman et al. 2018).

However, there is no standardized method for obtaining cognitive complaints, and cognitive screening tests often lack specificity for PD. This creates a need for more normative data, particularly for individuals from lower educational or socioeconomic backgrounds.

As mentioned, standardized functional assessments are also lacking, although some PD-specific tools exist (Goldman et al. 2018). Commonly used tools used for cognitive assessment include scales like Mini-Mental State Examination (MMSE) or Montreal Cognitive Assessment (MoCA) (Goldman et al. 2018). MoCA has proven to be more effective than some other cognitive tests (Smith et al. 2007; Zadikoff et al. 2008).

Behavioral and Temporal Data

Because impairment of motor control in PD is major issue, it is important to read cognitive data together with behavioral and temporal data that can provide more detailed insights into the disease (Lugtenborg et al. 2022). This approach assumes to directly measure motor performance and its degradation.

This can be done with a few methods, including the use of multiple, so-called *tracking tasks* (Lugtenborg et al. 2022). Tracking task involves visual monitoring and motor control to complete the sequences. These tasks are designed to evaluate the precision and consistency of motor movements over time, providing insights into the progression of motor control impairment. Additionally, the integration of behavioral and temporal data with cognitive assessments allows for a comprehensive analysis of how motor and cognitive functions interact in PD.

Affected Data

Moreover, in addition to cognitive testing, it's important to evaluate other factors that might affect cognition, such as mood disorders (depression, anxiety), behavioral issues (like psychosis), sleep

disturbances, and overall functional abilities. Interpretation of these elements is necessary for a complete approach to PD (Goldman et al. 2018).

Mood disorders, for instance, can worsen cognitive decline and affect the patient's quality of life. Behavioral issues, such as psychosis, can complicate treatment plans and patient management. Sleep disturbances are also critical to monitor, as they can significantly impact cognitive functions and daily living activities. Assessing overall functional abilities provides insight into the patient's capability to perform everyday tasks and maintain independence. Integrating these diverse factors in datasets is required for a comprehensive approach to PD.

Multidimensional Data

Furthermore, some screening techniques can help to address multiple dimensions in a consolidated way. For example, The Trail Making Test (TMT) is a neuropsychological test that measures diverse cognitive functions, including attention, visual search and scanning, sequencing, mental flexibility, and motor speed. TMT is commonly used to evaluate cognitive function in individuals with neurological conditions. It consists of two parts: A and B.

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.

Using these multidimensional assessments, it is possible to identify specific cognitive deficits, but also observe how these deficits interact with each other, resulting in the degradation of motor skills and emotional well-being. Tools like the TMT can, therefore, enhance the accuracy of diagnoses.

Imaging Data

Additionally, imaging data—the test results obtained from imaging tests—are often incorporated into PD research. This data often comes from Magnetic Resonance Imaging (MRI). Imaging biomarkers in PD are increasingly important for monitoring progression in clinical trials and also have the potential to improve clinical care and management (Mitchell et al. 2021).

MRI scans provide 3D images of brain structures and can show specific changes in brain anatomy associated with PD. These imaging biomarkers help in tracking the progression of the disease, and even identifying early signs of neurodegeneration. Incorporating imaging data into PD research and clinical practice allows for a more precise and individualized approach to patient care.

Digital Phenotyping

With the observable progress in machine learning, digital approach and applied informatics is becoming increasingly important in neurodegenerative disease research. This is supported by research presented in **[A5]**, which confirms that machine learning and digital tools can help in the early detection of neurodegenerative diseases, potentially enabling earlier interventions that could slow disease progression.

Current research focuses on validating digital biomarkers and ML-based predictive models in identifying early-stage neurodegenerative diseases. These approaches analyze data from various sensors and computerized versions of classical tests, such as the computerized Trail Making Test (cTMT) and computerized Montreal Cognitive Assessment (cMoCA). Additionally, digital tests can measure changes in attributes related to disease phenomena, collectively referred to as the "digital phenotype." This includes analyzing facial expressions (as disturbances may suggest hypomimia, the reduction of emotional facial expression), eye movements (which can be disrupted in PD), and data from accelerometers in wearables (which can potentially detect tremors or movement disruptions).

These digital studies in NDs are conducted by multiple research groups around the world, including the Digital Biomarkers in Neurodegenerative Disease Group (nd.pja.edu.pl), an initiative started by Przybyszewski in 2018, with focus on basic and applied research for the early diagnosis of neurodegenerative diseases. Recent research includes study **[C3]**, which was designed to identify AD biomarkers in healthy individuals by analyzing cognitive attributes from BIOCARD data, showing patterns indicative of potential early stages of AD. Using granular computing methods, the research predicted that some healthy subjects might develop varying degrees of dementia, validating the potential of machine learning for early AD detection and prevention. Furthermore, Śledzianowski (2024), in his PhD thesis, demonstrated that PD progression could be observed in eye movement and facial expression properties using machine learning methods (Śledzianowski 2024). Earlier, Szymański (2019) explored various digital biomarkers, such as MRI/DTI imaging and local blood flow in the brain, proving their effectiveness in classifying PD symptoms, including both motor and non-motor symptoms (Szymański 2019). His research also

investigated the use of eye movement data for classification models, suggesting broader applications for other neurodegenerative diseases. Collectively, these studies validate the use of digital technologies in neurodegenerative disease research and provide input for further investigation in the field, and findings described in **[A1-A8]** continue and extend research done within this group.

In summary, given the opportunities presented by previous studies in using selected methods, as well as the extreme challenge of managing and precisely detecting PD, this thesis focuses on three types of data sources: diffusion tensor imaging (DTI), eye-tracking, and online cognitive testing, along with selected ML algorithms to model the patterns in the data.

MRI and Machine Learning

Magnetic Resonance Imaging (MRI) plays a crucial role in the modern study of neurodegenerative diseases. Variations of this technique, such as Diffusion Weighted Imaging (DWI), provide insights into the structural details of the brain, allowing to visualize and quantify changes in brain anatomy and connectivity. These parameters can then be associated with diseases like Parkinson's and Alzheimer's, helping to model their progression using machine learning.

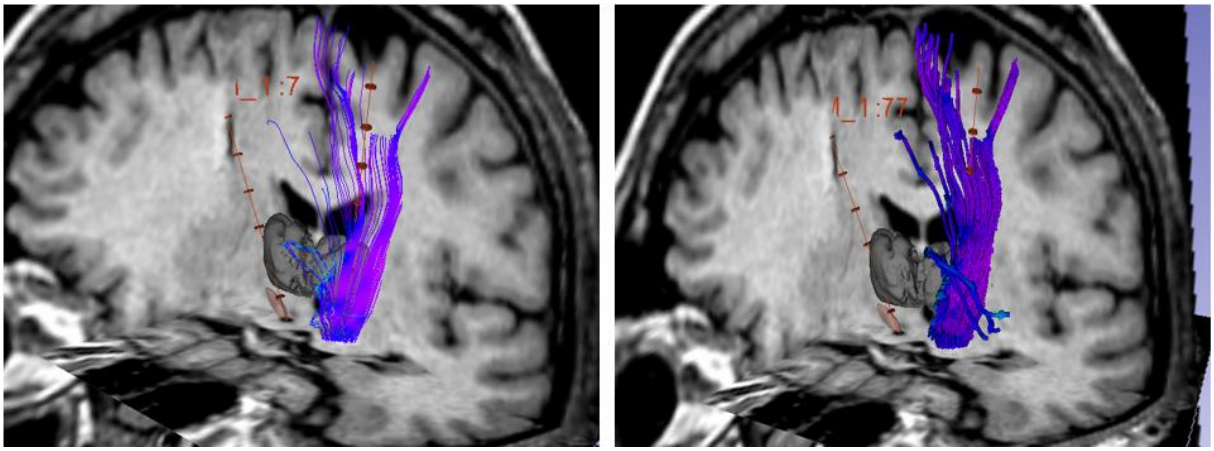

*Figure 2. Screenshot from 3DSlicer presenting a scan of a patient with exposed electrodes and neural tracts considered in the DTI study **[A1]** showing connectivity at different DBS stimulation amplitudes.*

In the context of PD, an especially interesting extension of DWI is Diffusion Tensor Imaging (DTI). This approach measures the behavior of water molecules over time, including both diffusion rates and directions, represented as 3D vectors (tensors).

DTI is particularly useful in neurodegenerative diseases because it captures tractography, showing the structure of nerve tracts. This is because water molecules show greater motion along the direction of nerve fibers, thus making DTI an effective tool for mapping these structures.

Furthermore, because some treatments for PD involve invasive procedures, mapping the brain's actual structure is crucial. Deep Brain Stimulation (DBS) involves inserting a thin electrode into the brain and is often associated with post-operative complications. DTI is a non-invasive method for visualizing neural structural connectivity, making it useful for understanding the impact of electrode contacts on the brain and potentially reducing complications (Przybyszewski et al. 2016).

Detailed mapping of brain structures is complex but required for planning and executing interventions effectively and safely, and some digital tools can support this process. For example study described in **[A1]** incorporated 3DSlicer for the registration of MRIs with an anatomical brain atlas (Figure 2). 3DSlicer (Harvard Medical School, Boston, MA) is software platform for the visualization, processing, segmentation, registration, and analysis of medical, biomedical, and other 3D images and meshes (Kikinis et al. 2014). It is peer-reviewed and a commonly used tool for planning and navigating image-guided procedures.

In the context of DBS, 3DSlicer allows for the overlay of preoperative MRI with postoperative DTI to study different pathways of connectivity between the DBS electrode and functional brain circuits. By combining these imaging modalities, it is possible to correlate stimulation parameters with the activation of specific brain structures for a more comprehensive understanding of both anatomical positioning and functional outcomes.

This approach is valuable for studying how different positions of the stimulating electrode contacts, and stimulation parameters influence brain functionality. This method differs from the traditional approach in which neurologists would assess the efficacy of DBS by observing patient symptoms over time, a process commonly referred to as "DBS tuning." However, analysis presented in **[A1]** shows that with advanced imaging techniques it is possible to predict and study stimulation effects in a more targeted and systematic manner.

Eye Movements

Following the importance of digital imaging, eye movements emerge as another crucial aspect in the study of neurodegenerative diseases. This is because of research which shows that simple reaction times, which decline in individuals with PD, can be effectively assessed through eye movement measurements. These measurements are useful in early diagnosis, monitoring disease progression, and differentiating between PD and other cognitive impairments.

Neurodegenerative conditions like PD affect the mechanisms controlling eye movements. For example, changes in saccadic delay have been linked to changes in the UPDRS III (motor

examination), indicating that disruptions in eye movement patterns are a sign of disease progression (Sledzianowski et al. 2019).

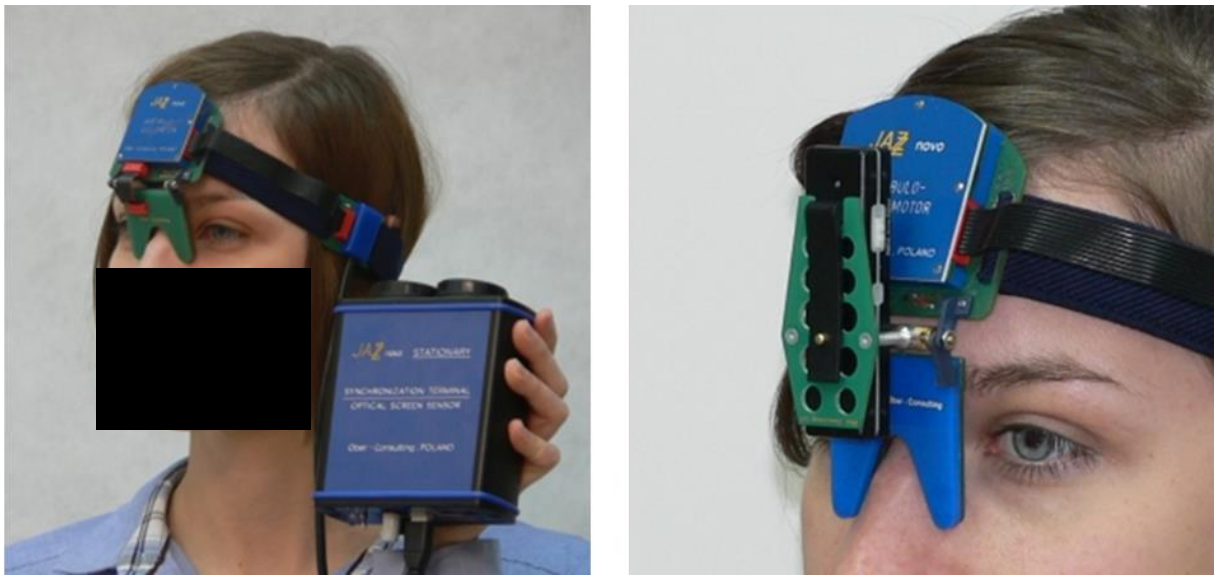

*Figure 3. Studies **[A3], [A4]** used a head-mounted saccadometer JAZZ (Ober Consulting, Poland), which was able to measure the reflexive saccades (RS) in the high frequency (1000 Hz). The advantage of this setup is lack of disruption from head movements.*

Eye movements can be categorized into saccades, smooth pursuits, and antisaccades, each with distinct characteristics. Saccades are rapid eye movements between targets and can be divided into visually-guided (reflexive – RS) saccades, and internally-guided (voluntary). Smooth pursuit refers to the continuous movement of the eyes as they follow a moving object, and antisaccades are eye movements in the opposite direction of a suddenly appearing target. When eyes stop scanning and focus on a single point it is called fixation.

For these types of eye movements, various parameters such as delay, duration, and maximum speed can be defined and measured. They are affected by age, as eye movement efficiency improves until the early 30s and then gradually declines (Hunfalvay et al. 2024). Interestingly, young children and elderly individuals show worse values in eye movements, while young adults had more stable eye movement patterns (including saccades, fixations, and pursuits).

The production of saccades is a complex process that depends on the coordinated excitatory and inhibitory inputs from various cortical and subcortical structures on the brainstem, where the motor commands for saccades are finalized (Klarendic and Kaski 2021). Dysfunction in this network impacts the delicate balance required for precise saccadic movements, leading to the characteristic eye movement abnormalities seen in neurodegenerative conditions.

These insights can be used for early detection and monitoring of neurological and cognitive disorders because eye movements are often susceptible to degradation in neurodegenerative diseases. Therefore, eye movements currently are a critical area of study in understanding and diagnosing neurodegenerative conditions such as Mild Cognitive Impairment (MCI) and Parkinson's disease (Liao et al. 2024).

Accurate measurement of eye movements requires specialized hardware and software. The most common method is video-based eye tracking, which can be performed using either specialized eye trackers or webcams. While webcams have lower frames per second (FPS) compared to dedicated eye trackers, **[A5]** describes that they can be used with convolutional neural networks (CNN) to achieve accuracy comparable to standalone eye trackers.

As a benchmark for these approaches, outputs from more specialized devices that can capture data with high frequency often serve as a standard. Therefore, studies described in **[A3]** and **[A4]** used a head-mounted device with an infrared (IR) sensor, specifically the JAZZ Novo from Ober Consulting, Poland. The advantage of this setup is lack of disruption from head movements and ability to capture eye movements at 1000 frames per second, providing high frequency and precision of data (Figure 3).

Computerized Tests

Because of the importance of not only motor but also cognitive data, and the extension of findings from eye tracking regarding the possibility to detect bradykinesia (slowness of movement) through eye movement, there is an opportunity for computerized tests that can capture not only cognitive but also motor changes using different input methods, such as cursor movements.

However, there is an ongoing discussion whether digital cognitive tools are comparable and trustworthy in the same way as their pen-and-paper equivalents. Therefore, Latendorf et al. (2021) evaluated that paper-based and digital versions of all test procedures correlated highly with each other (Latendorf et al. 2021). This validity was further concluded by others, including Linari et al. who validated the computer version of TMT (cTMT) showing that the reaction times (RTs) and performance profiles are consistent with the classic TMT (Linari et al. 2022).

Moreover, Linari et al. (2022) presented an identical saccade duration, saccade amplitude, and fixation duration suggest a similar visual mechanism between parts A and B (Linari et al. 2022). The observed difference in the number of fixations might originate in a more complex processing of the task, more related to higher-order cognitive processes than visual mechanisms. In other

words, to solve both parts of the task, subjects seem to use their visual machinery in a curiously similar way, except that TMT-B requires a more intensive scanning of the visual scene.

Interestingly, more complex tests, including digital version of Montreal Cognitive Assessment (cMoCA) require longer processing time in their pen-and-paper counterparts (Snowdon et al. 2015; Latendorf et al. 2021). This is related to the use of digital devices or user-friendliness because questions include tasks such as drawing a cube and a clock, which can be challenging depending on the approach used.

To validate this approach, studies **[A6-A8]** describe a custom online platform to administer a group of neuropsychological tests, including computer versions of both the MoCA and the Trail Making Test (Figure 4), along with other tests evaluating well-being and cognitive state.

This custom platform incorporated additional measures, specifically both instrumental reaction time (IRT) and time-to-submit (TTS) for each task, as adding extra temporal measures to cognitive tests (such as TMT-B) can improve the accuracy of predicting UPDRS III group classification. This platform is an example of a test that integrated behavioral, temporal, and cognitive data without requiring specific hardware.

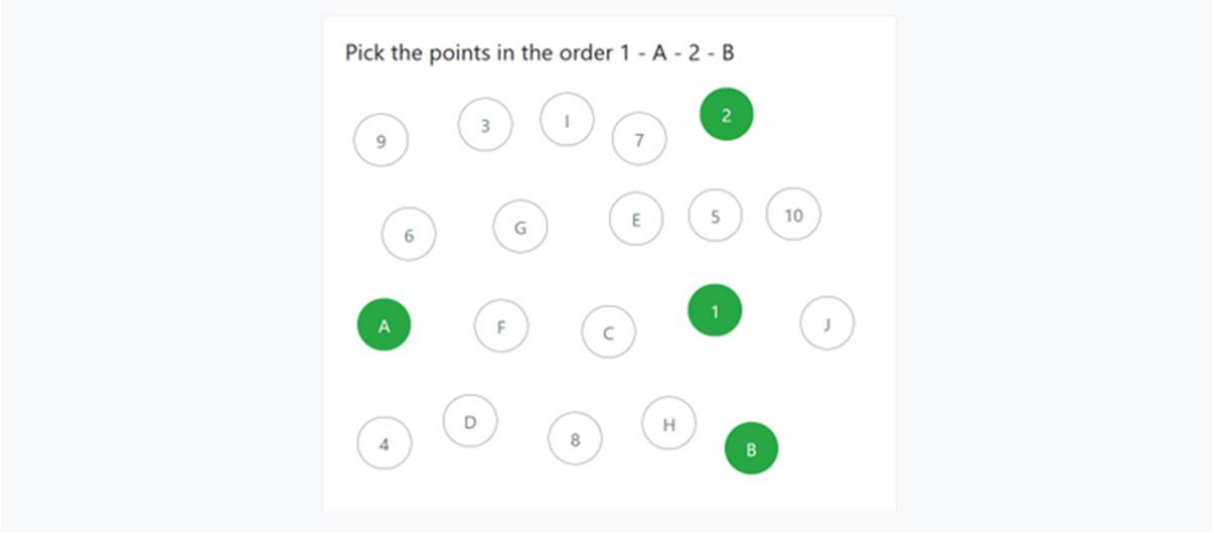

*Figure 4. The image presents a single selected screen from interactive cognitive task screen where the user is asked to pick points in a specific sequence. Web TMT B measures errors, initiation time and execution time **[A8]**.*

III. Models for Classification of Parkinson's Disease

To describe better the challenge of PD classification, there are two primary approaches to data modelling (Figure 5). The first, most common is binary classification. This approach focuses on differentiating between individuals with Parkinson's Disease and those without (healthy controls - HC). The goal is to develop models or algorithms that can accurately classify a given individual into one of these two categories based on various types of data. The second approach involves disease stratification, determine the stage or severity of Parkinson's Disease in patients. It often involves creating models to predict disease progression or to categorize patients into different stages of the disease.

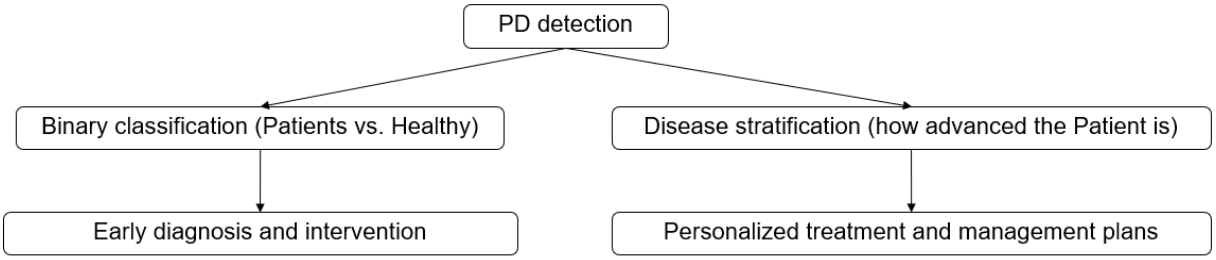

Figure 5.The two main approaches to PD detection and their respective outcomes.

Proposed Workflow

Despite differences in objectives, as well as in ML models, there is a common approach to evaluation of machine learning workflow. Importantly, managing limited datasets requires careful assessment (Figure 6).

Therefore, the process begins with the sensor, which is able to register relevant parameters. Data from this sensor is then placed in the database, to keep the record of the patient. The next step in the workflow involves data preprocessing (profiling, cleansing, reduction, transformation, enrichment, and validation - including statistical analysis).

After confirmation that the dataset is complete, a train_test_split approach should allocate around 30% of the data for testing before modeling. This split is done randomly to ensure that 70% of subjects were used for training and 30% for testing, preventing systematic bias in the distribution.

To further mitigate the risk of overfitting, cross-validation techniques are necessary to apply during the model training phase. Thanks to this approach, it is possible to assess the model's performance more accurately and ensure its generalizability to new, unseen data.

Selected model-specific techniques can help to address the issue of overfitting. For models like Logistic Regression, **[A7]** describes how to apply regularization methods to penalize complexity and discourage the learning of noise from the training set. This step helps to prevent overfitting, making models more robust to unseen data. The final step is the application of testing dataset on tuned models in order to validate the predictions.

In practice, this workflow helps to minimize the risk of overfitting and data leakage and impacts the reliability of the results.

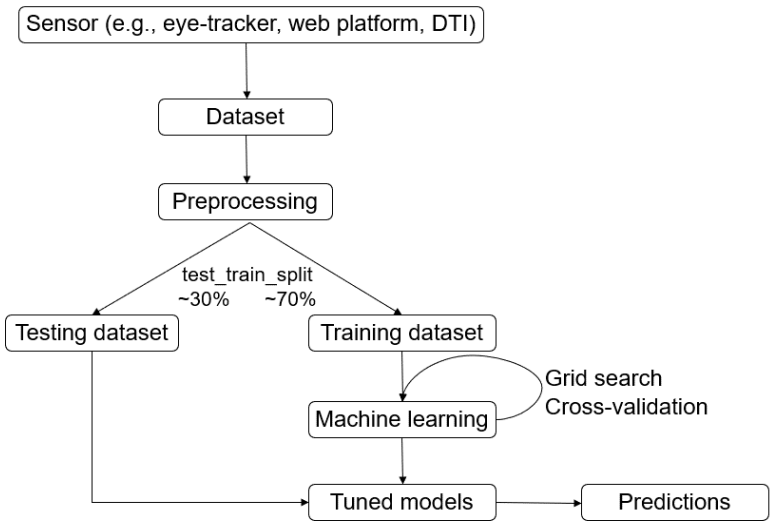

Figure 6. Workflow for ML evaluation. Starting from sensor data collection, the workflow proceeds through data preprocessing (profiling, cleansing, reduction, transformation, enrichment, and validation). The dataset is then split into training (70%) and testing (30%) sets. Machine learning models are trained with cross-validation and grid search for hyperparameter tuning. Model-specific methods, such as regularization, are applied to prevent overfitting. The tuned models are validated using the testing dataset to generate predictions.

Statistical Analysis

As presented in the workflow, the main step after dataset assembly is data preprocessing. An important part of this step includes statistical analysis. Summarizing the statistical main features of a dataset helps to review collected data. This includes calculating measures of central tendency (mean, median, mode), measures of variability (range, variance, standard deviation), and creating visual representations like histograms and charts. Testing differences between two

groups by means of attributes (e.g., using t-tests, chi-square tests) provides an overview of similarities and differences between them.

Two approaches assume either using specialized software (e.g., IBM SPSS Statistics) or calculating statistics using programming languages, such as Python. Python offers open-source libraries such as Pandas for data manipulation, NumPy for numerical computations, Matplotlib and Seaborn for data visualization, and various libraries for statistical analysis (Hunter 2007; McKinney 2010; Pedregosa et al. 2011; Harris et al. 2020; Waskom 2021).

Studies **[A6-A8]** used both approaches, combining specialized statistical software and code-based methods. Using SPSS for basic statistical analysis gives a good overview of the dataset. However, it is worth visualizing more in-depth connections between variables, for example using the Pearson's correlation coefficient, which can be calculated using Python and Pandas library. Furthermore, the Seaborn library can help to visualize these correlations in the form of a heatmap (Waskom 2021). Pearson's correlation is considered an effective method for assessing associations due to its dependence on covariance, making it easily interpretable.

After processing and dividing the data, the next step is machine learning modeling. This is important because statistics are often not enough to draw conclusions from the complexity and variability seen in digital biomarkers. In essence, when statistical tests fail to find something "significant," machine learning models can start to detect more complex, nonlinear relationships that are not immediately obvious.

Machine Learning

Often when we look at data, it can be difficult to make sense of the information it contains (Mahesh 2020). The patterns or insights might not be immediately clear, especially if we have collected a lot of variables. In these situations, we can turn to machine learning, which is a type of technology that helps us find hidden patterns and make predictions based on the data (Mahesh 2020). Machine learning algorithms can analyze complex datasets and extract meaningful information that might not be obvious to the human eye.

However, conclusions have to be understandable for scientists and practitioners due to the necessity for precision medicine. Therefore, there is a need for systems that can process information in a way similar to human thinking, which often involves abstraction and dealing with incomplete information. This need drove the development of various machine learning models

(function-approximation approach), and its subset of methods described by granular computing (rule-based approach).

Machine learning techniques can be broadly categorized into supervised learning, unsupervised learning, and reinforcement learning. In the context of PD research, supervised learning is often used, where the algorithm is trained on a labelled dataset, learning to make predictions or decisions based on input data. Studies **[A3]**, **[A4]**, **[A7]**, **[A8]** employed a range of these ML techniques to model the patterns in medical data.

Primarily, ML algorithms can approximate functions that map inputs to outputs, thereby identifying patterns. This is crucial because, in computational terms, the task of PD detection can be defined as a multiclass classification problem, with the objective of predicting the disease severity level across various groups of individuals. Therefore, the following sections present the foundations of computational theory (formalization of a problem) and the selection of the best-performing algorithms, including Logistic Regression, Random Forest, and Rough Set approach.

Computational Approach

Computationally, a medical dataset consists of N training samples $(x_1, y_1), \ldots, (x_N, y_N)$, where each x_i represents the vector of features (e.g. results of tests) for the i -th sample (e.g. for each patient), and y_i is a corresponding class label (e.g. the class of disease severity). The goal is to find a function $g: X \to Y$. This function is mapping from the input space X to the output space Y. Here, the function g belongs to a set of potential functions G, referred to as the hypothesis space. This is the set of all functions that the learning algorithm can choose from.

In the light of that, machine learning algorithm creates and adjusts a mapping function from input features (e.g., patient results) to output labels (e.g., medical diagnoses), using a given set of training samples. This is a non-trivial task and thus there are a couple of specific approaches to this general problem, resulting in a wide selection of machine learning algorithms.

This means however that there is no single universal solution to all machine learning problems. The choice of model (e.g., Decision Trees, Neural Networks, Support Vector Machines, Logistic Regression, Random Forest, RSES) depends on the nature of the data and the specific characteristics of the disease patterns. Ultimately, the best-performing models can be embedded in a broader ecosystem of “artificial intelligence” (AI) - a system capable of making decisions or predictions that we perceive as “intelligent.”

Optimization and Verification

Because of this challenge with model selection, there is a group of studies that focus on the analysis and the evaluation of outcomes using selected machine learning algorithms in the context of Parkinson's disease classification (Sharanyaa et al. 2020; Ouhmida et al. 2022; Biswas et al. 2023). This analysis is important, because it leads to the selection of the proper tool for the prediction of PD. This comparison can be done in multiple ways.

Therefore, studies **[A3], [A4], [A7],** and **[A8]** selected multiple algorithms to compare them side by side. By evaluating different models, they target to identify the most effective approach for accurately predicting disease severity and progression in PD.

A commonly used and simple metric is the accuracy of prediction. This parameter gives a fundamental overview of the performance of each classifier as it is easily understandable. Furthermore, accuracy score impacts multiple metrics as well. These other metrics, like precision, recall, F1 score, ROC (Receiver Operating Characteristic curve), and AUC (Area under the ROC curve) are often verified for better overall performance overview.

In PD, by the means of these metrics, the literature reveals that common selection of best performing algorithms includes Logistic Regression, Random Forest, Naïve Bayes, Decision Trees, K-Nearest Neighbours, Artificial Neural Networks, Support Vector Machine, and Gradient Boosting Machines (Sharanyaa et al. 2020; Ouhmida et al. 2022; Biswas et al. 2023).

Studies described in **[A3]**, **[A4]**, **[A7]**, and **[A8]** evaluated selected ML models, often together with granular computing approach. In the classic ML approach, the highest performance (in terms of accuracy and precision) was consistently achieved with Logistic Regression and Random Forest. However, as discussed later, these methods were often outperformed by the Granular Computing approach utilizing rough-set rules. In light of these findings, the following sections provide a brief overview of the selected algorithms.

Logistic Regression

Logistic Regression (LR) predicts categorical outcomes by modeling a linear relationship between the input variables and the log-odds of the outcome. The core of Logistic Regression is the sigmoid function, which is defined as: $p = \frac{1}{1+e^{-(b_0+b_1X)}}$. Here, p is the probability of the positive class, b_0 represents the baseline log-odds when all predictor variables are zero (intercept term), b_1 represents the strength and direction of the association between X and the outcome (coefficient for the predictor variable X). This formula is useful because it transforms the log-odds

into output values between 0 and 1, modeling them as probabilities. While the basic form of Logistic Regression is binary, it can be extended to manage multinomial outcomes. LR is particularly popular in healthcare analysis and medical research and yielded good results in the study **[A6]**.

Random Forest

Study **[A4]** shows that Random Forest (RF) presents good performance in complex PD classification. The core of Random Forest are outcomes of multiple Decision Trees. Each individual tree provides a prediction based on a random subset of the data and features. In RF, the final prediction for an observation is usually obtained by aggregating the predictions from all individual trees, either by averaging (for regression tasks) or by taking the mode (for classification tasks). Notably, Random Forest provides a metric called "feature importance" that indicates how much each feature contributes to the model's predictions, helping to identify the most influential variables in the dataset. As with Logistic Regression, the approach offered by RF is important for making informed clinical decisions and for identifying potential areas for intervention.

Granular Computing

Analysis of classical ML methods' performance versus the Rough-Set Theory (RST) described in **[A3]** suggests that granular computing approach using rough-set rules is able to provide good predictions on medical datasets (Pawlak 1982).

In the concept of granular computing (GrC), granules are collections of elements grouped together based on their similarity, functionality, or other criteria. These can be data points, objects, or pieces of information. Objects characterized by the same information are considered indiscernible. Any set of all similar objects forms a basic granule (or atom) of knowledge about the universe. Each set can be described by rules that define which objects certainly belong to the set (lower approximation) and which objects "possibly" belong to the set (upper approximation). Rough Set Theory can be applied to small datasets. Moreover, generated rules often are discovered in scenarios where comprehensive models might not be feasible.

Computational Theory of Mind

In the discussion about models that can process information in a way similar to human thinking, the brain's functioning is often compared to that of a digital computer or a Universal Turing Machine, which processes symbols (Turing 1937). However, psychophysical experiments and humans' ability to recognize complex objects, such as faces, in various contexts and lighting conditions, suggest otherwise. This challenges the idea of symbolic representation and instead

supports the notion that concept representation based on similarities may be a more accurate model for how the brain works.

While others interpreted principles of universal computing machine as a model of the brain, Turing himself had a different perspective on how the brain functions. Besides his development of the UTM, Turing contributed to the field of morphogenesis (Turing 1990). He proposed that natural patterns like stripes, spots, and spirals can arise from the interaction of chemical substances, which he called "morphogens." This is a process where chemicals move between cells causing them to "transform," causing the organism to develop its shape or specific pattern. This theory offers a dynamic view of biological development that contrasts with the fixed, predetermined computational process implied by the Universal Turing Machine.

Biological systems, including those studied in morphogenesis, often involve uncertain information. Rough Set Theory manages data where some information may be imprecise or missing. Therefore, **[A5]** argues that simple rules correspond with morphogenesis principles and allow for a nuanced understanding of cognitive processes.

This comparison can be explained using visual classification tasks to understand how the brain processes information. Visual classification involves specific neural interactions, such as asynchronous process integration in the retina. The visual classification model is based on the receptive field properties of neurons in different visual areas and uses both feedforward and feedback interactions between them (Figure 7).

In higher visual brain areas, people use processes similar to granular computing to identify upper and lower approximations of the retinal image. These approximations are compared with different object models stored in the visual cortex. As object recognition progresses, lower visual areas are tuned to extract properties of the selected model, and the gap between the upper and lower approximations narrows. When the border set becomes empty, the object is successfully recognized. This approach can be applied to propose various models that approximate the actual (and future) state of tested subjects.

Specifically, the feedforward pathways use "driver logical rules" to combine properties extracted in each area into hypotheses related to possible objects. Then, the feedback pathways use "modulator logical rules" to help transform weak concepts of objects' physical properties into crisp classifications in psychophysical space. This process approximates how the brain utilizes logical rules to transform blurred object concepts into clear categorizations.

Studies on macaques, such as those by Przybyszewski et al. (2000) show that back-projection pathways from the striate cortex (V1) to the lateral geniculate nucleus (LGN) enhance neuron responses based on visual contrast, supporting the theory of visual processing and object recognition. (Przybyszewski et al. 2000). These are connections that go from one part of the brain and back to an area that supplies input to it. Research shows that these back connections make neuron responses stronger, depending on the contrast of what is seen. Here as well, as object recognition progresses, lower visual areas are tuned to extract properties of the selected model, and the gap between the upper and lower approximations narrows (object is recognized).

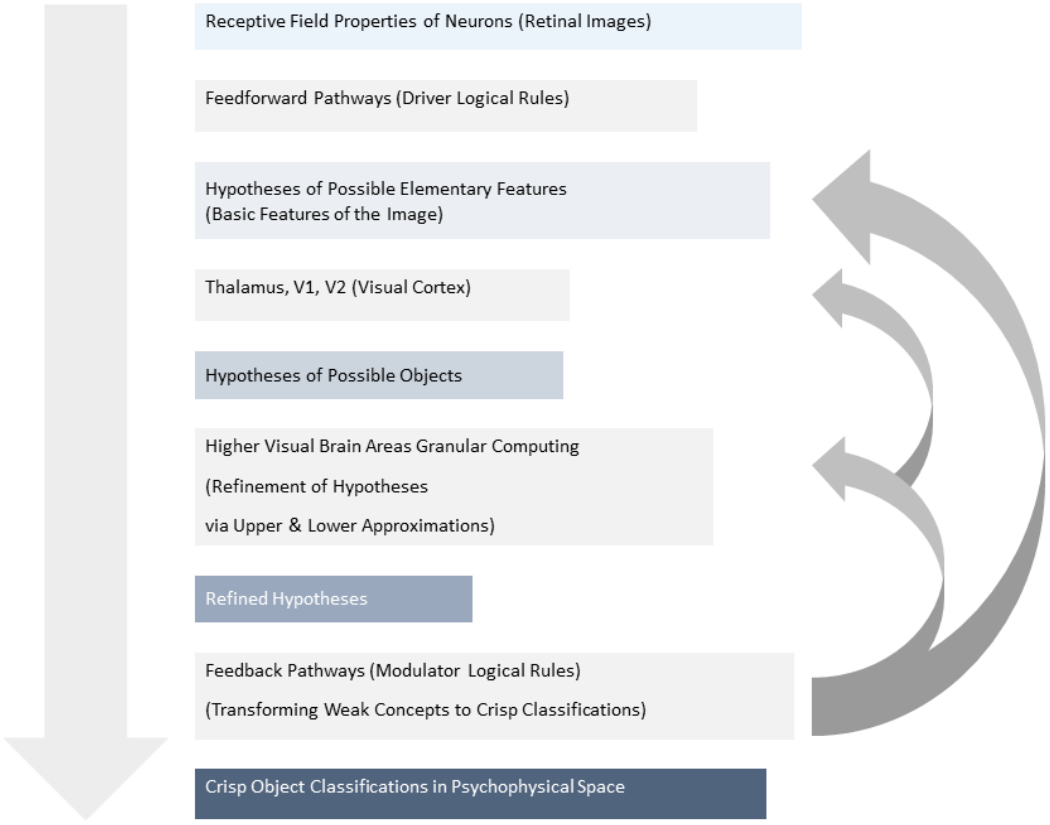

Figure 7. The receptive field properties of neurons feed into both feedforward and feedback pathways. Feedforward pathways use "driver logical rules" to create hypotheses of possible objects, while feedback pathways employ "modulator logical rules" to refine these hypotheses into crisp classifications. Higher visual brain areas then utilize GrC to compare upper and lower approximations of retinal images with stored object models, leading to object recognition as the gap between these approximations narrows ***[A5]***.

This observation gives an opportunity to reconsider the algorithms we use to develop digital diagnostic tools. Insights from how biology works, as suggested by Turing's theories and recent research on xenobots (so-called "living robots," Levin et al. 2021), suggest that simple rules can

create a complex organism (Blackiston et al. 2021). This view is maintained throughout all papers in this thesis, and it is contrary to computationally heavy, AI-driven "black box" methods.

IV. Contributions

The results of this research combine three distinct approaches towards data collection: using diffusion tensor imaging, eye tracking, and online cognitive testing. Each method was integrated with machine learning algorithms to analyze the data.

DTI and Machine Learning

The first study, **[A1]** ("DTI Helps to Predict Parkinson's Patient's Symptoms Using Data Mining Techniques", 2019), focused on postoperative (after DBS) patients. This approach involved ML to identify brain abnormalities through DTI analysis (Figure 2). In this work, I conducted the analysis of RST model results and provided a detailed description of the methods and findings. The initial analysis included discovering correlations between given attributes aligned with their importance using correlation heatmap. Then, based on the obtained values, it was possible to select relevant attributes for the data mining process, performed in RSES software.

The study included three subsets of the dataset, each targeting a different UPDRS II or UPDRS III marker: "#8: handwriting distortion", "#20: tremor at rest: face, lips, chin," and "#21: action or postural tremor of hands - left hand." For each subset, I conducted experiments based on defined attributes, leading to conclusions about possible UPDRS values after DBS treatment. The results of the prediction for the left-hand tremor based on given DTI parameters revealed an accuracy of 0.967 with a coverage of 0.425. The prediction result for the "face, chin, lips tremor" achieved an accuracy of 0.824 with a coverage of 0.7. The last experiment, detecting UPDRS value changes for the "handwriting distortion," achieved an accuracy of 0.878 with a coverage of 0.667.

Using rough set rules, the model detected issues such as handwriting distortion after deep brain stimulation treatment with notable accuracy. DTI analysis focused on finding ways (connections) to determine the structure of the motor cortex depending on the stimulation parameters and contact with electrodes. This approach provided insights into specific brain microstructural changes that were different in PD patients.

Further research with independent cohorts is necessary to validate and extend these findings, because this approach provided valuable insights into disease progression and helped differentiate between PD subclasses.

Eye-Tracking and Machine Learning

The second approach involved analyzing rapid eye movements, specifically saccades, to identify patterns that differentiate PD patients. Saccades were measured using head-mounted eye-tracker (Figure 3).

Study **[A2]** provides analysis of this approach, and reviews various tests, using saccades, antisaccades, and pursuit, to predict disease progression using ML methods. The predictive algorithms described in the study included a range of techniques, such as granular computing, Naive Bayes, Decision Trees, Logistic Regression, C-/Linear Support Vector Classifiers, K-Nearest Neighbors Classifier, and Random Forest. The review demonstrated that by applying ML algorithms to high-dimensional eye movement data, it is possible to detect significant changes related to neurodegenerative symptoms, proposing new insights into the underlying mechanisms of these diseases.

Furthermore, in the study **[A3]** ("Comparison of Different Data Mining Methods to Determine Disease Progression in Dissimilar Groups of Parkinson's Patients", 2020), the objective was to predict the results of different PD patients' treatments to find an optimal one. In this study I compared Rough Sets and various machine learning models, including Logistic Regression, Random Forest, Decision Tree, KNN, SVC, GradBoost, and GaussianNB, to describe and predict disease progression expressed as UPDRS values. This study showed that eye-tracking can play a crucial role in identifying different stages of PD. Because of the extensive granulation of the task, I coded the pipeline for the evaluation of ML models (Rough Sets were modeled independently), where each model was given the same task and the ability to hypertune itself in the best possible way. When each model was prepared, they reported their scores and hypertuned parameters to analyze their behavior. This study included three groups of PD patients: 23 BMT (Best Medical Treatment) patients on medication, 24 DBS (Deep Brain Stimulation) patients on medication and DBS therapy after surgery, and 15 POP (Postoperative) patients who had undergone surgery earlier. Each patient had three visits every six months. The study estimated UPDRS changes based on disease duration, saccadic eye movement parameters, and neuropsychological tests (PDQ39 and Epworth Sleepiness Scale). The results showed that RS rules predicted UPDRS values with global accuracies of 70% for BMT visits, 56% for DBS , and 67% and 79% for POP visits. ML models provided similar accuracy ranges, with the best performance in specific sessions. However, in contrast to RS rules, ML models could predict the results of one session (session 3 - MedON) in POPW1 patients based on DBSW2 session three results. This is a limited case but

shows some common mechanisms between DBS and POP groups. In this work, I was responsible for the creation and evaluation of ML models, as well as the description of the results.

This was further investigated in **[A4]** ("Eye-Tracking and Machine Learning Significance in Parkinson's Disease Symptoms Prediction", 2020). Here, I used data of patients from the BMT group (third visit), DBS (3rd visit), and POP (1st visit). The independent test set consisted of the POP group from the second visit. Using an upgraded version of the pipeline code for ML evaluation, I selected, hyper-tuned, and observed the behavior of ML models under this task, and I described the outcomes. In this study, by combining eye-tracking data with ML algorithms, the model achieved a 75% accuracy rate in predicting subclasses of UPDRS for patients in advanced stages of the disease. The overall accuracy for distinguishing all classes was moderate at 57%. However, it is important to highlight that this model was designed to solve a multi-class (four-class) classification problem among dissimilar PD groups. While using fewer groups would lead to better overall accuracy, the multi-class approach provides a more granular view of the data, capturing rules among the different stages and characteristics of PD.

Despite the scores presented by these results, it is worth noting that the study was limited by sample size, which may affect the generalizability of the findings. Further research with larger cohorts is necessary to validate these results and fully establish the utility of eye-tracking as a non-invasive tool for PD diagnosis and monitoring. Nevertheless, the ML models were able to detect distinctive eye movement patterns characteristic for different stages of the disease.

Digital Cognitive Tests and Machine Learning

The newest approach involves using data from web-based cognitive tests to identify motor and cognitive signs of Parkinson's Disease. This is an outcome of the review **[A5]** that explored digital biomarkers (tools designed for remote neurocognitive data collection and ML analysis) as a potential solution for early diagnosis of NDs. Integration of web digital tools holds a potential to transform ND diagnostics, making early detection tools more cost-effective and globally accessible.

For the evaluation of this approach, I created a web test to acquire specific motor (instrumental reaction time - IRT and time to submit - TTS) and cognitive (Trail Making Test A and B, Montreal Cognitive Assessment) data as biomarkers for disease progression. In 2023, I presented first preliminary findings of the web platform and initial model for data. Study **[A6]** ("Investigating the Impact of Parkinson's Disease on Brain Computations: An Online Study of Healthy Controls and PD Patients", 2023), collected data from individuals with PD and healthy controls using an online

platform and multiple neuropsychological tests. Using Logistic Regression, the model achieved an accuracy rate of 91.1% in differentiating PD patients from healthy controls. However, two PD patients were misclassified as healthy subjects, and two healthy individuals were misclassified as PD patients. I also utilized multinomial Logistic Regression to predict the UPDRS III group of patients and healthy individuals, achieving similarly high accuracy. These findings suggest that web cognitive and behavioral tests can detect changes in brain computations, potentially indicating the onset of PD before clinical symptoms appear.

In the subsequent study **[A7]** ("Classification of Parkinson's Disease Using Machine Learning with MoCA Response Dynamics", 2024), I examined the diagnostic efficiency of ML models using MoCA test results to classify scores of people with PD and healthy subjects. I implemented both rule-based modeling using Rough Set Theory and classic ML techniques such as Logistic Regression, Support Vector Machines, and Random Forests. The diagnostic accuracy of the best-performing model (RST) increased from 80.0% to 93.4%, and diagnostic specificity increased from 57.2% to 93.4% only when the MoCA score was combined with proposed temporal metrics (IRT and TTS). This highlights that online platforms can detect subtle signs of bradykinesia (slowness of movement, a hallmark symptom of PD) and use this as a biomarker to provide more precise and specific diagnoses. Despite the constrained number of participants (15 PD patients and 16 healthy controls), the results suggest that incorporating time-based metrics into cognitive screening algorithms may significantly improve their diagnostic capabilities.

Finally, in 2024, I used the same platform to distinguish between symptoms of mild-PD and advanced-PD. In study **[A8]** ("Recognizing Patterns of Parkinson's Disease using Online Trail Making Test and Response Dynamics – Preliminary Study", 2024), I evaluated the effectiveness of an online version of TMT A and TMT B, incorporating time-based measures to recognize cognitive and motor manifestations of PD severity. For validation, this research was conducted with 15 PD patients. The study applied TMT sensitivity to executive function impairments by measuring response and reaction times and correlating these with stages of PD severity. I used various ML models, including Naïve Bayes, Logistic Regression, Support Vector Machine, and Random Forest, to predict disease severity based on TMT performance. Among these, Random Forest was the most effective, achieving 80% accuracy (AUC=0.92), indicating good performance in distinguishing between mild and advanced stages of PD. Although limited by sample size, this preliminary study highlights the role of digital tools in capturing specific changes in PD that can be relevant for diagnostics and monitoring of PD.

Consequently, findings from these studies present the utility of web-based platforms in assessing cognitive and motor functions in PD patients. By integrating temporal dynamics with traditional cognitive tests, we can achieve higher diagnostic accuracy and specificity.

This approach not only assists early detection but also enables continuous monitoring of disease progression in a non-invasive manner. Importantly, the results suggest that the inclusion of digital tools in routine clinical practice can support the detection and management of neurodegenerative diseases.

Research in Virtual Reality

Furthermore, it is worth noting that patients are experiencing navigation problems in 3D space in both Alzheimer's disease and Parkinson's disease. Thus, I also evaluated the possible impact of virtual reality (VR) on early disease detection as a parallel direction for diagnosis (Figure 8).

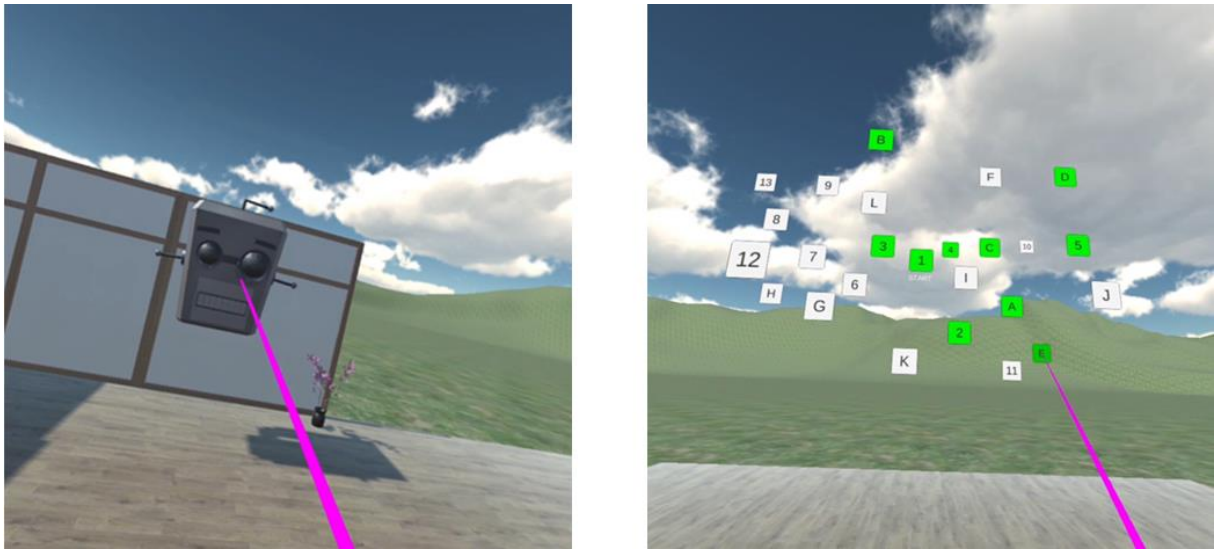

Figure 8. The prototype of the application assessing the recognition of numbers and letters in the right order (TMT B) implemented in the VR environment on the Oculus Go device (demo online: nd.pja.edu.pl/ar-vr.html).

As confirmed in web platform studies, TMT Part B is sensitive to executive functioning since the test requires multiple abilities to complete it, and thus it is more difficult in 3D. My role in this area was to develop a prototype application for assessing the recognition of numbers and letters in the correct order within a VR environment using the Unity engine (demonstration is available online: nd.pja.edu.pl/ar-vr.html). In this context, the challenges of the virtual environment, such as cybersickness, are described in a "Further Research" section of **[A2]**. The potential of VR technology, however, should not be dismissed.

As discussed in the referenced literature, VR applications have shown significant promise in cognitive assessments, particularly in tasks involving visuospatial processing, which is critical in the early detection of Alzheimer's disease and other neurodegenerative conditions. But accessibility issues and hardware requirements cannot be overlooked. Thus, future research should focus on optimizing these VR tools for broader clinical use, especially because it creates an immersive environment.

This is important because recent research has shown that engaging in action video games can lead to structural changes in the brain, particularly in regions associated with sensory processing, attention, and spatial navigation. Interestingly, it is possible to induce changes in gray matter thickness while playing action video games, as found in **[C2]**. The structural plasticity of the insula could be linked to the quality of the gaming experience and the skills developed during gameplay. However, due to the additional hardware overhead required for VR, I suggest that the web approach holds promise for being more accessible by not needing dedicated hardware and software.

Summary of Contributions

In summary, my contributions to the field include the development of digital assessment platforms integrated with machine learning models for complex datasets, as well as the implementation, evaluation, analysis, and interpretation of the results.

I implemented and evaluated various ML models for the specific challenges posed by complex Parkinson's disease datasets. This work involved researching algorithms, tuning model parameters, and ensuring that the models effectively captured and predicted patterns related to disease progression.

Furthermore, I conducted a detailed analysis and interpretation of the results generated by these ML models. This included measuring performance and improving the explainability of the models, making their predictions more understandable and actionable.

A significant aspect of my contributions was the development of a web-based cognitive assessment platform designed to collect and process both cognitive and motor data from patients. This platform provided data for the ML workflow and demonstrated its capability to generate insightful datasets that register subtle changes in disease progression.

Additionally, I proposed and validated two new behavioral digital biomarkers: Instrumental Reaction Time (IRT) and Time To Submit (TTS). These biomarkers were incorporated into the

assessment process and significantly enhanced the accuracy and specificity of the test results. By capturing these temporal metrics, the platform provided a more nuanced understanding of motor and cognitive impairments associated with Parkinson's disease.

Each method provided computational insights into disease progression and contributed to determining PD severity based on digital data and machine learning models. Thus, the methods described in this thesis contribute to the field of computer science through the design and development of digital platforms and explainable machine learning workflows that can be applied to complex and limited datasets, such as those related to neurodegenerative diseases.

V. Discussion

The integration of DTI, eye tracking, and online cognitive testing with machine learning ML algorithms presented the ability to draw conclusions relevant for PD diagnosis and treatment monitoring. Each method demonstrated unique strengths in identifying and predicting PD symptoms, although they are limited by sample sizes, thus requiring further research for validation, including longitudinal studies.

The study on DTI and ML focused on postoperative DBS patients, employing Rough Set Theory to predict symptoms such as tremors and handwriting distortion. The results showed high accuracy in predictions, particularly for left-hand tremors (0.967 accuracy) and handwriting distortion (0.878 accuracy). This approach highlighted the ability of DTI to reveal brain abnormalities linked to PD and provided insights into the microstructural changes associated with the disease.

Furthermore, eye-tracking studies focused on rapid eye movements (saccades) to differentiate PD patients. By comparing various ML models and rough sets, the studies aimed to predict disease progression and treatment outcomes. RS rules predicted UPDRS values with good accuracy, while ML models showed similar accuracy ranges and were able to predict specific session results in certain cases. A subsequent study achieved a 75% accuracy rate in predicting subclasses of UPDRS using eye-tracking data. Again, despite the promising results, the sample size again calls for further validation for the broader applicability of these findings. Nonetheless, these studies underscore the potential of eye-tracking as a non-invasive diagnostic tool for PD.

Web-based cognitive testing emerged as a powerful tool for detecting motor and cognitive signs of PD. Preliminary findings showed a high accuracy rate (91.1%) in differentiating PD patients from healthy controls using Logistic Regression. Further studies combined cognitive test scores with temporal metrics, significantly improving diagnostic accuracy and specificity. The inclusion

of time-based metrics in cognitive screening algorithms highlights the potential for online platforms to capture subtle signs of PD, enabling early detection and continuous monitoring.

Additional direction explored in these studies was the use of virtual reality for the implementation of TMT B. While VR environments present challenges such as cybersickness, web-based approaches offer greater accessibility and do not require specialized hardware. Furthermore, the integration of temporal dynamics with traditional cognitive tests via web platforms can enhance diagnostic accuracy and support disease management.

In conclusion, this discussion offers a brief overview of the key points covered in articles included in this thesis. It serves as a summary of the more detailed discussions presented in each of publications, and I encourage the reader to refer to those original works for a comprehensive analysis.

VI. Conclusions

This work presents the feasibility and validity of digital methods and machine learning for the detection of Parkinson's disease. While many challenges remain, the results indicate that machine learning workflows can be effectively applied to complex and limited disease datasets. This has the potential to advance the development of improved digital diagnostic approaches.

In conclusion, selected digital biomarkers combined with machine learning can be applied to detect PD symptoms in an objective and reliable manner. Further research should focus on evaluating these methods with independent samples and exploring their longitudinal application.

Acknowledgements

I would like to express my gratitude to the Polish-Japanese Academy of Information Technology for the opportunity to conduct research under its affiliation.

I am very thankful to prof. Andrzej Przybyszewski for his supervision and suggestions provided over many years, as well as his central role in the development of this methodology.

I would like to thank my co-authors and colleagues in the research group, whose expertise has greatly influenced my understanding of this field of research.

I am very grateful to my wife for her constant support throughout this study. Her words have been a great source of strength and inspiration.

Glossary

Phrase	Definition
3DSlicer	An open-source software platform for visualization, processing, segmentation, registration, and analysis of medical and biomedical images.
Accuracy	A metric used to evaluate the performance of a classification model.
AD	Alzheimer's disease.
Amantadine	A medication originally used as an antiviral drug that improves motor features of PD.
Antisaccades	Eye movements in the opposite direction of a suddenly appearing target.
AUC	Area under the ROC curve, a measure of the overall performance of a binary classification model.
Binary Classification	A classification task that involves differentiating between two categories.
BMT	Best Medical Treatment, a group of patients receiving optimal medical therapy.
Bradykinesia	Slowness in initiating or performing movements, which is a common symptom of Parkinson's disease.
cMoCA	Computerized Montreal Cognitive Assessment.
CNN	Convolutional neural networks, a class of deep neural networks commonly used in analyzing visual data.
COMT inhibitors	Medications that inhibit the breakdown of dopamine.
Cross-validation	A technique for assessing how a model will generalize to an independent dataset.
cTMT	Computerized Trail Making Test.
DBS	Deep brain stimulation, a surgical treatment involving implanting electrodes in specific brain areas to modulate neural activity.
Decision Trees	A hierarchical structure used for classification and regression tasks.

Digital Biomarkers	Data collected from interactions with technology to understand human behavior or cognitive function.
Disease Stratification	The process of categorizing patients into different stages or severity of a disease.
Dopamine	A neurotransmitter essential for smooth, coordinated muscle activity and various brain functions.
Dopamine agonists	Medications that mimic dopamine's effects.
DTI	Diffusion tensor imaging, a type of MRI used to measure neural structural connectivity.
Epworth	A scale used to measure daytime sleepiness.
F1 score	The harmonic means of precision and recall.
Fixations	Moments when the eyes stop scanning and focus on a single point.
GaussianNB	Gaussian Naive Bayes, a probabilistic classifier based on Bayes' theorem.
GradBoost	Gradient Boosting, a machine learning technique used for regression and classification tasks.
Granular Computing	A computational approach used to analyze patterns in data for disease prediction.
Granule	A collection of elements grouped together based on similarity or functionality.
Hoehn and Yahr Scale	A scale used to describe the progression of Parkinson's disease in stages, from mild symptoms to severe disability.
Hypomimia	Reduction of emotional facial expression.
IR sensor	Infrared sensor used for precise measurement of eye movements.
IRT	Instrumental reaction time.
JAZZ Novo	A head-mounted device with an infrared sensor used for capturing eye movements.
Levodopa	A medication used to replenish depleted dopamine levels in the brain.

Logistic Regression	A classification algorithm used to predict categorical outcomes based on input variables.
Machine Learning	A technique which focuses on learning patterns from data, even when those patterns are too subtle to be identified by traditional statistical techniques.
MAO-B inhibitors	Medications that inhibit the breakdown of dopamine.
MCI	Mild Cognitive Impairment.
MedON	Medication ON, a state where PD patients are evaluated while on medication.
MedOFF	Medication OFF, a state where patients are evaluated while they are off their medication.
MMSE	Mini-Mental State Examination, a tool used for cognitive assessment.
MoCA	Montreal Cognitive Assessment, a cognitive test more effective than some others for assessing cognitive function in PD.
MSA	Multiple system atrophy.
NDs	Neurodegenerative diseases.
PD	Parkinson's disease.
PDD	Dementia in Parkinson's Disease.
PDQ39	Parkinson's Disease Questionnaire-39, a tool used to measure the quality of life in PD patients.
Pearson's	A measure of linear correlation between two sets of data.
POP	Postoperative, a group of patients who have undergone surgery.
Precision	A metric that measures the accuracy of the positive predictions made by the model.
PSP	Progressive supranuclear palsy.
Random Forest (RF)	An ensemble learning method that uses multiple decision trees for classification and regression tasks.

Recall	A metric that measures the ability of the model to identify all relevant instances in the dataset.
Regularization	Techniques used to prevent overfitting in machine learning models by penalizing large coefficients.
ROC curve	A graphical representation of the diagnostic ability of a binary classifier.
Rough Set Theory (RST)	Rough Set Theory, a mathematical approach to deal with vagueness and uncertainty in data.
RSES	Rough Set Exploration System, software used for data mining and analysis using Rough Set Theory.
Saccades	Rapid eye movements between targets.
Smooth pursuit	Continuous movement of the eyes as they follow a moving object.
SN	Substantia Nigra, a deeply pigmented gray matter situated in the midbrain.
SNc	Substantia Nigra pars compacta, a specific region within the Substantia Nigra.
SPSS	Statistical software used for data management and analysis.
STN	Subthalamic nucleus, a brain region involved in movement control.
SVC	Support Vector Classifier, a type of support vector machine used for classification tasks.
TMT	Trail Making Test, a neuropsychological test that measures various cognitive functions.
train_test_split	A technique for dividing a dataset into training and testing subsets.
TTS	Time-to-submit, a measure used in cognitive tests.
Unity engine	A cross-platform game engine used to develop simulations and interactive experiences.
UPDRS	Unified Parkinson Disease Rating Scale, a comprehensive tool used to measure the severity and progression of Parkinson's disease.
VR	Virtual Reality, simulated experience that employs 3D near-eye displays.

References

Agnello L, Ciaccio M (2022) Neurodegenerative Diseases: From Molecular Basis to Therapy. Int J Mol Sci 23:12854. https://doi.org/10.3390/ijms232112854

Alfalahi H, Dias SB, Khandoker AH, et al (2023) A scoping review of neurodegenerative manifestations in explainable digital phenotyping. NPJ Parkinsons Dis 9:49. https://doi.org/10.1038/s41531-023-00494-0

Armstrong MJ, Okun MS (2020) Diagnosis and Treatment of Parkinson Disease. JAMA 323:548. https://doi.org/10.1001/jama.2019.22360

Baschi R, Luca A, Nicoletti A, et al (2020) Changes in Motor, Cognitive, and Behavioral Symptoms in Parkinson's Disease and Mild Cognitive Impairment During the COVID-19 Lockdown. Front Psychiatry 11:. https://doi.org/10.3389/fpsyt.2020.590134

Biswas SK, Nath Boruah A, Saha R, et al (2023) Early detection of Parkinson disease using stacking ensemble method. Comput Methods Biomech Biomed Engin 26:527–539. https://doi.org/10.1080/10255842.2022.2072683

Brem A-K, Kuruppu S, de Boer C, et al (2023) Digital endpoints in clinical trials of Alzheimer's disease and other neurodegenerative diseases: challenges and opportunities. Front Neurol 14:. https://doi.org/10.3389/fneur.2023.1210974

Butson CR, Cooper SE, Henderson JM, McIntyre CC (2007) Patient-specific analysis of the volume of tissue activated during deep brain stimulation. Neuroimage 34:661–670. https://doi.org/10.1016/j.neuroimage.2006.09.034

Chougar L, Arsovic E, Gaurav R, et al (2022) Regional Selectivity of Neuromelanin Changes in the Substantia Nigra in Atypical Parkinsonism. Movement Disorders 37:1245–1255. https://doi.org/10.1002/mds.28988

Crosby NJ, Deane K, Clarke CE (2003) Amantadine in Parkinson's disease. Cochrane Database of Systematic Reviews 2010:. https://doi.org/10.1002/14651858.CD003468

Farrow SL, Cooper AA, O'Sullivan JM (2022) Redefining the hypotheses driving Parkinson's diseases research. NPJ Parkinsons Dis 8:45. https://doi.org/10.1038/s41531-022-00307-w

Fröhlich H, Bontridder N, Petrovska-Delacréta D, et al (2022) Leveraging the Potential of Digital Technology for Better Individualized Treatment of Parkinson's Disease. Front Neurol 13:. https://doi.org/10.3389/fneur.2022.788427

Garcia Santa Cruz B, Husch A, Hertel F (2023) Machine learning models for diagnosis and prognosis of Parkinson's disease using brain imaging: general overview, main challenges, and future directions. Front Aging Neurosci 15:. https://doi.org/10.3389/fnagi.2023.1216163

Goetz CG, Tilley BC, Shaftman SR, et al (2008) Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. Movement Disorders 23:2129–2170. https://doi.org/10.1002/mds.22340

Goldman JG, Vernaleo BA, Camicioli R, et al (2018) Cognitive impairment in Parkinson's disease: a report from a multidisciplinary symposium on unmet needs and future directions to maintain cognitive health. NPJ Parkinsons Dis 4:19. https://doi.org/10.1038/s41531-018-0055-3

Harris CR, Millman KJ, van der Walt SJ, et al (2020) Array programming with NumPy. Nature 585:357–362. https://doi.org/10.1038/s41586-020-2649-2

Hoehn MM, Yahr MD (1967) Parkinsonism. Neurology 17:427–427. https://doi.org/10.1212/WNL.17.5.427

Hunfalvay M, Bolte T, Singh A, et al (2024) Age-Based Developmental Biomarkers in Eye Movements: A Retrospective Analysis Using Machine Learning. Brain Sci 14:686. https://doi.org/10.3390/brainsci14070686

Hunter JD (2007) Matplotlib: A 2D Graphics Environment. Comput Sci Eng 9:90–95. https://doi.org/10.1109/MCSE.2007.55

Iturria-Medina Y, Evans AC (2015) On the central role of brain connectivity in neurodegenerative disease progression. Front Aging Neurosci 7:. https://doi.org/10.3389/fnagi.2015.00090

Janssen Daalen JM, van den Bergh R, Prins EM, et al (2024) Digital biomarkers for non-motor symptoms in Parkinson's disease: the state of the art. NPJ Digit Med 7:186. https://doi.org/10.1038/s41746-024-01144-2

Kikinis R, Pieper SD, Vosburgh KG (2014) 3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support. In: Intraoperative Imaging and Image-Guided Therapy. Springer New York, New York, NY, pp 277–289

Klarendic M, Kaski D (2021) Deep brain stimulation and eye movements. European Journal of Neuroscience 53:2344–2361. https://doi.org/10.1111/ejn.14898

Langdon AJ, Sharpe MJ, Schoenbaum G, Niv Y (2018) Model-based predictions for dopamine. Curr Opin Neurobiol 49:1–7. https://doi.org/10.1016/j.conb.2017.10.006

Latendorf A, Runde LM, Salminen T, Steinert A (2021) Digitization of neuropsychological diagnostics: a pilot study to compare three paper-based and digitized cognitive assessments. Aging Clin Exp Res 33:1585–1597. https://doi.org/10.1007/s40520-020-01668-z

Latif S, Jahangeer M, Maknoon Razia D, et al (2021) Dopamine in Parkinson's disease. Clinica Chimica Acta 522:114–126. https://doi.org/10.1016/j.cca.2021.08.009

Liao X, Yao J, Tang H, et al (2024) Deciphering Parkinson's Disease through Eye Movements: A Promising Tool for Early Diagnosis in the Face of Cognitive Impairment. Int J Clin Pract 2024:. https://doi.org/10.1155/2024/5579238

Lima AA, Mridha MF, Das SC, et al (2022) A Comprehensive Survey on the Detection, Classification, and Challenges of Neurological Disorders. Biology (Basel) 11:469. https://doi.org/10.3390/biology11030469

Linari I, Juantorena GE, Ibáñez A, et al (2022) Unveiling Trail Making Test: visual and manual trajectories indexing multiple executive processes. Sci Rep 12:14265. https://doi.org/10.1038/s41598-022-16431-9

Liss JL, Seleri Assunção S, Cummings J, et al (2021) Practical recommendations for timely, accurate diagnosis of symptomatic Alzheimer's disease (MCI and dementia) in primary care: a review and synthesis. J Intern Med 290:310–334. https://doi.org/10.1111/joim.13244

Lugtenborg LA, Pel JJM, Pool DM (2022) Identifying Behavioural Changes due to Parkinson's Disease Progression in Motor Performance Data. IFAC-PapersOnLine 55:102–107. https://doi.org/10.1016/j.ifacol.2022.10.239

Mahesh B (2020) Machine Learning Algorithms - A Review. International Journal of Science and Research (IJSR) 9:381–386. https://doi.org/10.21275/ART20203995

McKinney W (2010) Data Structures for Statistical Computing in Python. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Mei J, Desrosiers C, Frasnelli J (2021) Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature. Front Aging Neurosci 13:. https://doi.org/10.3389/fnagi.2021.633752

Mitchell T, Lehéricy S, Chiu SY, et al (2021) Emerging Neuroimaging Biomarkers Across Disease Stage in Parkinson Disease. JAMA Neurol 78:1262. https://doi.org/10.1001/jamaneurol.2021.1312

Mollick JA, Hazy TE, Krueger KA, et al (2020) A systems-neuroscience model of phasic dopamine. Psychol Rev 127:972–1021. https://doi.org/10.1037/rev0000199

Ouhmida A, Raihani A, Cherradi B, Lamalem Y (2022) Parkinson's disease classification using machine learning algorithms: performance analysis and comparison. In: 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET). IEEE, pp 1–6

Pawlak Z (1982) Rough sets. International Journal of Computer & Information Sciences 11:341–356. https://doi.org/10.1007/BF01001956

Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12:108–122

Prince M, Wimo A, Guerchet M, et al (2015) World Alzheimer Report 2015.The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends.

Przybyszewski AW, Nowacki JP, Drabik A, et al (2021) Concept of Parkinson Leading to Understanding Mechanisms of the Disease. 456–466. https://doi.org/10.1007/978-3-030-88081-1_34

Rascol O, Fabbri M, Poewe W (2021) Amantadine in the treatment of Parkinson's disease and other movement disorders. Lancet Neurol 20:1048–1056. https://doi.org/10.1016/S1474-4422(21)00249-0

Reneman L, van der Pluijm M, Schrantee A, van de Giessen E (2021) Imaging of the dopamine system with focus on pharmacological MRI and neuromelanin imaging. Eur J Radiol 140:109752. https://doi.org/10.1016/j.ejrad.2021.109752

Rosenblum S, Meyer S, Gemerman N, et al (2020) The Montreal Cognitive Assessment: Is It Suitable for Identifying Mild Cognitive Impairment in Parkinson's Disease? Mov Disord Clin Pract 7:648–655. https://doi.org/10.1002/mdc3.12969

Sequeira L, Benfeito S, Fernandes C, et al (2024) Drug Development for Alzheimer's and Parkinson's Disease: Where Do We Go Now? Pharmaceutics 16:708. https://doi.org/10.3390/pharmaceutics16060708

Sharanyaa S, Renjith PN, Ramesh K (2020) Classification of Parkinson's Disease using Speech Attributes with Parametric and Nonparametric Machine Learning Techniques. In: 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). IEEE, pp 437–442

Sheng X, Chen H, Shao P, et al (2021) Brain Structural Network Compensation Is Associated With Cognitive Impairment and Alzheimer's Disease Pathology. Front Neurosci 15:. https://doi.org/10.3389/fnins.2021.630278

Śledzianowski A (2024) The application of data mining and machine learning techniques to study eye movement properties in the context of neurodegenerative diseases and changes in the emotional system (PhD Thesis). Polish-Japanese Academy of Information Technology

Sledzianowski A, Szymanski A, Drabik A, et al (2019) Measurements of Antisaccades Parameters Can Improve the Prediction of Parkinson's Disease Progression. 602–614. https://doi.org/10.1007/978-3-030-14802-7_52

Smith T, Gildeh N, Holmes C (2007) The Montreal Cognitive Assessment: Validity and Utility in a Memory Clinic Setting. The Canadian Journal of Psychiatry 52:329–332. https://doi.org/10.1177/070674370705200508

Snowdon A, Hussein A, Kent R, et al (2015) Comparison of an Electronic and Paper-based Montreal Cognitive Assessment Tool. Alzheimer Dis Assoc Disord 29:325–329. https://doi.org/10.1097/WAD.0000000000000069

Step K, Ndong Sima CAA, Mata I, Bardien S (2024) Exploring the role of underrepresented populations in polygenic risk scores for neurodegenerative disease risk prediction. Front Neurosci 18:. https://doi.org/10.3389/fnins.2024.1380860

Szymański A (2019) Application of Data Mining Methods For Symptom Evaluation in Parkinson's Disease (PhD Thesis). Polish-Japanese Academy of Information Technology

Tolosa E, Garrido A, Scholz SW, Poewe W (2021) Challenges in the diagnosis of Parkinson's disease. Lancet Neurol 20:385–397. https://doi.org/10.1016/S1474-4422(21)00030-2

Turing AM (1937) On computable numbers, with an application to the entscheidungsproblem. Proceedings of the London Mathematical Society s2-42:. https://doi.org/10.1112/plms/s2-42.1.230

Turing AM (1990) The chemical basis of morphogenesis. Bull Math Biol 52:. https://doi.org/10.1007/BF02459572

Waskom M (2021) seaborn: statistical data visualization. J Open Source Softw 6:3021. https://doi.org/10.21105/joss.03021

Watts J, Khojandi A, Shylo O, Ramdhani RA (2020) Machine Learning's Application in Deep Brain Stimulation for Parkinson's Disease: A Review. Brain Sci 10:809. https://doi.org/10.3390/brainsci10110809

Yilmaz R, Hopfner F, van Eimeren T, Berg D (2019) Biomarkers of Parkinson's disease: 20 years later. J Neural Transm 126:803–813. https://doi.org/10.1007/s00702-019-02001-3

Zadikoff C, Fox SH, Tang-Wai DF, et al (2008) A comparison of the mini mental state exam to the montreal cognitive assessment in identifying cognitive deficits in Parkinson's disease. Movement Disorders 23:297–299. https://doi.org/10.1002/mds.21837

DBND - Digital Biomarkers in Neurodegenerative Diseases. https://nd.pja.edu.pl/. Accessed 13 Jul 2024

Abstracts

DTI Helps to Predict Parkinson's Patient's Symptoms Using Data Mining Techniques

A. **Chudzik**, A. Szymański, J. P. Nowacki, and A. W. Przybyszewski

1. **Type:** Conference paper presented in Yogyakarta, Indonesia, during the Asian Conference on Intelligent Information and Database Systems - Special Session on Analysis of Image, Video, Movements and Brain Intelligence in Life Sciences (Springer, 2019)
2. **Score:** 70 points (MEiN list of January 5, 2024).
3. **Abstract:** Deep Brain Stimulation (DBS) is commonly used to treat, inter alia, movement disorder symptoms in patients with Parkinson's disease, dystonia, or essential tremor. The procedure stimulates a targeted region of the brain through implanted leads that are powered by a device called an implantable pulse generator (IPG). The mentioned targeted region is mainly chosen to be subthalamic nucleus (STN) during most of the operations. STN is a nucleus in the midbrain with a size of 3 mm × 5 mm × 9 mm that consists of parts with different physiological functions. The purpose of the study was to predict Parkinson's patient's symptoms defined by Unified Parkinson's Disease Rating Scale (UPDRS) that may occur after the DBS treatment. Parameters had been obtained from 3DSlicer (Harvard Medical School, Boston, MA), which allowed us to track connections between the stimulated part of STN and the cortex based on the DTI (diffusion tensor imaging).
4. **Contribution**: I conducted the analysis of RST model results and prepared the description of the methods and findings.
5. **Reference:** A. Chudzik, A. Szymański, J. P. Nowacki, and A. W. Przybyszewski, "DTI Helps to Predict Parkinson's Patient's Symptoms Using Data Mining Techniques," in Intelligent Information and Database Systems, F. L. and H. T.-P. and T. B. Nguyen Ngoc Thanh and Gaol, Ed., Cham: Springer International Publishing, 2019, pp. 615–623. doi: 10.1007/978-3-030-14802-7_53.

Machine Learning and Eye Movements Give Insights into Neurodegenerative Disease Mechanisms

A. W. Przybyszewski, A. Śledzianowski, A. **Chudzik**, S. Szlufik, and D. Koziorowski

1. **Type:** Article published in Sensors (MDPI, 2023).
2. **Score:** 100 points (MEiN list of January 5, 2024).
3. **Abstract:** Humans are a vision-dominated species; what we perceive depends on where we look. Therefore, eye movements (EMs) are essential to our interactions with the environment, and experimental findings show EMs are affected in neurodegenerative disorders (ND). This could be a reason for some cognitive and movement disorders in ND. Therefore, we aim to establish whether changes in EM-evoked responses can tell us about the progression of ND, such as Alzheimer's (AD) and Parkinson's diseases (PD), in different stages. In the present review, we have analyzed the results of psychological, neurological, and EM (saccades, antisaccades, pursuit) tests to predict disease progression with machine learning (ML) methods. Thanks to ML algorithms, from the high-dimensional parameter space, we were able to find significant EM changes related to ND symptoms that gave us insights into ND mechanisms. The predictive algorithms described use various approaches, including granular computing, Naive Bayes, Decision Trees/Tables, Logistic Regression, C-/Linear SVC, KNC, and Random Forest. We demonstrated that EM is a robust biomarker for assessing symptom progression in PD and AD. There are navigation problems in 3D space in both diseases. Consequently, we investigated EM experiments in virtual space and how they may help find neurodegeneration-related brain changes, e.g., related to place or/and orientation problems. In conclusion, EM parameters with clinical symptoms are powerful precision instruments that, in addition to their potential for predictions of ND progression with the help of ML, could be used to indicate the different preclinical stages of both diseases.
4. **Contribution**: I wrote parts of the article, including the challenges of virtual environments in the Further Research section.
5. **Reference:** A. W. Przybyszewski, A. Śledzianowski, A. Chudzik, S. Szlufik, and D. Koziorowski, "Machine Learning and Eye Movements Give Insights into Neurodegenerative Disease Mechanisms," Sensors, vol. 23, no. 4, p. 2145, Feb. 2023, doi: 10.3390/s23042145.

Comparison of Different Data Mining Methods to Determine Disease Progression in Dissimilar Groups of Parkinson's Patients

A. W. Przybyszewski, A. **Chudzik**, S. Szlufik, P. Habela, and D. M. Koziorowski

1. **Type:** Article published in Fundamenta Informaticae (IOS Press, 2020).
2. **Score:** 70 points (MEiN list of January 5, 2024).
3. **Abstract:** Parkinson's disease (PD) is the second after Alzheimer's most popular neurodegenerative disease (ND). Cures for both NDs are currently unavailable. OBJECTIVE: The purpose of our study was to predict the results of different PD patients' treatments to find an optimal one. METHODS: We have compared rough sets (RS) and others, in short, machine learning (ML) models to describe and predict disease progression expressed as UPDRS values (Unified Parkinson's Disease Rating Scale) in three groups of Parkinson's patients: 23 BMT (Best Medical Treatment) patients on medication; 24 DBS patients on medication and on DBS therapy (Deep Brain Stimulation) after surgery performed during our study; and 15 POP (Postoperative) patients who had had surgery earlier (before the beginning of our research). Every PD patient had three visits approximately every six months. The first visit for DBS patients was before surgery. Based on the following condition attributes: disease duration, saccadic eye movement parameters, and neuropsychological tests: PDQ39 (Parkinson's Disease Questionnaire - disease-specific health-related quality-of-life questionnaire), and Epworth Sleepiness Scale tests we have estimated UPDRS changes (as the decision attribute). RESULTS: By means of RS rules obtained for the first visit of BMT/DBS/POP patients, we have predicted UPDRS values in the following year (two visits) with global accuracy of 70% for both BMT visits; 56% for DBS, and 67%, 79% for POP second and third visits. The accuracy obtained by ML models was generally in the same range, but it was calculated separately for different sessions (MedOFF/MedON). We have used RS rules obtained in BMT patients to predict UPDRS of DBS patients; for the first session DBSW1: global accuracy was 64%, for the second DBSW2: 85% and the third DBSW3: 74% but only for DBS patients during stimulation-ON. ML models gave better accuracy for DBSW1/W2 session S1(MedOFF): 88%, but inferior results for session S3 (MedON): 58% and 54%. Both RS and ML could not predict UPDRS in DBS patients during stimulation-OFF visits because of differences in UPDRS. By using RS rules from BMT or DBS patients we could not predict UPDRS of POP group, but with certain limitations (only for MedON), we derived such predictions for the POP group from results of DBS patients by using ML models (60%). SIGNIFICANCE:

Thanks to our RS and ML methods, we were able to predict Parkinson's disease (PD) progression in dissimilar groups of patients with different treatments. It might lead, in the future, to the discovery of universal rules of PD progression and optimize the treatment.

4. **Contribution**: I developed the pipeline for evaluating machine learning models with consistent tasks and optimal hyperparameter tuning, and I prepared the corresponding sections of the paper describing these models.
5. **Reference**: A. W. Przybyszewski, A. Chudzik, S. Szlufik, P. Habela, and D. M. Koziorowski, "Comparison of Different Data Mining Methods to Determine Disease Progression in Dissimilar Groups of Parkinson's Patients," Fundam Inform, vol. 176, no. 2, pp. 167–181, Dec. 2020, doi: 10.3233/FI-2020-1969.

Eye-Tracking and Machine Learning Significance in Parkinson's Disease Symptoms Prediction

A. **Chudzik**, A. Szymański, J. P. Nowacki, and A. W. Przybyszewski

1. **Type:** Article presented online during the Asian Conference on Intelligent Information and Database Systems (ACIIDS) - Special Session on Interactive Analysis of Image, Video and Motion Data in Life Sciences (Springer, 2020).
2. **Score:** 70 points (MEiN list of January 5, 2024).
3. **Abstract:** Parkinson's disease (PD) is a progressive, neurodegenerative disorder characterized by resting tremor, rigidity, bradykinesia, and postural instability. The standard measure of the PD progression is Unified Parkinson's Disease Rating (UPDRS). Our goal was to predict patients' UPDRS development based on the various groups of patients in the different stages of the disease. We used standard neurological and neuropsychological tests, aligned with eye movements on a dedicated computer system. For predictions, we have applied various machine learning models with different parameters embedded in our dedicated data science framework written in Python and based on the Scikit Learn and Pandas libraries. The models proposed by us reached 75% and 70% accuracy while predicting subclasses of UPDRS for patients in advanced stages of the disease who respond to treatment, with a global 57% accuracy score for all classes. We have demonstrated that it is possible to use eye movements as a biomarker for the assessment of symptom progression in PD.
4. **Contribution**: I selected, hyper-tuned, and observed the behavior of ML models for this task and described the results in the article.
5. **Citation:** A. Chudzik, A. Szymański, J. P. Nowacki, and A. W. Przybyszewski, "Eye-Tracking and Machine Learning Significance in Parkinson's Disease Symptoms Prediction," in Intelligent Information and Database Systems, K. and S. A. and T. B. and C. S. Nguyen Ngoc Thanh and Jearanaitanakij, Ed., Cham: Springer International Publishing, 2020, pp. 537–547. doi: 10.1007/978-3-030-42058-1_45.

Machine Learning and Digital Biomarkers Can Detect Early Stages of Neurodegenerative Diseases

A. **Chudzik**, A. Śledzianowski, and A. W. Przybyszewski

1. **Type**: Article published in Sensors (MDPI, 2024).
2. **Score:** 100 points (MEiN list of January 5, 2024).
3. **Abstract**: Neurodegenerative diseases (NDs) such as Alzheimer's Disease (AD) and Parkinson's Disease (PD) are devastating conditions that can develop without noticeable symptoms, causing irreversible damage to neurons before any signs become clinically evident. NDs are a major cause of disability and mortality worldwide. Currently, there are no cures or treatments to halt their progression. Therefore, the development of early detection methods is urgently needed to delay neuronal loss as soon as possible. De spite advancements in Medtech, the early diagnosis of NDs remains a challenge at the intersection of medical, IT, and regulatory fields. Thus, this review explores "digital biomarkers" (tools designed for remote neurocognitive data collection and AI analysis) as a potential solution. The review summarizes that recent studies combining AI with digital biomarkers suggest the possibility of identifying pre-symptomatic indicators of NDs. For instance, research utilizing convolutional neural networks for eye tra cking has achieved significant diagnostic accuracies. ROC-AUC scores reached up to 0.88, indicating high model performance in differentiating between PD patients and healthy controls. Similarly, advancements in facial expression analysis through tools have demonstrated significant potential in detecting emotional changes in ND patients, with some models reaching an accuracy of 0.89 and a precision of 0.85. This review follows a structured approach to article selection, starting with a comprehensive database search and culminating in a rigorous quality assessment and meaning for NDs of the different methods. The process is visualized in 10 tables with 54 parameters describing different approaches and their consequences for understanding various mechanisms in ND changes. However, these methods also face challenges related to data accuracy and privacy concerns. To address these issues, this review proposes strategies that emphasize the need for rigorous validation and rapid integration into clinical practice. Such integration could transform ND diagnostics, making early detection tools more cost-effective and globally accessible. In conclusion, this review underscores the urgent need to incorporate validated digital health tools into mainstream medical practice. This integration could indicate a new era in the early diagnosis of neurodegenerative diseases,

potentially altering the trajectory of these conditions for millions worldwide. Thus, by highlighting specific and statistically significant findings, this review demonstrates the current progress in this field and the potential impact of these advancements on the global management of NDs.

4. **Contribution**: I prepared the original draft of the review, including parts of the theoretical computational aspects of neurodegenerative diseases.
5. **Reference**: A. Chudzik, A. Śledzianowski, and A. W. Przybyszewski, "Machine Learning and Digital Biomarkers Can Detect Early Stages of Neurodegenerative Diseases," Sensors, vol. 24, no. 5, p. 1572, Feb. 2024, doi: 10.3390/s24051572.

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

A. **Chudzik**, A. Drabik, and A. W. Przybyszewski

1. **Type:** Conference paper presented online in Phuket, Thailand, during the Asian Conference on Intelligent Information and Database Systems - Special Session on AI in Neurology to Brain Computations (Springer, 2023).
2. **Score:** 70 points (MEiN list of January 5, 2024).
3. **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.
4. **Contribution**: I coded the web platform, conducted the analysis, and prepared the original draft of the article.
5. **Reference:** A. Chudzik, A. Drabik, and A. W. Przybyszewski, "Investigating the Impact of Parkinson's Disease on Brain Computations: An Online Study of Healthy Controls and PD Patients," in Intelligent Information and Database Systems: 15th Asian Conference, ACIIDS 2023, Phuket, Thailand, July 24–26, 2023, Proceedings, Part II, Berlin, Heidelberg: Springer-Verlag, 2023, pp. 235–246. doi: 10.1007/978-981-99-5837-5_20.

Classification of Parkinson's Disease Using Machine Learning with MoCA Response Dynamics

A. **Chudzik** and A. W. Przybyszewski

1. **Type**: Article published in Applied Sciences (MDPI, 2024).
2. **Score:** 100 points (MEiN list of January 5, 2024).
3. **Abstract**: Neurodegenerative diseases (NDs), including Parkinson's and Alzheimer's disease, pose a significant challenge to global health, and early detection tools are crucial for effective intervention. The adaptation of online screening forms and machine learning methods can lead to better and wider diagnosis, potentially altering the progression of NDs. Therefore, this study examines the diagnostic efficiency of machine learning models using Montreal Cognitive Assessment test results (MoCA) to classify scores of people with Parkinson's disease (PD) and healthy subjects. For data analysis, we implemented both rule-based modeling using Rough Set Theory (RST) and classic machine learning (ML) techniques such as Logistic Regression, support vector machines, and Random Forest. Importantly, the diagnostic accuracy of the best performing model (RST) increased from 80.0% to 93.4% and diagnostic specificity increased from 57.2% to 93.4% when the MoCA score was combined with temporal metrics such as IRT—instrumental reaction time and TTS—submission time. This highlights that online platforms are able to detect subtle signs of bradykinesia (a hallmark symptom of Parkinson's disease) and use this as a biomarker to provide more precise and specific diagnosis. Despite the constrained number of participants (15 Parkinson's disease patients and 16 healthy controls), the results suggest that incorporating time-based metrics into cognitive screening algorithms may significantly improve their diagnostic capabilities. Therefore, these findings recommend the inclusion of temporal dynamics in MoCA assessments, which may potentially improve the early detection of NDs.
4. **Contribution**: I coded the web platform, conducted the analysis, and prepared the original draft of the article.
5. **Reference**: A. Chudzik and A. W. Przybyszewski, "Classification of Parkinson's Disease Using Machine Learning with MoCA Response Dynamics," Applied Sciences, vol. 14, no. 7, p. 2979, Apr. 2024, doi: 10.3390/app14072979.

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

A. **Chudzik**, J. P. Nowacki, and A. W. Przybyszewski

1. **Type:** Conference paper that will be presented in Kolkata, India, during the International Conference on Pattern Recognition - ICPR (2024).
2. **Score:** 70 points (MEiN list of January 5, 2024).
3. **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, 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.
4. **Contribution**: I coded the web platform, conducted the analysis, and prepared the original draft of the article.
5. **Reference:** A. Chudzik, J. P. Nowacki, and A. W. Przybyszewski, "Recognizing Patterns of Parkinson's Disease using Online Trail Making Test and Response Dynamics – Preliminary Study," ICPR 2024. PREPRINT.

Additional Contributions

Amantadine treatment in Parkinson's disease patients as a modulatory factor of SARS-Cov-2 infection

S. Szlufik, A. **Chudzik**, A. Przybyszewski, and D. Koziorowski

1. **Type:** Abstract published in Parkinsonism & Related Disorders (Elsevier, 2023).
2. **Abstract:** Background: Amantadine has been used for the prevention and treatment
3. of viral influenza A, but more recently it is used mainly in PD patients. Previous studies showed a possible impact of amantadine on COVID-19 severity in patients using this drug due to other (neurological diseases, mainly PD). Therefore the aim of this study was to evaluate the possible impact of amantadine on the SARS-Cov-2 infection in Parkinson's disease (PD) patients. Methods: It was a nation-wide survey performed in Polish PD population from 01.2021 till 01.2022. All members of Polish PD foundations have been asked to answer a survey e 140 PD patients filled the questionnaire consisting of 35 questions concerning the amantadine treatment, Parkinson's disease and SARS-Cov-2 infection history. The patients were divided into 2 groups: group A+ which was treated with amantadine (57 cases) and group A- (83 non-amantadine takers). Results: We have observed more slight symptoms and progression of SARS-Cov-2 infection in PD patients taking amantadine (8 patients COVID-19+) than in PD patients not taking amantadine (12 patients COVID-19+). The symptoms of COVID-19 in A+ group were slight weakness, sweating or none of symptoms whereas group A- mainly demonstrated cough, smell loss, high temperature e one of group A- patients was hospitalized. Conclusions: Amantadine treatment in PD patients can reduce the severity of SARS-Cov-2 infection in PD patients.
4. **Contribution**: I created an online questionnaire and analyzed the results.
5. **Reference:** S. Szlufik, A. Chudzik, A. Przybyszewski, and D. Koziorowski, "Amantadine treatment in Parkinson's disease patients as a modulatory factor of SARS-Cov-2 infection," *Parkinsonism Relat Disord*, vol. 113, Aug. 2023, doi: 10.1016/j.parkreldis.2023.105606.

Structural neuroplasticity induced by training in the form of a first-person shooter video game

N. Kowalczyk, M. Myśliwiec, M. Skorko, P. Dobrowolski, A. **Chudzik**, and A. Brzezicka

1. **Type:** Abstract published in 9th Annual Conference Aspects of Neuroscience Abstract Book (Warsaw University, 2019).
2. **Abstract:** Over the last few years, computer games have evolved from an unrealistic and straightforward two-dimensional environment to experience becoming more similar to a real-life activity. Due to the increasing impact that video games playing has been having on global society, the cognitive effects of video games have become an exciting matter to be considered on a scientific level. While many studies show that playing action video games has a positive impact on a vast range of cognitive skills, numerous studies show negligible or no cognitive effect. There is nevertheless only a handful of studies that focus on possible structural changes as an effect of playing a video game. Those do show that playing action video games induces grey matter thickness changes in structures like the parahippocampal cortex, somatosensory cortex, superior parietal lobule or insula. The first aim of the presented study was to see whether or not approximately 30 hours of training is sufficient for any cognitive and structural changes to occur. The region of interest (ROI) approach was used to select structures of the brain that could be related to the gaming experience. The second aim of the study was to see whether or not the cortical thickness in specific structures can be predictive of the quality of the training process. A strong effect of insula did occur – while its right side negatively changed its thickness as an effect of training, the left side correlated positively with the game achievements.
3. **Contribution**: I wrote a parser for demos (replays of user sessions) that extracted relevant data about the performance of each participant.
4. **Reference:** N. Kowalczyk, M. Myśliwiec, M. Skorko, P. Dobrowolski, A. Chudzik, and A. Brzezicka, "Structural neuroplasticity induced by training in the form of a first-person shooter video game," in *9th Annual Conference Aspects of Neuroscience Abstract Book*, 1st ed., Paź Marta and Karpińska Magdalena, Eds., Warsaw: Warsaw University, 2019, pp. 149–149, ISBN: 978-83-954288-1-4.

How to Cure Alzheimer’s Disease

A. W. Przybyszewski and A. **Chudzik**

1. **Type:** Commentary published in Journal of Alzheimer's Disease (IOS Press, 2024).
2. **Abstract:** There has been a lot of buzz surrounding new drug discoveries that claim to cure Alzheimer's disease (AD). However, it is crucial to keep in mind that the changes in the brain linked to AD start occurring 20-30 years before the first symptoms arise. By the time symptoms become apparent, many areas of the brain have already been affected. That's why experts are focusing on identifying the onset of neurodegeneration processes to prevent or cure AD effectively. Scientists use biomarkers and machine learning methods to analyze AD progressions and estimate them "backward" in time to discover the beginning of the disease.
3. **Contribution**: I wrote a part of this article.
4. **Reference:** A. W. Przybyszewski and A. Chudzik, "How to Cure Alzheimer’s Disease," Journal of Alzheimer’s Disease, pp. 1–3, May 2024, doi: 10.3233/JAD-240231.

Full Texts of Articles Constituting the Thesis

DTI Helps to Predict Parkinson's Patient's Symptoms Using Data Mining Techniques

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

[1] Polish-Japanese Academy of Information Technology, Koszykowa 86 St., 02-008 Warsaw, Poland
{artur.chudzik, artur.szymanski, nowacki, przy}@pjwstk.edu.pl
[2] Department of Neurology, University of Massachusetts Medical School, 65 Lake Avenue, Worcester, MA 01655, USA
andrzej.przybyszewski@umassmed.edu

Abstract. Deep Brain Stimulation (DBS) is commonly used to treat, inter alia, movement disorder symptoms in patients with Parkinson's disease, dystonia or essential tremor. The procedure stimulates a targeted region of the brain through implanted leads that are powered by a device called an implantable pulse generator (IPG). The mentioned targeted region is mainly chosen to be subthalamic nucleus (STN) during most of the operations. STN is a nucleus in the midbrain with a size of 3 mm × 5 mm × 9 mm that consist of parts with different physiological functions. The purpose of the study was to predict Parkinson's patient's symptoms defined by Unified Parkinson's Disease Rating Scale (UPDRS) that may occur after the DBS treatment. Parameters had been obtained from 3DSlicer (Harvard Medical School, Boston, MA), which allowed us to track connections between the stimulated part of STN and the cortex based on the DTI (diffusion tensor imaging).

Keywords: Subthalamic nucleus · UPDRS · RSES · MRI · DTI · DBS · Parkinson's disease · Data mining

1 Introduction

Neurodegenerative diseases, in which we could distinguish Parkinson Disease (PD), have their background in neurodegeneration which could be described as progressive loss of structure or function of neurons, including the death of neurons. PD is primarily related to the substantia nigra degeneration which leads to dopamine insufficiency. Standard medication in PD is L-DOPA, which is a precursor of dopamine. However, disease progression affects in L-DOPA efficiency decay which may be revealed in on-off symptom fluctuation.

Thus, the neurologist has often to extend standard medication therapy to DBS (Deep Brain Stimulation) surgery [1]. DBS treatment depends on stimulation of the subthalamic nucleus (STN) which is dorsal to the substantia nigra and medial to the internal capsule. STN is also being known as a "hyper direct pathway" [2] of motor

N. T. Nguyen et al. (Eds.): ACIIDS 2019, LNAI 11432, pp. 615–623, 2019.
https://doi.org/10.1007/978-3-030-14802-7_53

control, contrasting with the direct and indirect pathways implemented elsewhere in the basal ganglia. However, the procedure of application the DBS electrode under the appropriate placement is challenging and may affect in different recovery time and treatment effectiveness.

The searching of localization of the subthalamic nucleus is done mainly by the registrations of neuronal activity via microelectrode recording (MER) [8]. MER is an intraoperative analysis of multi-unit activity (MUA). The commonly used criteria for electrophysiological localization of the STN are qualitative and mainly based on visual and acoustic observations of changes in spike frequencies and background activity. The characteristics of spike trains change during the whole path of brain structures and differ when the electrode passes through the thalamus, zona incerta, lenticular fasciculus, subthalamic nucleus, and the substantia nigra. Bursts in the background activity and sudden increases in the frequency of neuronal spiking are signs that electrode is near to STN. To obtain additional confirmation of the correct electrode placement, supplementary kinesthetic responses measurements aligned with microstimulation are being proceeded. There are two main strategies in searching for STN. First one depends on a single microelectrode which leads to the necessity of multiple passes for correct localization of the motor region. The second one uses 4–5 microelectrode insertions simultaneously. It has to be noticed that any stimulation or manipulation of the non-motor STN region is usually avoided since it can provoke psychiatric and cognitive dysfunctions [7].

To predict the neurological effects related to different electrode-contact stimulations, we have extracted specific parameters acquired from diffusion tensor imaging (DTI). We have demonstrated that with the data mining methods, supported with the rough set theory, it is possible to predict Parkinson's patient's symptoms, according to Unified Parkinson's Disease Rating Scale (UPDRS) [9].

2 Methods

In this research, the subject of study was data acquired from nine patients with advanced PD, which have had DBS electrodes implanted. The primary step was the analysis of the data acquired from the DTI by 3DSlicer software. Those parameters were: two technical values (fiducial region size which determines the tractography radius for selected electrode contact; stopping value - the value of ceasing for the generation of the given tract) and an amount of tract reaching the proximity of given ROI, distinguished between left and right side for every region.

The process of the tractography generation was described in previous works [3, 4]. The generation was carried separately for each contact, and it was on DTI data from the pre-OP DWI (pre-operational diffusion-weighted imaging). The DWI to DTI (diffusion tensor imaging) data was estimated by the use of least squares function approximation. Then, to generate relevant tracts, a proper ROI (region of interest) has been set for each patient, based on electrode contacts. Next, a module called Tractography Interactive Seeding has been applied in order to generate tracts. For every patient, two sets of data

have been generated- with a minimal and large (over 30) number of tracts into primary and supplementary motor cortices (Fig. 1). The parameters that were used during the creation of individual tractography were fiducial region size and stopping value, mentioned previously.

The analysis included discovering of the correlation between given attributes aligned with the importance. This operation was performed with the usage of *pandas* library, which is a Python tool for data analysis and statistics [6]. Based on the obtained values, it was possible to select the attributes relevant in the data mining process, which was performed in RSES software.

We have used the RSES 2.1.1 (Rough System Exploration Program) with implementation of RS rules to process our data. An information system [5] can be considered as a pair:

$$S = (U, A)$$

Where:

$U = universe\ of\ objects$
$A = set\ of\ attributes$
$V = set\ of\ values$
$a(u) = unique\ element\ of\ V$
$a \in A$
$u \in U$
A decision table for S is the 3-tuple:

$$S = (U, C, D)$$

Where:

$C = condition\ attributes$
$D = decision\ attributes$

Information table contains rows, where each denotes a particular rule that connects condition and decision attributes for a single measurement of a specific patient.

For results evaluation, we have used a technique called cross-validation, which is a suitable method for estimating the performance of a predictive model, selection of features or parameters adjustment. It is based on the approach of partitioning a data set X into n subsets X_i. Then, given algorithm is performed n times, each time using a different training set $X - X_i$ and validating the results on X_i.

The classifier could be considered to be relevant because of attributes selection that had been done by an algorithm itself. For example, when the left hand tremor is taken as a prediction value, the RSES assumed that relevant attributes are related to, inter alia, hand tracts.

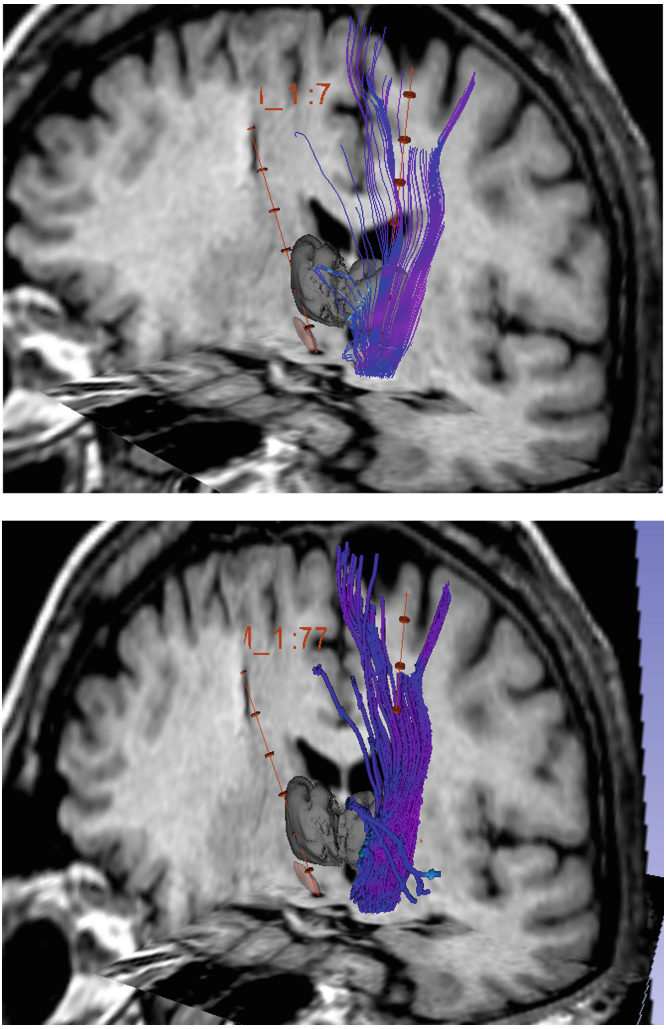

Fig. 1. Screenshots taken from Slicer 3D project of sample patient. We can observe trajectories of both implanted electrodes marked by orange lines, which are aligned with dead tissue visible on MRI slices as a result of surgery. Position of the electrodes is visualized with relevant neighboring structures: STN, Globus Pallidus and Thalamus. DTI tractography is generated from left STN showing connections going to SM and M1 areas of cortex. (Color figure online)

3 Results

The first step was to create a decision table that consists of data attributes obtained from 3DSlicer, based on diffusion tensor imaging, as described previously in the section on Methods. Then, RSES methods were applied, to get decision rules. To achieve that, rows and columns of Table 1 must have been exchanged so that parameters of different patients were in rows, and their results (attributes) were in columns.

Table 1. A fragment of the input dataset. UPDRS <code> - UPDRS value for specific motoric classification of patient' condition; Slicer L/R fiducial region size – Slicer tractography radius for selected electrode contact (in millimeters); Slicer L/R stopping value – Slicer parameter for ceasing generation of the given tract; Slicer L/R tracts lip/hand/foot – number of tract reaching proximity of lip/hand/foot ROI.

Patient ID	10	10	20
UPDRS 21 L Hand action or postural tremor	0	1	2
Slicer R fiducial region size	5	5	5
Slicer R stopping value	0.21	0.21	0.21
Slicer R tracts hand	2	3	3
Slicer R tracts lip	4	15	2
Slicer R tracts foot	15	2	2

In all experiments, we had used the Unified Parkinson Disease Rating Scale (UPDRS) which gave us information about the disease progression in various parts of the body in the context of disease-dependent factor (tremor).

The dataset has been spited into three, smaller subsets, where each was targeted to a different UPDRS marker ("left-hand tremor", "face, lips, chin tremor", "handwriting distortion"). On every subset, we had conducted an experiment based on defined attributes that led to the conclusion of possible UPDRS value after application of DBS treatment on the patient.

3.1 Prediction of the Left-Hand Tremor

Attributes that were relevant during the discretization process were strictly related to Slicer data acquired from the right side of the area, such as fiducial region size, stopping value, tracts for hand, lip and foot (Fig. 2).

The results of the prediction of the left-hand tremor based on given DTI parameters revealed the accuracy of 0.967 with the coverage of 0.425 (Table 2).

3.2 Prediction of the Facial Area Tremor

For this task, chosen attributes were related both to the left and right part of the area. From the left side: fiducial region size, stopping value and lip tracts. From the right side: stopping values, and tracts for lips and foot (Fig. 3).

Prediction result of the "face, chin, lips tremor" was with the accuracy as high as 0.824 with the coverage of 0.7 (Table 3).

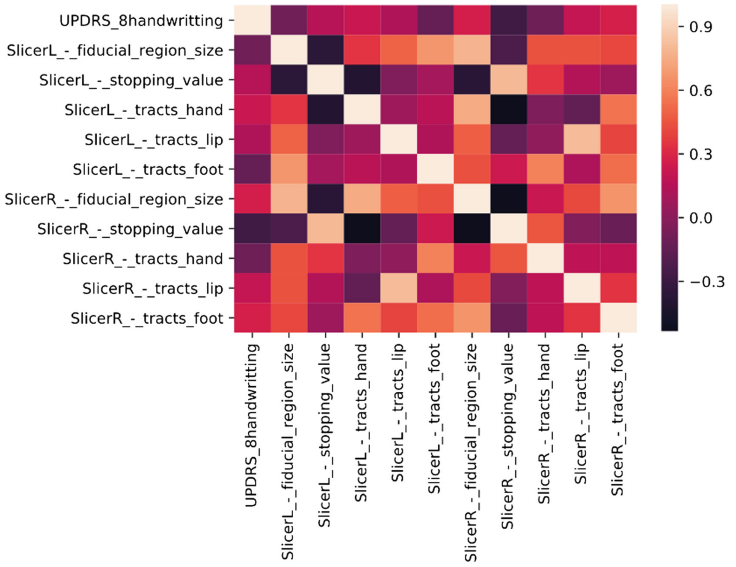

Fig. 2. Computed correlation between Slicer parameters and UPDRS examination: action or postural tremor of Hands – Left hand. The similarity between the two parameters is correlated with the color of the cell in the matrix (lighter shade represents higher similarity and contrariwise). (Color figure online)

Table 2. Confusion matrix of UPDRS #21: action or postural tremor of hands - left hand by rules obtained from 3DSlicer values based on DTI data. Number of tested subjects: 8. Accuracy (total): 0.967. Coverage (total): 0.425. TPR stands for "true positive rate".

		Predicted					
		0	**1**	**2**	**No. of obj.**	**Accuracy**	**Coverage**
Actual	**0**	2.6	0.0	0.0	5.2	**0.8**	0.48
	1	0.2	0.6	0.0	2.0	**0.5**	0.30
	2	0.0	0.0	0.0	0.8	**0.0**	0.00
	TPR	**0.76**	**0.6**	**0.0**			

3.3 Prediction of the Handwriting Disturbances

The last experiment with the detection of UPDRS value change of the "handwriting" test was with the accuracy of 0.878 by the coverage 0.667 (Fig. 4).

The relevant attributes were: DBS and BMT. Furthermore, mainly parameters from the left side Slicer were observed as relevant, such as fiducial region size, stopping value, hand and lip tracts. From the right side, only one technical value (stopping) was taken under consideration (Table 4).

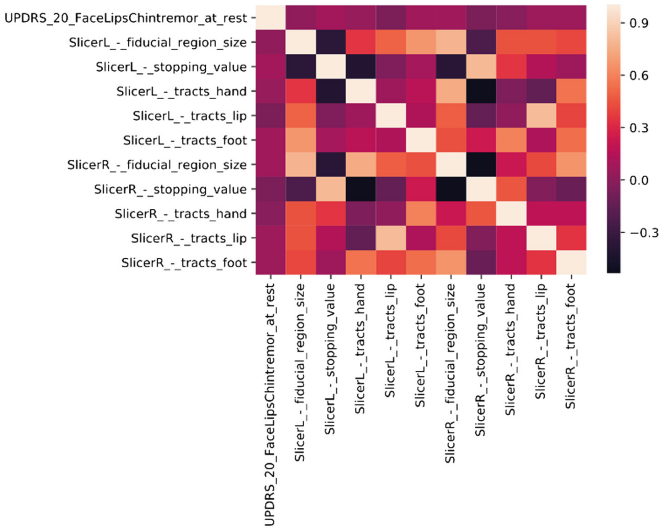

Fig. 3. Computed correlation between Slicer parameters and UPDRS examination: tremor at rest, face, lips, and chin. The similarity between the two parameters is correlated with the color of the cell in the matrix (lighter shade represents higher similarity and contrariwise). (Color figure online)

Table 3. Confusion matrix of UPDRS #20: tremor at rest, face, lips, chin by rules obtained from 3DSlicer values based on DTI data. Total number of tested subjects: 10. Accuracy (total): 0.824. Coverage (total): 0.7.

		Predicted						
		0	**1**	**2**	**3**	**No. of obj.**	**Accuracy**	**Coverage**
Actual	**0**	4.75	0.50	0.00	0.25	6.5	**0.867**	0.866
	1	0.25	0.75	0.00	0.00	1.5	**0.750**	0.833
	2	0.00	0.00	0.00	0.00	0.5	**0.000**	0.000
	3	0.25	0.00	0.00	0.25	1.5	**0.125**	0.250
	TPR	**0.93**	**0.50**	**0.00**	**0.25**			

4 Discussion

Deep brain stimulation is currently widely applied as a surgical choice of treatment for patients with advanced PD. The benefits of STN stimulation are due to combined mechanisms and involve several adjacent structures. To improve the success of the procedure, more selectivity is needed and both topographical level and stimulation parameters must be enhanced [1].

This article represents the continuation of previous findings presented in [7] that are useful to the surgeon as a tool for confirmation that the subthalamic nucleus is near to the microelectrode path. Furthermore, it extends them even more with data mining techniques to predict the neurological effects related to different electrode-contact stimulations.

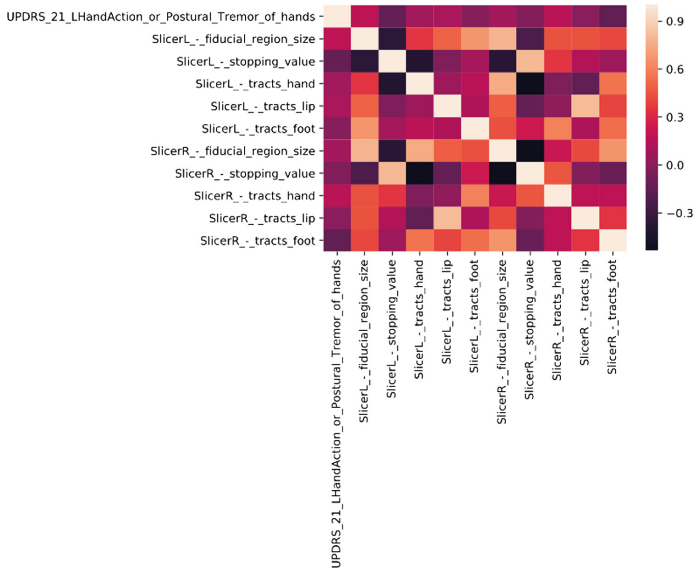

Fig. 4. Computed correlation between Slicer parameters and UPDRS examination: handwriting. The similarity between the two parameters is correlated with the color of the cell in the matrix (lighter shade represents higher similarity and contrariwise). (Color figure online)

Table 4. Confusion matrix of UPDRS #8: handwriting distortion by rules obtained from 3DSlicer values based on DTI data. Total number of tested subjects: 6. Accuracy (total): 0.878. Coverage (total): 0.667.

		Predicted							
		0	**1**	**2**	**3**	**4**	**No. of obj.**	**Accuracy**	**Coverage**
Actual	**0**	3.33	0.17	0.00	0.00	0.00	4.00	**0.944**	0.878
	1	0.00	0.00	0.00	0.00	0.00	0.83	**0.000**	0.000
	2	0.00	0.00	0.17	0.00	0.00	0.50	**0.167**	0.083
	3	0.00	0.00	0.33	0.00	0.00	0.33	**0.000**	0.167
	4	0.00	0.00	0.00	0.00	0.00	0.33	**0.000**	0.000
	TPR	**1.00**	**0.00**	**0.17**	**0.00**	0.00			

5 Conclusions

Our recent research described above was meant to determine if data mining can predict possible Parkinson's patient's symptoms based only on the DTI data of patients who go through the DBS surgery. We have applied the rough set theory on the data obtained from DTI after the operation to conclude whether it is possible to create a system that is unbiased of human opinion.

The results have shown that it is possible to introduce a new, autonomous, doctor independent and a highly accurate method of disease course prediction.

What is more, this approach enables a new way for a deduction of an impact of a region-specific stimulation of STN and its effect on patients.

However, since this results have been based on a small data set, further work is required to perform more credible statistics and verification in the sake of elimination of overfitting problem.

References

1. Benabid, A.L., et al.: Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease. Lancet Neurol. **8**(1), 67–81 (2009)
2. Nambu, A., Tokuno, H., Takada, M.: Functional significance of the cortico–subthalamo–pallidal 'hyperdirect' pathway. Neurosci. Res. **43**(2), 111–117 (2002)
3. Szymański, A., Przybyszewski, A.W.: Rough set rules help to optimize parameters of deep brain stimulation in Parkinson's patients. In: Ślęzak, D., Tan, A.-H., Peters, J.F., Schwabe, L. (eds.) BIH 2014. LNCS (LNAI), vol. 8609, pp. 345–356. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-09891-3_32
4. Szymański, A., Kubis, A., Przybyszewski, A.W.: Data mining and neural network simulations can help to improve deep brain stimulation effects in Parkinson's disease. Comput. Sci. **16**(2), 199 (2015)
5. Pawlak, Z.: Rough set theory and its applications. J. Telecommun. Inf. Technol., 7–10 (2002)
6. McKinney, W.: Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference, vol. 445 (2010)
7. Przybyszewski, A.W., et al.: Multi-parametric analysis assists in STN localization in Parkinson's patients. J. Neurol. Sci. **366**, 37–43 (2016)
8. Benazzouz, A., et al.: Intraoperative microrecordings of the subthalamic nucleus in Parkinson's disease. Mov. Disord. Official J. Mov. Disord. Soc. **17**(S3), S145–S149 (2002)
9. Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease: The unified Parkinson's disease rating scale (UPDRS): status and recommendations. Mov. Disord. **18**(7), 738–750 (2003)

sensors

MDPI

Review

Machine Learning and Eye Movements Give Insights into Neurodegenerative Disease Mechanisms

Andrzej W. Przybyszewski [1,2,*], Albert Śledzianowski [1], Artur Chudzik [1], Stanisław Szlufik [3] and Dariusz Koziorowski [3]

[1] Polish-Japanese Academy of Information Technology, The Faculty of Information Technology, 86 Koszykowa Street, 02-008 Warsaw, Poland
[2] Department of Neurology, University of Massachusetts Medical School, 65 Lake Avenue, Worcester, MA 01655, USA
[3] Department of Neurology, Faculty of Health Science, Medical University of Warsaw, 8 Kondratowicza Street, 03-242 Warsaw, Poland
* Correspondence: przy@pjwstk.edu.pl or andrzej.przybyszewski@umassmed.edu

Abstract: Humans are a vision-dominated species; what we perceive depends on where we look. Therefore, eye movements (EMs) are essential to our interactions with the environment, and experimental findings show EMs are affected in neurodegenerative disorders (ND). This could be a reason for some cognitive and movement disorders in ND. Therefore, we aim to establish whether changes in EM-evoked responses can tell us about the progression of ND, such as Alzheimer's (AD) and Parkinson's diseases (PD), in different stages. In the present review, we have analyzed the results of psychological, neurological, and EM (saccades, antisaccades, pursuit) tests to predict disease progression with machine learning (ML) methods. Thanks to ML algorithms, from the high-dimensional parameter space, we were able to find significant EM changes related to ND symptoms that gave us insights into ND mechanisms. The predictive algorithms described use various approaches, including granular computing, Naive Bayes, Decision Trees/Tables, logistic regression, C-/Linear SVC, KNC, and Random Forest. We demonstrated that EM is a robust biomarker for assessing symptom progression in PD and AD. There are navigation problems in 3D space in both diseases. Consequently, we investigated EM experiments in the virtual space and how they may help find neurodegeneration-related brain changes, e.g., related to place or/and orientation problems. In conclusion, EM parameters with clinical symptoms are powerful precision instruments that, in addition to their potential for predictions of ND progression with the help of ML, could be used to indicate the different preclinical stages of both diseases.

Keywords: Alzheimer's disease; Parkinson's disease; eye movements; Rough Set; machine learning

Citation: Przybyszewski, A.W.; Śledzianowski, A.; Chudzik, A.; Szlufik, S.; Koziorowski, D. Machine Learning and Eye Movements Give Insights into Neurodegenerative Disease Mechanisms. *Sensors* **2023**, *23*, 2145. https://doi.org/10.3390/s23042145

Academic Editors: Lin Meng, Hiroyuki Tomiyama, Kenshi Saho and Xiangbo Kong

Received: 5 January 2023
Revised: 7 February 2023
Accepted: 13 February 2023
Published: 14 February 2023

1. Introduction

Neurodegenerative brain changes start about two decades before the first detectable symptoms [1,2]. During this period, everyone develops various plastic compensatory brain mechanisms. The rates and processes of neurodegenerative disease (ND) progressions have a vast patient-specific spectrum. The main aim of this review is to look, from this perspective, at the early processes related to neurodegenerative changes and to precisely characterize them by using granular computing (GC) or other ML methods. Our leading candidate for a possible early biomarker is related to eye movements (EM). We have concentrated on reflexive EM, such as reflexive saccades (RS), antisaccades (AS), and pursuit EM. Generally, these movements slow down and become less precise with aging, but these changes are more pronounced due to ND. It means that one needs to differentiate between processes related to aging and those related to ND, which is one of the major problems with early ND biomarkers. We base our approach on identifying similarities between different groups of patients and whether patients' EM and symptoms become

more like those of patients in the advanced stages of ND with disease progression. We have demonstrated this approach in different groups of Parkinson's disease (PD) patients. In the next step, we assumed that more minor cognitive and motor symptoms might be related to more minor changes in EM parameters. This could be a possible approach for the early detection of preclinical ND.

There are three kinds of symptoms in neurodegenerative disease. As different symptoms are related to different structures in the brain, there is an essential question as to whether ND processes are independent of or specific to different parts of the brain. The three kinds of symptoms are as follows:

- Cognitive symptoms are dominant in Alzheimer's disease (AD) but are secondary in Parkinson's disease (PD);
- Motor symptoms are characteristic of PD and less evident for AD;
- Emotional symptoms are mostly related to depression and are observed in both diseases but are characteristic of LOAD (late-onset AD) [3,4].

It is widely recognized that neurodegenerative processes start from the basal ganglia in PD and typically from the hippocampus in AD. In PD, the structure that degenerates first is the substantia nigra (part of the basal ganglia), whose neurons are responsible for releasing dopamine. The lack of dopamine slows down and affects movement regulation (GO—NOGO antagonism) [5]. In addition, there is a connection between the striatum and the prefrontal cortex that influences memory and cognition [5]. Finally, as dopamine is the reward neurotransmitter, it affects the reward systems (pleasure feelings) and might lead to depression [3].

The place cells in the hippocampus are responsible for orientation deficiency (navigation problems) in AD, and connections between the hippocampus and the frontal cortex are responsible for memory and cognitive problems. Through its connections to the striatum, the prefrontal cortex might also affect movement in AD. Depression is the main reason for the late onset of AD (LOAD) in older subjects (over 65 years of age) [4].

This review is mainly based on work with Parkinson's patients under different therapies and at various stages of the disease. We have demonstrated how different ML algorithms can help to predict disease development and how, by using ML, we can compare PD symptoms of less and more advanced patients. Therefore, the main question is: if we can predict disease development, can we also predict the start of ND disease? This question might also be related to looking for sensitive biomarkers that might indicate neurodegenerative brain changes even before the first observable symptoms. We have extensively tested and evaluated eye movement (EM) as a potential biomarker in PD.

Multiple studies have demonstrated that EM helps to predict peripheral movement disorders as well as cognitive and emotional symptom progressions. EM is also affected in AD, so it could be an excellent biomarker for both diseases. One very characteristic parameter of EM in both PD and AD is the delay of the reflexive saccades (RSLat—the time difference between the light spot position change and the start of eye movement). However, RSLat is also related to aging and, to be a significant biomarker, must also be linked to changes in many other, still unidentified, parameters.

To find practical consequences of changes in EM and other ND-related parameters for everyday life, we can immerse the user in a virtual reality (VR) world. The advantage of VR in comparison to the real environment is that one can control significant features of the VR world and test the influence of different parameters on the subject's behavior.

Hence, one needs to test a high-dimensional and noisy parameter space, and we suggest that ML methods are the right tool for it.

2. Eye Movements and Neurodegenerative Diseases

2.1. Standard Neurological Approach

Experienced neurologists use their clinical knowledge and experience from many years of practice to estimate symptom development and the best treatment for an individual Parkinson's disease (PD) patient. However, because of the long period of compensatory

mechanisms unique to each patient, there is a famous saying that "No two people face Parkinson's in quite the same way." Therefore, a neurologist must consider not only motor symptoms, but many others, and must even try to understand a patient's cognitive and emotional symptoms through the Theory of Mind space [6,7] to estimate disease progression in the individual patient.

Typically, eye movement analysis coexists in the broader context of other neuropsychological measurements. They can be related to the following tests: PDQ-39/PDQ-8—a 39/8-question test related to health difficulties in everyday living (summaries of both tests are strongly correlated), ESS—Epworth sleeping scale results (related to sleepiness problems, predominantly during the day), BDI—Beck's depression inventory (21-item quantitative measure of symptoms of depression), TMT A and TMT B tests (Trail Making Test—part A measures psychomotor speed and part B is related to executive function). Another test, AIMS—Abnormal Involuntary Movement Score—measures the involuntary movements of patients. Additionally, neurologists can use more wide-ranging cognitive tests, such as MoCA (Montreal cognitive assessment) for the detection of MCI (mild cognitive impairment) for PD and AD or dementia (AD), or the similar, but shorter, MMSE (Mini-Mental State Examination), which is less sensitive than MoCA, but has established clinical values. Another important test, the FAS (Phonemic Verbal Fluency) test involves orally producing words that start with the letters F, A, and S, evaluates the cognitive function and measures language-related executive functioning. The most common decisive attribute is the UPDRS (Unified Parkinson's Disease Rating Scale), an essential neurological test for the effects of PD long-treatment effects. The UPDRS score estimates daily activity (non-motor—UPDRS I and motor—UPDRS II) and motor-related problems (UPDRS III). It is the so-called "Gold Standard" for determining PD progressions. Alzheimer's disease progression is mainly related to cognitive changes (normal, MCI, dementia), and disease progression is determined by the CDR (Cognitive Dementia Rating) scale as the decision attribute.

2.2. EM in PD—Saccades

There is pervasive, clinically oriented, literature related to reflexive saccade (RS) latency in PD in comparison to same-age healthy controls showing different, contradictory results [8]. However, a meta-analytic review [9] demonstrated, by analyzing 47 representative studies (from 1529 references), that the RS latency depends on which method was used by different authors: gap, step, or overlap. Different methods are determined by the time difference (Δt) between the fixation point disappearance and the target appearance: if $\Delta t > 0$, then it is the gap method; if $\Delta t = 0$, then it is the step method; and if $\Delta t < 0$, it is the overlap method. By using quantitative pooling analysis, Chambers and Prescott [9] demonstrated that the slowed response time in PD, compared to a control, is strongest in the step method ($\Delta t = 0$), weaker in the gap method ($\Delta t > 0$), and negligible in the overlap method ($\Delta t < 0$).

In a study [10], standard neurological attributes and PDQ39, Epworth, and AIMS were measured, as well as EM—reflexive saccades (Figure 1). The authors compared different algorithms for the classification of UPDRS, UPDRS II, and UPDRS III in 10 PD patients. A six-fold cross-validation method was used. For the accuracy measure of UPDRS, the best results were given by RSES (Rough Set Exploration System based on Rough Set theory, which is an important implementation of GC—granular computing) with a global accuracy of 0.90, the second was Random Forest, which had a global accuracy of 0.68, and the third was a Tree Ensemble, with 0.65 global accuracy. For UPDRS II, the best results were for the Random Forest with an accuracy of 0.80, the second was RSES with 0.79, and the third was Bayesian classification, which gave an accuracy of 0.77. For the UPDRS III, the best was RSES classification, which gave a 0.82 accuracy, and the Decision Table (with Weka), which gave an accuracy of 0.77.

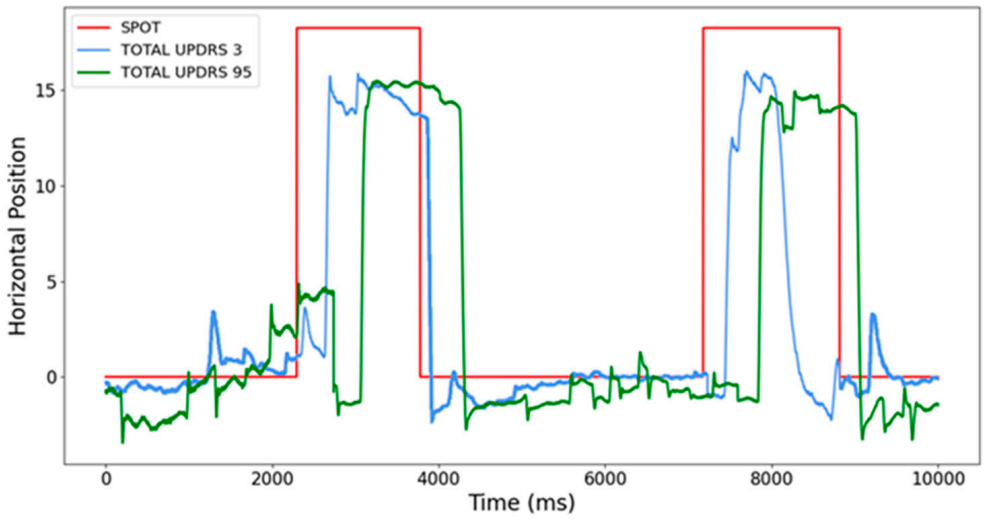

Figure 1. Saccades at various stages of PD. Recordings were performed with different UPDRS measured in clinical conditions under a doctor's supervision. The red line is related to light spot movements, and the blue and green curves are related to reflexive saccades performed by two different PD patients in different stages of the disease: the blue curve denotes EM at the beginning of the disease, and the green curve at advanced PD. Notice the significant difference in response latencies [11].

The results above agree in general with a recent publication [12] where they tested horizontal and vertical saccades and antisaccades (AS) for healthy control and PD subjects in different stages of the disease, measured by Hoehn and Yahr stages (H&Y2 and H&Y3). The PD group displayed decreased vertical saccade amplitudes and increased vertical saccade and AS latencies. AS latency increased for H&Y2 and H&Y3 patients, but AS errors (correlated with MoCA—Montreal cognitive assessment test's score) were similar for control and H&Y2 subjects, but larger for H&Y3 subjects. Levodopa has increased vertical saccade latency but decreased AS latency. This work has described many different difficult-to-analyze mechanisms that might be easier to put through machine learning methods.

In one study [13], it was demonstrated that saccades could predict cognitive decline in PD patients. In 140 PD and 90 age-matched participants, the authors evaluated differences in RS metrics between early-PD and healthy age-matched adults. They assessed RS and cognition at baseline at 18, 36, and 54 months. RS parameters were latency, duration, amplitude, peak velocity, and average velocity, and the cognitive assessment contained executive function, attention, fluctuating attention, and memory. RS parameters, with the help of linear mixed-effects models, were used as predictors of cognitive decline over 54 months. At the baseline, RS was impaired in PD patients compared to the control group. RS parameters predicted a decline in global cognition, executive function (verbal fluency), attention, and memory over 54 months in PD patients. However, only reductions in global cognition and attention were predicted by RS parameters in age-matched subjects, which means that cognitive changes were not just age-related [13]. The dependence between RS latency and executive functions was also confirmed earlier [14]. In addition, manual and saccadic performances are uncorrelated in the average population, but both are similarly affected by PD [15].

Abasi et al. [16] tested whether vestibular therapy exercises might recover both oculomotor functions and postural control (in the upright position) in patients with PD. Vestibular therapy is a set of exercises that affect the basic elements of central sensorimotor integration. They tested 11 idiopathic PD patients voluntarily contributing to the survey based on the following criteria: central vestibular dysfunction and Hoehn and Yahr scale scores ≤ 3. Videonystagmography (VNG) and Berg Balance Scale (BBS) scores were measured as the control. PD patients undertook vestibular rehabilitation training for 24 sessions (3 sessions

per week). VNG and BBS scores were appraised again 48 h after completing the last session of the exercise. Investigators detected significant advances in balance ($p \leq 0.015$), and eye-tracking and gaze performance were statistically enhanced in seven and six patients, respectively.

Wong et al. [17] investigated the relationship between EM parameters and execution in different cognitive tests in Parkinson's patients. In the eye-tracking experiment, subjects were asked to look for a number embedded in an array of alphabets distributed randomly on a workstation screen. Researchers calculated the average amplitude of saccades and fixation duration and correlated these data with the results of cognitive tests. It was established that prolonged fixation time was linked with inferior performance in verbal fluency, as well as in visual and verbal memory, showing that EM parameters are alternate markers for cognitive function in PD patients.

Wong et al. [18] studied the correlations between EM and cognitive functions in patients with Parkinson's disease. A total of 62 patients with non-demented Parkinson's disease and 62 controls of the same sex, similar age, and equivalent education were exposed to cognitive and oculomotor tests. Researchers observed a negative correlation between the length of eye fixation and functioning in semantic verbal fluency, as well as verbal and visual memory undertakings. Researchers concluded that increased duration of visual fixation correlates with poor results in semantic verbal fluency and verbal and visual memory tasks in non-dementia PD. In another 2-year follow-up study, 49 of the primary 62 patients (15 cases/31% were classified as MCI (with mild cognitive impairment)) were examined in the context of the relationship between domain-specific cognitive impairment and the progression of visual fixation duration. Researchers noticed that the duration of fixations increased significantly after two years. For the analysis, they used ANOVA with repeated measurements, and according to the results, impairment of semantic verbal fluency, visual and verbal recognition memory, and indicative attention function had a significant effect on the extension of the duration of visual fixation [18]. Repeated measures beyond two years showed a correlation between prolonged visual fixation and different types of cognitive impairment associated with cholinergic dysfunction. According to the authors, it provides preliminary evidence that the observed eye-tracking model is a surrogate marker for the cholinergic deficit in Parkinson's disease.

Archibald et al. [19] assessed the error rates and visual exploration tactics of subjects with Parkinson's disease, in relation to the extension of their cognitive impairment, while performing a battery of visuo-cognitive tasks. They found that error rates were significantly higher in those PD groups with either MCI ($p = 0.001$) or dementia ($p < 0.001$) in comparison to cognitively normal subjects with Parkinson's. When matched to cognitively normal Parkinson's disease patients, the exploration tactic, as measured by EM-tracking variables, was least efficient in the dementia group and inefficient to a lesser extent in MCI patients. In control subjects and cognitively normal subjects with Parkinson's, it was established that saccade amplitude was drastically reduced in the groups with MCI or dementia.

The fixation period was stretched in all PD patients in relation to healthy control subjects but was most prolonged for cognitively impaired PD groups. The average fixation period was strongly related to disease severity. The authors concluded that the increase in the fixation period, existing even in cognitively normal patients with PD, implies a disease-specific influence on the systems directing visual search.

2.3. EM in PD—Antisaccades

One of the well-proven experimental models used to examine the inhibition of automatic reflexive responses is the antisaccade task (AS) [20].

In a study [21], the significance of antisaccade (AS) parameters for the classification of Parkinson's disease motor and motor variations (UPDRS II and UPDRS IV) was tested. There were 11 PD patients examined in 4 sessions. In addition to the standard neurological attributes, AS parameters such as delay, duration, and maximum speed were measured. RSES was used for the data discretization and attribute reduction and to perform a 5-fold

cross-validation. The best result was obtained by the RSES Decomposition Tree, which splits the dataset into fragments represented as a tree's leaves [21]. The UPDRS III classification results indicated an accuracy of 0.85 with a coverage of 0.48. Surprisingly, the UPDRS IV was estimated with an accuracy of 0.91 and coverage of 0.39, so UPDRS IV showed a more significant correlation with antisaccade parameters. Thus, UPDRS IV showed greater sensitivity in predicting antisaccade parameters [21]. From the results, it also emerged that attributes describing methods of patient treatment (again, the session attribute) and mean duration were most sensitive in predicting the scores of both UPDRS III and IV. An example of AS in two different PD patients is shown in Figure 2.

As described above, different analysis methods influence the RS latency [9]. In a review meta-analysis, Waldthaler et al. [22] analyzed the influence of the paradigm (gap, step, overlap) on AS latency and errors. They [22] compared the results of 703 PD patients with 600 healthy controls for antisaccade latency and 831 patients and 727 healthy controls for antisaccade error rate. Over 60% of studies excluded PD with dementia. Like RS latencies, the mean AS latency was 339.8 ms in the PD patients and 294.2 ms in the healthy group in the gap paradigm, and 411.7 ms in the PD patients and 368.6 ms in the healthy group in the step paradigm. This was measured for PD patients with disease duration between 0.7 and 14.7 years and UPDRS III scores between 5 and 85, from early to advanced disease stages. In a meta-analysis, the authors [22] demonstrated that AS latency increases with disease severity, and an increase in the levodopa dosage influences the AS error rate (negatively moderating effect).

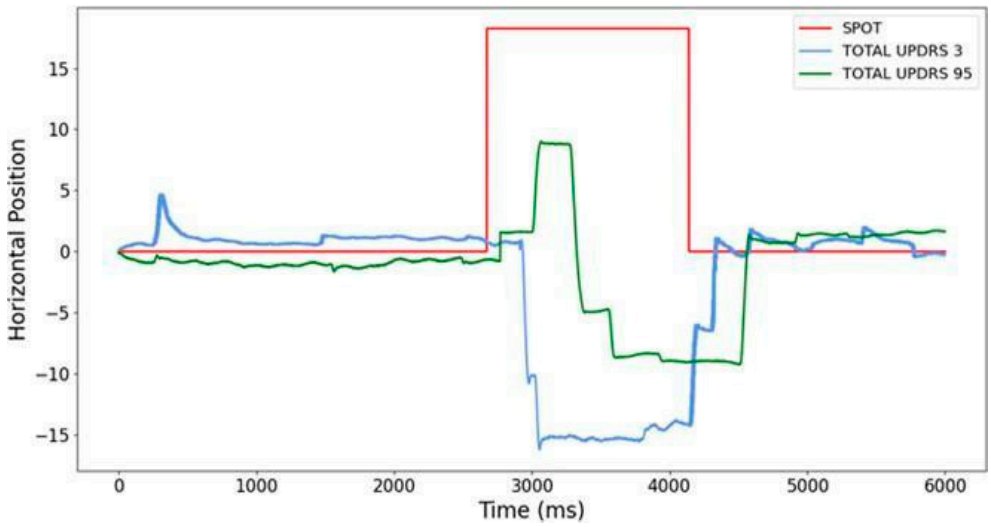

Figure 2. Antisaccades (AS) at various stages of PD. The red line is related to the light spot movements, and the blue and green curves are related to reflexive AS performed by two different PD patients in different stages of the disease: the blue curve denotes EM at the beginning of the disease, and the green curve in the advanced PD stage. Notice the significant difference not only in response latencies, but also in AS speed, and that a more advanced PD patient started with a saccade that changed to AS [21].

A study by Waldthaler et al. [23] tested whether patients with Parkinson's taking dopaminergic medication performed better at response inhibition during antisaccade tasks. Levodopa intake has favorable or harmful effects on dopamine-dependent cognitive tasks based on essential basal dopamine intensities in ventral segments of the striatum, agreeing with the dopamine overdose theory. Thirty-five patients with Parkinson's (and 30 healthy subjects) completed antisaccade tasks in OFF and ON medication conditions. Investigators computed multiple linear regressions to forecast the alterations in antisaccade delay and directive mistakes, and to express saccade rate based on age at Parkinson's disease onset, disease duration, levodopa-equivalent circadian amount, motor indicator difficulties, and

executive functions. According to their results, earlier disease onset and milder motor symptoms in the OFF-medication status were related to diminished inhibition ability response after levodopa intake, mirrored in enlarged express saccades and mistakes. They concluded that levodopa might have opposite results on oculomotor reaction inhibition contingent on the age at Parkinson's disease onset and motor disease gravity.

During their next study, Waldthaler et al. [24] examined whether there was any correlation between the development of motor and cognitive indications in 25 patients and Parkinson's disease (age: 61.4 +/− 6.8, disease duration: 6.0 +/− 4.5 years). A total of 10 patients from all 25 PD patients received subthalamic nucleus DBS (deep brain stimulation) during the follow-up period (from DBS surgery to follow-up visit: 4.5 +/− 2.1 months). All PD patients were examined in ON medication and ON-DBS states, and modifications of dopaminergic treatment were permitted during the follow-up epoch. PD patients without DBS who displayed substantial improvement in motor signs after one year also received higher levodopa equivalent dosages at follow-up. Generally, the antisaccade (AS) delay (baseline: 339 +/− 72 ms, mean change: 95 +/− 1.1 ms) and mistake rate (baseline: 0.52, mean change: −0.02 +/− 0.3) stayed steady in the non-DBS group. In the DBS group, the AS delay tended to increase (baseline: 295 +/− 78 ms, mean change: 48 +/− 75 ms ($p = 0.09$)), but the mistake rate improved at follow-up (baseline: 0.76, mean change: −0.21 +/− 0.3 ($p = 0.048$)). The change in AS delay was connected to change in MDS-UPDRS III in both groups (non-DBS group baseline: 25.7 +/− 13, mean change: 0.3 +/− 7.4; DBS group baseline: 27.2 +/− 16.5, mean change: −5.2 +/− 17.8) and with the change in MoCA score in the non-DBS group (25.8 +/− 3.1, mean change 1.3 +/− 3.1). The authors indicate that AS delay may be sensitive to the development of motor and cognitive signs over time in Parkinson's disease patients.

2.4. EM in PD—Saccades and Antisaccades

The same group of patients as in [21] was used for UPDRS prediction based on RS and AS measurements. The best accuracy of 0.89 was achieved by Decision Trees [11]. The results showed that the accuracy of the predictions increased with the number of significant attributes that were obtained by, for example, averaging RS and AS duration or by adding the averaged standard deviations of each patient's latencies [11].

The authors of [22] demonstrated that RS and AS latencies were correlated with the results of neuropsychological tests in 65 PD patients, but only the results for AS latencies concerning patients' cognitive impairment were statistically significant. In a study [25], 19 drug-naïve PD patients and 20 age-matched controls were examined. Patients had clinically probable idiopathic disease within three years of disease onset. Their RS latencies were like those of the controls, but AS error rates differed significantly (PD 15% vs. 8.7% for controls).

Fooken et al. [26] studied different tasks and conditions in which the oculomotor function in Parkinson's patients is preserved. A total of 16 patients with Parkinson's disease and 18 healthy, age-matched controls performed a set of tasks of saccades (RS), anti-saccades (AS), pursuits, and rapid 'go/no-go' manual interventions. Compared to the control group, PD patients showed regular impairment in tasks with fixed targets: prosaccades were hypometric, and AS were wrongly started towards the indicated target in 35% of the trials compared to 14% of errors in the control group. In PD subjects, task errors were linked with short-latency saccades, demonstrating anomalies in inhibitory control. However, the patients' EMs in response to dynamic targets were well-preserved. Parkinson's disease patients can track and predict a moving target and make quick go/no-go decisions with the same precision as healthy people. The intercepting hand movements of the patients were slower on average but indicated adaptive processes compensating for the motor slow down. Researchers concluded that the preservation of eye and hand movement functions in PD is linked to a separate functional pathway through the upper colliculus–brainstem loop that detours the frontal–basal ganglia network.

Kocoglu et al. [27] investigated how social processes and behaviors change in PD during spatial signaling tasks. Socially relevant directional cues, such as photos of people looking left or right, have been found to redirect attention. In conclusion, the basal ganglia can play a role in responding to such directional signals. In this research, patients and healthy controls performed pro- and anti-saccade tasks in which different directional signs preceded the appearance of the target. They analyzed reaction time, prediction errors, and correlations with PD severity and cognitive assessment scores. Patients displayed increased errors and answer times with the AS (antisaccade) task, but not with the RS (saccade) task. The control subjects made the most predictive errors in the finger-pointing trials, and the PD patients were mostly affected by the arrow, gaze, and pointing clues. It has been found that PD patients have a reduced ability to suppress responses to directional signals, but this effect is not specific to social signals.

Munoz et al. [28] studied whether bilateral deep stimulation of the basal ganglia–subthalamic nucleus (STN DBS) may affect the control of inhibition of eye movement in PD. They investigated the effect of DBS amplitude on inhibitory power during an antisaccade procedure on 10 PD patients after their DBS surgery. Subjects without medication (12 h, overnight) performed the antisaccade tasks with a set of different DBS stimulation amplitudes (from 0—no stimulation to 5—higher levels). The prosaccade error rate (related to a saccade at the beginning of the antisaccade) increased with increasing DBS stimulation amplitude ($p < 0.01$). Moreover, the saccade error rate increased with the decrease in the modeled volume of tissue activated (VTA) and decreased overlap of the STN stimulation area, but this connection was determined by the stimulation amplitude ($p = 0.04$). They concluded that the directional prosaccade error rate during the antisaccade task indicated impaired inhibitory control and suggested that higher stimulation amplitude settings can be modulatory for inhibitory control.

2.5. EM in PD—Pursuit

Another study tested how effective diagnostic parameters of slow (pursuit) eye movements are for the prediction of PD symptom development [29,30]. Horizontal pursuit EM with three different sinusoidal movement speeds was measured. The gain and accuracy (EM measurement section) were estimated. The discretization and attribute reduction with RSES demonstrated that the significant attributes were precise for the accuracy of the fastest sinusoidal movement speed, and gains decreased for medium and high sinusoidal movement light spot speeds [29,30]. The result of the 4-fold cross-validation gave a global accuracy of 0.77 for the UPDRS III prediction. An accuracy of 0.8 for the session number prediction (different treatments) in 10 PD patients was found. The above predictions were obtained for a sample of 20 patients using different binning methods (KNIME auto-binner), which allowed the grouping of UPDRS III data in intervals of equal frequencies. A 90% accuracy in predictions on these data was achieved with the RSES and 5-fold cross-validation [30]. When comparing the accuracy results of different classifiers, the RSES is in first place in the ranking, ahead of SVM (59%), Naive Bayes (55%), and Random Forest (52%) [30].

In this context, in her review, Frei [31] analyzed 29 articles (from 819 found) on smooth-pursuit eye movements in PD patients and compared them to those in normal subjects. She found that in 18 articles, the gain was measured and reduced in PD patients compared to controls in 16 of these papers. In two papers, the gain was reduced for higher target velocities. In three articles, accuracy was measured and found to be reduced in PD. There were also correcting saccades during smooth-pursuit EM that were more dominant in more advanced PD and for faster smooth pursuits, but quantification of saccades was difficult [31].

In another study, deep brain stimulation (DBS) increased smooth-pursuit accuracy ($p < 0.001$) and smooth-pursuit gain ($p = 0.005$), especially for faster smooth pursuits ($p = 0.034$) [32].

In their study, Farashi et al. [33] observed eye movements (EMs) during inactive states (eyes closed and eyes open), measuring EM using vertical electrooculography (VEOG). They performed the analysis in the time, frequency, and time–frequency axes of the VEOG time series. The authors completed a categorization by comparing healthy subjects and PD patients in OFF and ON medication conditions. They used an SVM (support vector machine) classifier and allowed multiple-differentiation-corrected p-values. The VEOG data achieved 69.10% and 87.27% discrimination precision for OFF and ON medication conditions, respectively. The authors established that PD patients' vertical EM had smaller amplitude changes than healthy subjects in OFF medication conditions. The levodopa treatment augmented such changes in vertical EM during the eyes-closed situation and diminished during the eyes-open situation. As a result of levodopa treatment, VEOG time series amplitudes may change, although vertical EM rates were not affected (frequency contents).

2.6. EM in PD—Pupillometry

Parkinson's disease patients develop a distorted pupillary response dependent on an abnormality in the retinal ganglion cells. Tabashum et al. [34] illustrated an arrangement for pupil size estimates that permits the discovery of pupil parameters to measure the post-illumination pupillary response (PIPR) with a Kalman filter estimating the pupil center and diameter over time. The pupillary reaction was estimated in the contralateral eye to two diverse light stimuli (470 and 610 nm) for 19 Parkinson's patients and 10 healthy subjects. Net PIPR displayed different reactions to wavelengths (0.13 mm for Parkinson's patients and 0.61 mm for healthy subjects, proving an extremely significant differentiation ($p < 0.001$)).

Tsitsi et al. [35] evaluated gaze constancy and pupil size in steady light surroundings, as well as eye movements (EMs) during constant fixation in a group of 50 Parkinson's disease subjects (66% males) with unilateral to mild symptoms (Hoehn and Yahr 1–3; Schwab and England 70–90%) and 43 control subjects (37% males) with an eye tracker (1200 Hz) and logistic regression analysis. They examined the potency of the relationship of EM measures with the ROC curve results of 0.817, 95% CI: 0.732–0.901, and concluded that eye-tracking-established amounts of gaze fixation and pupil reaction might be valuable biomarkers of Parkinson's disease indications.

2.7. EM in PD—Multimodal Approach

Bonnet et al. [36] investigated how connections between vision and posture are exaggerated in Parkinson's patients. PD subjects have been shown to display unusually low levels of synergy in their posture self-control. These impaired reactions are related to the neurodegeneration processes in Parkinson's disease that affect the basal ganglia, which facilitate the integration of both types of movements. They tested 20 PD patients (mean age: 60) on levodopa and 20 age-matched-healthy subjects (mean age: 61) with a detailed visual assignment (target-seeking scenarios in an image) and an inaccurate control task (arbitrarily viewing an image) in which pictures were projected onto a large screen. Lower back, upper back, head, and EM were registered simultaneously. To analyze behavioral synergies, the authors computed Pearson correlations between EM and postural actions. The associations between EM and upper- and lower-back movements were diminished in Parkinson's subjects. The healthy control subjects did not display important correlations between EM and postural activities. Generally, their results revealed that the Parkinson's subjects were unable to correct and change their postural rigidity to achieve success in the visual task. Moreover, these problems may occur in the early stages of Parkinson's (an early biomarker opportunity).

Zhang et al. [37] investigated 49 Parkinson's patients, including 35 early-stage (Hoehn and Yahr: 1–2 staging) and 14 advanced PD subjects (Hoehn and Yahr scale: 3 to 5 staging) and 23 healthy subjects. In addition to clinically significant PD symptoms, videooculography was used to measure EM features such as eye fixation stability, horizontal and

vertical reflexive saccade (RS), and horizontal and vertical smooth-pursuit movements. The authors discovered that five EM features—specifically square wave jerk frequency, vertical RS delays, the accuracy of the vertical–upward RS, and the horizontal smooth-pursuit RS gain—were meaningfully different in Parkinson's and normal subjects. By merging all five features, the authors achieved a symptomatic sensitivity of 78.3% and a specificity of 95.2%. The study discovered that more deficiencies in upward–vertical RS than in other directions were related to disease duration and the stage of development of Parkinson's disease.

Perkins et al. [38] investigated whether Sleep Behavior Disorder (RBD) indicates PD. With video-based eye tracking, researchers tested saccade, pupillary, and blink responses in RBD and isolated REM (rapid eye movement) with 22 PD and 22 RBD patients and 74 healthy controls. They found that RBD patients did not have significantly different saccades compared to healthy controls, but PD patients differed from both healthy controls and RBD patients. They concluded that RBD and PD patients had altered pupil and blink behavior compared to healthy controls. Because RBD saccade parameters were comparable to healthy controls, brain areas responsible for pupil and blink control may be impacted before saccadic control areas, making them a potential prodrome of PD.

2.8. Prediction of Disease Progression in Different PD Groups

The goal in [39] was to predict Parkinson's disease progression in advanced-stage patients based on data obtained from patients under different treatments and at different stages of the disease. Patients from the BMT group (only on medication, third visit), DBS group (after recent deep brain stimulation surgery, third visit), and POP group (after older DBS surgery, first visit) were used as a training dataset—a model. The model was tested on the data obtained from the POP group during the second visit. A dedicated data science framework written in Python was used and based on the Scikit Learn and Pandas libraries that implemented different multiclass strategies, such as k-Nearest Neighbors Classifier, Support Vector Classifier, Decision Tree Classifier, and Random Forest Classifier. In this trial, the Random Forest Classifier achieved the highest overall accuracy score of 0.75 and an accuracy of 0.7 when predicting subclasses of UPDRS for patients in advanced stages of the disease who responded to treatment, with a global 0.57 accuracy score for all classes [39].

The purpose of another study [40] was to predict the results of different PD patient treatments to find the optimal one. The study compared the intelligent methods based on Rough Set theory with several different machine learning algorithms, namely Gaussian Naive Bayes, Decision Tree, Logistic Regression, C-Support Vector, Linear SVC, and Random Forest. Generally, the Rough Set method gave better accuracy, but less coverage, than other algorithms. On the other hand, the Rough Set-based approach allows the creation of more general rules without the necessity of additional data splitting (into different sessions), which was required in the other ML models to obtain accuracies similar to those obtained by RS. An example is the prediction of UPDRS in a DBS patient group from rules obtained from BMT patients. Global accuracy for DBS patients was 0.64 for the first visit, 0.85 for the second visit, and 0.74 for the third visit. Other methods gave accuracies of 0.88, 0.58, and 0.54, respectively [40].

The principal conclusion from this comparison is the observation that RS is a much more universal method when considering medical data. Finally, it was demonstrated that it is possible to estimate symptoms and their time development in populations treated differently, which may, in the future, lead to the discovery of universal rules of PD progression and to the optimization of treatment.

2.9. Prediction of Disease Progression Related to Motor, Cognitive, and Emotional Longitudinal Changes in PD Patients

In [41], two BMT groups of patients (only on medication) were tested. The first one, less advanced, was tested three times every half year (visit 1, visit 2, visit 3). In the second BMT group, more advanced patients were tested only once. All tests were performed with the following condition attributes: PDQ39, Epworth, depression score (Beck test), TMT A

and B, disease duration, and fast EM. The decision attribute was UPDRS. With the help of Rough Set theory (RSES), rules describing the more advanced BMT group were constructed and used to predict disease progression over three visits in the less advanced BMT group of patients. Using all condition attributes, general rules gave accuracies as follows: visit 1—0.68, visit 2—0.86, and visit 3—0.88. When rules were related only to motor attributes, the accuracies were as follows: visits 1—0.80, 2—0.93, and 3—1.0. For rules related to cognitive attributes, the results were as follows: visit 1—0.50, visit 2—0.60, and visit 3—0.64. The higher accuracy can be interpreted as more similar patient symptoms. General and motor-related accuracies increased with disease progression (visit numbers), which means that the less advanced group of patients became more like the advanced group. However, this was not the case for cognition-related symptoms that gave lower accuracies, which means that their progressions were not as strongly correlated with disease development.

The influence of the patient's emotions on the accuracy of the predictions of disease progression in the same group or different groups of patients was also tested through the depression score (Beck test) [42]. The progressions of the BMT group (only on medication) for visits 2 and 3 and the DBS group (deep brain stimulation) for visit 1 were compared based on the BMT symptoms during visit 1. The predictions were performed with the help of RSES and with standard neurological testing and EM parameters. Based on rules from first visit BMT patients, the prediction of symptoms (UPDRS) of BMT for visits 2 and 3 had accuracies of 0.7 and 0.7, but by adding the depression score, accuracies increased to 0.77 and 0.80 [42]. Similar predictions were calculated for the DBS group progression based on first visit BMT rules. Accuracies obtained for the DBS group were as follows: visit 1—0.64, visit 2—0.77, and visit 3—0.74. Adding the depression score to all attributes, improved accuracies of visit 1 to 0.77, visit 2 to 0.85, and visit 3 to 0.8 were demonstrated [42]. In summary, the depression score has a significant influence on predicting Parkinson's disease progression.

2.10. EM in AD vs. PD

The impairment of the oculomotor system in AD manifested with longer RS latency along with higher variability in accuracy and speed [43]. Yang et al., 2012, found similarities between three groups: AD patients, patients with amnestic mild cognitive impairment (aMCI), and healthy elderly subjects [43]. All groups showed shorter latencies in the gap tests (when there is a time delay between the disappearance of the fixation spot and the appearance of the light spot in the periphery) than in the overlap tests (when the above spots' appearance overlaps in time). However, in both tests, AD patients showed abnormally long saccade latencies. Although there was no significant difference in the accuracy (gain) and the velocity (both mean and peak velocity) between the three groups of subjects, AD patients showed an abnormally high coefficient of variation in the latency, accuracy, and speed of the reflexive saccades. There was a significant correlation between scores for the Mini-Mental State Examination (MMSE) and latencies of the saccades when comparing the MCI subjects to healthy elderly subjects [43].

Wilcockson et al. [44] explored AS eye movements in patients with amnestic and non-amnestic variants of MCI. There were 68 patients with dementia due to AD, 42 had amnestic MCI (aMCI), 47 had non-amnestic MCI (naMCI), and 92 were age-matched healthy controls (HC). The latencies for AS correction in the AD group were significantly longer than those for the HC and naMCI groups, but AS latencies in the AD group did not differ significantly from latencies in the aMCI group, even after age difference corrections [44]. They obtained similar results for the percentage of uncorrected AS errors. The AD and aMCI groups had similar and higher error rates than the naMCI and HC groups. This demonstrated that MCI patients are more likely to develop dementia due to AD than age-matched healthy adults. People with aMCI are at the highest risk of progressing to AD [45], and AS measurements might be an additional prognostic tool for predicting which people with MCI are more likely to progress to AD. It is worth noting that AS latency is a sensitive measure of the inhibitory process and is related to disease progression in the early stages of AD and PD.

In research by Pereira et al. [46], MCI sufferers were similarly impaired in their voluntary saccadic reaction times compared to AD sufferers, with a longer time to correct erroneous saccades.

Boxer et al. [47] compared saccade and antisaccade parameters in patients with frontotemporal dementia (FTD), patients with AD, and healthy subjects. The patients with AD showed an increased saccade latency compared to the FTD group during the horizontal saccade tasks. This might be related to the different dorsal parietal lobe roles in these two groups of patients [47]. In the AS task, all FTD and AD patients were impaired relative to the healthy subjects. The AD patients made fewer correct AS than controls, and they had more difficulty correcting saccade direction when they began from saccade instead of AS [46].

The relationships between AS parameters and measures of inhibitory control, attention, working memory, and self-monitoring showed correlations and common patterns reflecting deficits in executive function, confirming cognitive impairment in MCI and AD patients [46].

In Figure 3, we compared the latency of the reflexive saccades for normal subjects with those of AD and PD patients. These are averaged values for patients in different stages of the disease. However, latencies for AD and PD patients look similar and they are significantly longer than the mean RS latency for normal subjects.

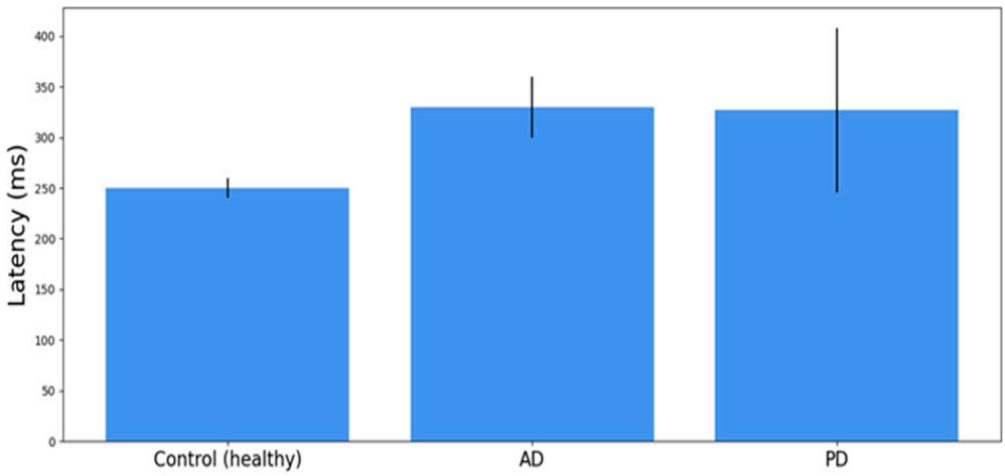

Figure 3. A comparison of the average results +/− SD of the AS latencies, obtained in a group of 27 age-matched normal controls (healthy people), 10 AD subjects, and 12 patients with PD. The averages of the control group and AD patients are derived from the research results obtained from [43] and the average of the PD patients from the results obtained from [11].

Three studies found a reduction in pursuit gain along with an increase in correction saccades in patients with Alzheimer's disease [48–50].

3. Further Research

The ultimate research goal is to identify unrecognized changes in the brain which can cause AD and PD. We think introducing the methods to a wider audience might enable the fulfillment of the following points:

- Results must be based on a broader control group.
- Tests must ensure repeatability and reproducibility in a non-experimental environment.
- Methods must be extended with new digital biomarkers that can be observed in a three-dimensional space.

Therefore, our research team aims to design, evaluate, and introduce modern methods of data aggregation based on online self-assessments and virtual reality environments.

Virtual Reality—A Research Opportunity

In recent years, advances in technology related to display, computation, and controllers have brought to the market new solutions that have changed digital content consumption, including virtual reality (VR) technology.

To classify this approach, Milgram et al. [51] introduced the term "virtuality continuum", relating to the mixture of classes of objects presented in any display situation. This representation describes a superset of the user's perception of the environment. The continuum starts from the authentic environments (consisting solely of natural objects) and ends in completely simulated, virtual environments. It includes different stages of representative forms, such as augmented reality (AR) and virtual reality (VR). These representations of the environment and the areas interpolated together comprise the term Extended Reality (XR). VR refers to devices that occlude the user's view of the physical world only, allowing sight of digitally rendered images. VR devices can mimic stereoscopic vision by presenting slightly different, separate images to both eyes. The main idea is to immerse the user in the virtual world by depriving external stimuli during a content presentation. VR devices are headsets that entirely cover the field of view (FoV) and project the image directly to both eyes.

The wide availability of devices and improving quality mean they are being used on an increasing scale. Therefore, VR technology could be the next gold standard in cognitive assessment. The need for new tools has emerged from criticism of the current cognitive screening tools because these tests often (30 of 50 classical screening tools) miss a visuospatial component, such as the Clock Drawing Test, the Cube Drawing Test, and the Intersecting Shapes Test [52]. Because visuospatial tasks demonstrate significant diagnostic and prognostic potential in AD [53], VR applications have great potential as an assessment tool in dementia [54]. One example is shown in Figure 4, where the standard executive function Trial B test is performed in 3D instead of 2D (on the paper). In this case, in addition to engaging executive processes related to 2D, the subject must also activate 3D orientation processes that often fail in ND (especially in AD).

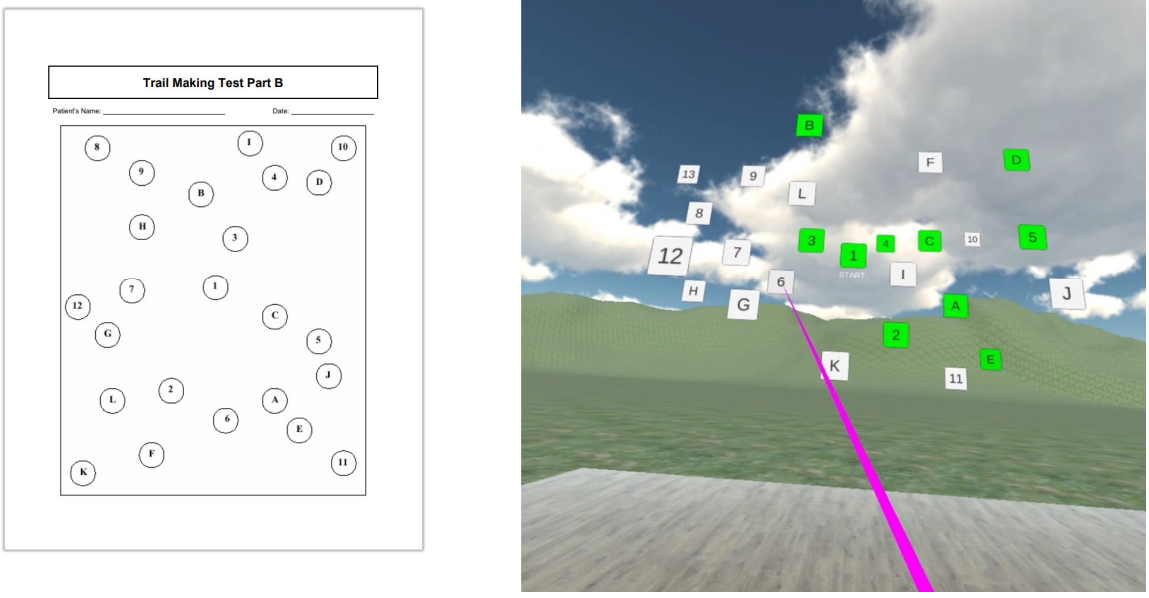

Figure 4. The prototype of an application testing the recognition of numbers and letters in the correct order implemented in a VR environment on an Oculus Go device. Trail B is generally thought to be sensitive to executive functioning since it requires a wide range of skills to complete. In 2D, it engages attention, memory, visual screening abilities, motor functioning, and cognitive processes, and it is likely to be even more difficult in 3D. Squares in green are already marked, and in white will be chosen in the proper order by the pointer in pink.

One of the main concerns of VR usage might be the physiological phenomenon called VR sickness (cybersickness). The most usual explanation is the sensory conflict theory [55]. According to this theory, VR sickness is caused by discrepancies in the sensory communication sent to the brain as the operator progresses through the virtual environment. Symptoms frequently reported include general distress, headache, eyestrain, stomach sensitivity, nausea, sweating, spite disorder (a.k.a. drowsiness), disorientation, and a nausea response. Symptoms can last from minutes to days post-exposure, with after-effects displaying as postural ataxia, visual displacement (e.g., altered vestibular-ocular reflex), and altered hand–eye harmonization, among other disorders. Cybersickness (CS) has been called the "elephant in the room" due to the possibility for it to radically limit VR equipment's uptake.

There are individual differences in susceptibility to VR sickness, and age is one of the factors. Brooks [56] presented that over-50s are more likely to experience virtual reality sickness than younger adults. Another factor is gender. Park [57] discovered that women are more vulnerable to virtual environments in terms of motion sickness (MS) or simulator sickness (SS). Furthermore, Kennedy [58] claimed that women are more susceptible than men, and the main reasons could be hormone differences and that women have a wider field of view. Depending on the immersive content, 20–95% of users typically experienced some form of cybersickness, ranging from a slight headache to an emetic response. This was not the conclusion in more recent research [59], which found no evidence that the incidence of motion sickness or the severity of motion sickness symptoms differed between the sexes.

These factors are common obstacles that impact the widespread use of technology, particularly among the elderly. However, with improving technology (higher resolution and frequencies), this problem seems likely to fade and should be eliminated in the future. Caserman [60] conducted a meta-analysis and compared different head-mounted displays (HMD) and stimuli. For example, Oculus Rift HMD vs. HTC Vive HMD and matched stimuli vs. unmatched stimuli. The meta-analysis results show that last-generation HMD devices have significantly fewer problems with CS, although they are still present. The findings reveal user experience (UX) flaws that could be obstacles in medical research. Thus, research group selection must be performed carefully and precisely until the technology can deal with cybersickness.

Immersive virtual environments allow researchers to create realistic environments while maintaining a high level of experimental control. For example, Garcia [61] suggests that it is possible to create experimental conditions where a virtual human modifies tone of voice while maintaining neutral facial expressions that would allow the study of the impact of tone of voice on persons with dementia. Furthermore, Flynn [62] presented findings that demonstrated that it is feasible to work in virtual environments with people with dementia. Bek et al. [63] found differences in eye gaze for emotional expressions which are static and dynamic. According to the researchers, PD may reduce the ability to utilize motion in emotion recognition, and eye movements reveal subtle effects of motion on emotion processing in PD. Researchers concluded that measuring eye gaze for moving faces enhances understanding of emotion recognition.

Additionally, as presented in [64,65], early detection of Alzheimer's disease can be supported by navigation tasks. This study used an immersive virtual reality paradigm in which participants walked through simulated environments to investigate path integration tasks. This study shows that a virtual reality navigation task can distinguish patients with moderate cognitive disabilities at low and high risk of developing dementia with better classification accuracy than classic cognitive measures. Virtual reality spatial cognition assessments have also been shown to be more sensitive than traditional visuospatial pencil-and-paper tests, such as the Mental Rotation Test, in detecting spatial navigation deficiencies [66]. Research papers also present a definite connection between virtual and real-world findings, for example, proof that wayfinding navigation performance on a

mobile app-based VR navigation task is closely correlated with real-world city street wayfinding performance [67].

A VR environment also provides vast possibilities for diminished curiosity research, which are behavioral changes that are extremely difficult to measure experimentally. A particular group of studies devoted to novel visual object perception (curiosity) associated with aging [68–71] presented that AD patients distributed their viewing time equally and spent significantly less time than controls looking at the novel (unpredictable) stimuli versus classical stimuli (in comparison to healthy subjects). Such novelties can be a horse that appears to have no hind legs or a lion that appears in a children's classroom, as in the classical study from 1992 [68]. Because a VR environment with proper hardware enables the simulation of real-world and synthetic objects with outstanding detail, one can automatically measure the subject's attention span based on eye movement registration in 3D space for prominent and less obvious examples of artifacts.

Moreover, Mandera [72] shared the opinion that virtual reality can be a supportive therapy for patients with MCI and various forms of dementia to improve adherence to cognitive training of older adults with cognitive impairment. This opinion is coherent with the outcome of the meta-analysis prepared by Kashif et al. [73]. Out of nine studies on motor function, six reported equal improvements in motor function compared to other groups. In addition, VR groups achieved superior results in improving static balance among patients with PD.

Our research group also evaluated the possible impact on early disease detection. We created a prototype of an application testing the recognition of numbers and letters in the correct order implemented in a VR environment. The test assumed that Trail B is generally sensitive to executive functioning since the test requires multiple abilities to complete it. Part B requires attention, memory, visual screening abilities, motor functioning, and cognitive processes in 2D, and we intuit that it is more difficult in 3D.

In the context of previous research, it is worth noting that there are VR headsets with eye tracking on the market, such as the HTC VIVE Pro Eye. Hence, they can be used as standalone research environments that allow us to control the experiment remotely, without on-site supervision. Furthermore, they add a spatial dimension to our research, connected to the motor reaction of the eyes combined with actions executed by the subject's hands.

4. Discussion

In this review, we have demonstrated that the parameters of reflexive eye movement are significant in estimating Parkinson's and Alzheimer's disease progressions. Therefore, they might also be good biomarkers in the preclinical stages of these ND diseases. Various saccade abnormalities were found in Parkinson's [74–77] and Alzheimer's diseases, e.g., in reviews [78] and using computational attention models under realistic scenarios for AD [79]. Even if many authors have demonstrated EM pathologies in AD and PD, they did not demonstrate how we can use EM parameters to predict the disease progression of many different patients or even patients with different treatments and symptoms. However, in [80], the authors used LR (logistic regression), SVM (support vector machine), and NB (Naive Bayes) algorithms to classify normal (NC), MCI, and AD subjects based on novelty preference (NP), pupil diameter (PD), saccade orientation (SO), and re-fixation (RE) and fixation duration (FD) differences between watching a familiar or a novel image (see discussion above related to [68–71]). The division into NC (n = 30 subjects), MCI (n = 10), and AD (n = 20) was assessed by clinicians based on standard assessments and neuropsychological tests [80]. The authors used a cross-validation method on 20 NC and all AD subjects to determine classification algorithms that distinguish between AD and NC. In the next step, they used this algorithm to distinguish NC from MCI. They repeated the classification process 100 times by changing partitions of NC subjects and, each time, testing different NC subjects against all MIC patients and averaging all results. The best results were obtained for all attributes: NP + PD + SO + RE + FD. For the SVM algorithm,

accuracy was 0.87 and sensitivity 0.97. These are great results, giving the basis for our proposal for VR testing.

In [81], the authors used oculomotor behavior to differentiate diagnoses between normal subjects, AD sufferers, and behavior variants of frontotemporal dementia (bvFTD) and a semantic variant of primary progressive aphasia (svPPA) groups. They tested RS, AS, and pursuit EM. By comparing RS latency, AS success (successfully performed AS), and spatial accuracies of RS, AS, and pursuit, the authors found that AD patients performed the worst in all these tests. Additionally, the mean MMSE was 16.7 +/− 5.2 and was the lowest for AD compared to other patients. MMSE = 15 and below signifies the probability of total impairment, and there are such patients in this group. SVM and k-Nearest Neighbors (k-NN) algorithms were used to classify AD vs. control, bvFTD vs. control, and AD v. bvFTD, and they obtained a mean accuracy above 0.92. The svPPA group was too small for the ML procedures [81].

As we have described above (see Figure 3), there are many similarities in the properties of eye movements between AD and PD. However, in our projects, we have not only predicted the PD stage (UPDRS) from reflexive EM and neuropsychological test results, but also tried to predict the progression of the disease (longitudinal studies). We have also demonstrated that we can take a group of more advanced Parkinson's disease patients as a "Model". This approach allows us to see PD progression as patients' symptoms change relative to the model, which means that the multidimensional sets of their values (granules) become more similar to the set (granule) of the model. This approach allows us to compare different rates of various patients' disease development and relate the effectiveness of different treatments (our different groups of PD: BMT, DBS, and POP).

Another critical issue is related to disease progression and related changes in different parts of the brain or symptoms related to the motor/cognitive and emotional systems [41]. We have estimated PD progression by all of our (general) attributes, by only motor-related attributes, and by only cognitive symptoms. The general and motor attributes predicted disease progression (UPDRS changes) well, but the correlation with cognitive attributes was much weaker [41]. The cognitive attributes do not change with disease progression in all PD patients, and this is the opposite of Alzheimer's disease, which is mainly related to cognitive changes (MMSE changes).

We have demonstrated for PD patients that the depression score (measured by Beck's depression inventory) is an essential attribute in the estimation of disease progression, and its value significantly increases the accuracy of UPDRS estimation [42]. In Alzheimer's disease, depression is a significant factor, especially for older subjects with late onset of AD (LOAD) and for patients over 65 years of age [82]. Another essential function related to preclinical AD relates to motor symptoms, and they can even predict MCI [83]. They are related to muscle strength decrease and deficient grip strength [83]. Physical frailty, gait and balance problems, and loss of other motor functions can all precede cognitive impairment by several years. Even the trajectory of gait speed can precede MCI by 12 years [84,85]. Therefore, experiments in VR using familiar and novelty objects could also test the subject's motor abilities and responses to different emotions.

There are two different ML classification approaches for predicting ND symptoms. The first one, more clinically oriented, is that based on neuropsychological and clinical tests, experienced neurologists decide each patient's disease stage: (1) in Alzheimer's disease—normal, MCI, dementia; (2) in Parkinson's disease—normal, early-stage, medium, or advanced, or they can use the Hoehn and Yahr scale. In the next step, different parameters of EM, with the help of ML methods, attempt to classify patients following the doctors' findings. In [80], the authors used five parameters related to novelty preference to differentiate normal and MCI subjects based on training normal subjects and AD patients. There were also five parameters related to RS, AS, and pursuit EM used to differentiate no disease from AD and AD from FTD (frontotemporal dementia) [81]. Our approach was different, as we took all eight neuropsychological and clinical parameters and four EM attributes together with ten to twelve parameters (in different studies) to classify four different ranges

of PD stages. An even higher dimension of parameters was used in the BIBIOCARD study, where there were 181 condition parameters used to classify four stages of the disease—normal = 1, impaired and not MCI = 2, MCI = 3, dementia = 4—in a longitudinal study lasting over twenty-five years [86]. However, they have only used nine attributes, namely age, education, two cognitive testing results, two MRI scan results, two cerebrospinal fluid parameters, and APOE genotype, to predict the probability of progression from normal to MCI in the next 5 years. We have used their 11 cognitive test results and APOE genotype to determine, with the granular computing approach, that some of their normal patients might already have mild/very mild dementia or questionable impairment [87].

5. Conclusions

Alzheimer's and Parkinson's diseases are two of the most common neurodegenerative age-dependent diseases for which, despite many years of intensive research, we still do not have a cure. Both diseases have complex etiology and many years of hidden, non-reversible neurodegenerative processes (ND) with devastating effects on the brain. When the first symptoms appear, a large part of the brain has already disappeared. One possible solution is to find the neurodegenerative processes early enough to test possible methods to slow them down or even cure them.

We have described many papers showing that eye movement parameters change with disease progression and that they are also sensitive to the early stages of both diseases. However, the parameter space describing different measured attributes and methods has a very high dimension. To find significant subspace(s) with parameters sensitive to disease progression, we have proposed the use of different AI (machine learning) algorithms. These algorithms describe disease symptoms more precisely than the standard approach and can also predict disease development. We have given several examples of such AI (ML) methods and have demonstrated their effectiveness for the most common neurodegenerative diseases.

Therefore, the major differences between classical (statistical) and AI approaches are not only opportunities to reduce the dimension of the parameter space, but also to ask diverse questions about the nature of the diseases. For example, as mentioned above, many authors found pathological saccade parameters in AD and PD, but using the AI approach, we ask a different question: can EM parameters predict the AD/PD progression in different individual patients having different treatments, and in diverse disease stages? Responses to this question are related to the mechanisms of the disease.

There is a related difference between subject testing in real and virtual worlds. In the real environment, we can find many differences in behaviors such as EM, emotions, or peripheral body movements between ND and control subjects. We cannot control significant features of the real world and determine how changes in this environment may influence the subject. However, in virtual reality, we can ask the question: how do different features in the surroundings influence the individual subject? Again, finding isolated elements in the surroundings that, in a unique way, influence the behavior of the individual subjects gives insight into ND mechanisms.

We think that the next research step should be into virtual reality. VR offers tremendous possibilities in the experimental environment. Researchers can simulate virtually every scenario in a strictly controlled environment where interactions can be controlled in real time. Furthermore, the patient is always safe because the experiment occurs in the laboratory or physician's office. Development environments, such as Unity 3D, are offered free of charge and enable the design and development of VR applications. Therefore, virtual reality applications can become a valuable tool for dementia assessment if only more interdisciplinary teams, combining health professionals and computer scientists, would pave the way to new applications in this area. Much research has been done already, and more research is required on the scale of its usability for patients.

Author Contributions: Every author has written different parts of this review; A.Ś. and A.C. made figures; S.S. and D.K. supplied clinical data related to this review, A.W.P. supervised this article. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: All studies were conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Department of Neurology, Medical University Warsaw.

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

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Younes, L.; Albert, M.; Moghekar, A.; Soldan, A.; Pettigrew, C.; Miller, M.I. Identifying Changepoints in Biomarkers During the Preclinical Phase of Alzheimer's Disease. *Front. Aging Neurosci.* **2019**, *11*, 74. [CrossRef] [PubMed]
2. Savica, R.; Rocca, W.A.; Ahlskog, J.E. When does Parkinson's disease start? *Arch. Neurol.* **2010**, *67*, 798–801. [CrossRef] [PubMed]
3. Reijnders, J.S.A.M.; Ehrt, U.; Weber, W.E.J.; Aarsland, D.; Leentjens, A.F.G. A systematic review of prevalence studies of depression in Parkinson's disease. *Mov. Disord.* **2007**, *23*, 183–189. [CrossRef] [PubMed]
4. Haaksma, M.L.; Vilela, L.R.; Marengoni, A.; Calderón-Larrañaga, A.; Leoutsakos, J.-M.S.; Rikkert, M.G.M.O.; Melis, R.J.F. Comorbidity and progression of late onset Alzheimer's disease: A systematic review. *PLoS ONE* **2017**, *12*, e0177044. [CrossRef] [PubMed]
5. Moustafa, A.A.; Sherman, S.J.; Frank, M.J. A dopaminergic basis for working memory, learning and attentional shifting in Parkinsonism. *Neuropsychologia* **2008**, *46*, 3144–3156. [CrossRef]
6. Foley, J.A.; Lancaster, C.; Poznyak, E.; Borejko, O.; Niven, E.; Foltynie, T.; Abrahams, S.; Cipolotti, L. Impairment in Theory of Mind in Parkinson's Disease Is Explained by Deficits in Inhibition. *Park. Dis.* **2019**, *2019*, 5480913. [CrossRef]
7. Przybyszewski, A.W. Theory of mind helps to predict neurodegenerative processes in Parkinson's disease In Proceedings of the International Conference on Computational Science. Krakow, Poland, 16–18 June 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 542–555.
8. Briand, K.A.; Strallow, D.; Hening, W.; Poizner, H.; Sereno, A.B. Control of voluntary and reflexive saccades in Parkinson's disease. *Exp. Brain Res.* **1999**, *129*, 38–48. [CrossRef]
9. Chambers, J.M.; Prescott, T.J. Response times for visually guided saccades in persons with Parkinson's disease: A meta-analytic review. *Neuropsychologia* **2010**, *48*, 887–899. [CrossRef]
10. Przybyszewski, A.W.; Kon, M.; Szlufik, S.; Szymanski, A.; Habela, P.; Koziorowski, D.M. Multimodal Learning and Intelligent Prediction of Symptom Development in Individual Parkinson's Patients. *Sensors* **2016**, *16*, 1498. [CrossRef]
11. Śledzianowski, A.; Szymanski, A.; Drabik, A.; Szlufik, S.; Koziorowski, D.; Przybyszewski, A.W. Combining results of different oculometric tests improved prediction of Parkinson's disease development. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Phuket, Thailand, 23–26 March 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 517–526.
12. Turner, T.H.; Renfroe, J.B.; Duppstadt-Delambo, A.; Hinson, V.K. Validation of a Behavioral Approach for Measuring Saccades in Parkinson's Disease. *J. Mot. Behav.* **2017**, *49*, 657–667. [CrossRef]
13. Stuart, S.; Lawson, R.A.; Yarnall, A.J.; Nell, J.; Alcock, L.; Duncan, G.W.; Khoo, T.K.; Barker, R.; Rochester, L.; Burn, D.J.; et al. Pro-Saccades Predict Cognitive Decline in Parkinson's Disease: ICICLE-PD. *Mov. Disord.* **2019**, *34*, 1690–1698. [CrossRef]
14. Perneczky, R.; Ghosh, B.; Hughes, L.; Carpenter, R.; Barker, R.; Rowe, J. Saccadic latency in Parkinson's disease correlates with executive function and brain atrophy, but not motor severity. *Neurobiol. Dis.* **2011**, *43*, 79–85. [CrossRef]
15. Antoniades, C.A.; Xu, Z.; Carpenter, R.; Barker, R. The relationship between abnormalities of saccadic and manual response times in parkin- son's disease. *J. Park. Dis.* **2013**, *3*, 557–563.
16. Abasi, A.; Hoseinabadi, R.; Raji, P.; Friedman, J.H.; Hadian, M.-R. Evaluating Oculomotor Tests before and after Vestibular Rehabilitation in Patients with Parkinson's Disease: A Pilot Pre-Post Study. *Park. Dis.* **2022**, *2022*, 6913691. [CrossRef]
17. Wong, O.W.; Fung, G.; Chan, S. Characterizing the relationship between eye movement parameters and cognitive functions in non-demented Parkinson's disease patients with eye tracking. *JoVE (J. Vis. Exp.)* **2019**, *151*, e60052.
18. Wong, O.W.; Chan, A.Y.; Wong, A.; Lau, C.K.; Yeung, J.H.; Mok, V.C.; Lam, L.C.; Chan, S. Eye movement parameters and cognitive functions in Parkinson's disease patients without dementia. *Park. Relat. Disord.* **2018**, *52*, 43–48. [CrossRef]
19. Archibald, N.K.; Hutton, S.B.; Clarke, M.P.; Mosimann, U.P.; Burn, D.J. Visual exploration in Parkinson's disease and Parkinson's disease dementia. *Brain* **2013**, *136*, 739–750. [CrossRef]
20. Everling, S.; Fischer, B. The antisaccade: A review of basic research and clinical studies. *Neuropsychologia* **1998**, *36*, 885–899. [CrossRef]

21. Sledzianowski, A.; Szymanski, A.; Drabik, A.; Szlufik, S.; Koziorowski, D.M.; Przybyszewski, A.W. Measurements of antisaccades parameters can improve the prediction of Parkinson's disease progression. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Yogyakarta, Indonesia, 8–11 April 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 602–614.
22. Waldthaler, J.; Stock, L.; Student, J.; Sommerkorn, J.; Dowiasch, S.; Timmermann, L. Antisaccades in Parkinson's Disease: A Meta-Analysis. *Neuropsychol. Rev.* **2021**, *31*, 628–642. [CrossRef]
23. Waldthaler, J.; Stock, L.; Krüger-Zechlin, C.; Timmermann, L. Age at Parkinson's disease onset modulates the effect of levodopa on response inhibition: Support for the dopamine overdose hypothesis from the antisaccade task. *Neuropsychologia* **2021**, *163*, 108082. [CrossRef]
24. Waldthaler, J.; Stock, L.; Sommerkorn, J.; Krüger-Zechlin, C.; Timmermann, L. Antisaccade Latency Is Sensitive to Longitudinal Change of Motor and Cognitive Symptoms in Parkinson's Disease. *Mov. Disord.* **2020**, *36*, 266–268. [CrossRef] [PubMed]
25. Antoniades, C.A.; Demeyere, N.; Kennard, C.; Humphreys, G.W.; Hu, M.T. Antisaccades and executive dysfunction in early drug-naive Parkinson's disease: The discovery study. *Mov. Disord.* **2015**, *30*, 843–847. [CrossRef] [PubMed]
26. Fooken, J.; Patel, P.; Jones, C.B.; McKeown, M.J.; Spering, M. Preservation of Eye Movements in Parkinson's Disease Is Stimulus- and Task-Specific. *J. Neurosci.* **2021**, *42*, 487–499. [CrossRef] [PubMed]
27. Koçoğlu, K.; Akdal, G.; Çolakoğlu, B.D.; Çakmur, R.; Sharma, J.C.; Ezard, G.; Hermens, F.; Hodgson, T.L. The effect of directional social cues on saccadic eye movements in Parkinson's disease. *Exp. Brain Res.* **2021**, *239*, 2063–2075. [CrossRef] [PubMed]
28. Munoz, M.J.; Goelz, L.C.; Pal, G.D.; Karl, J.A.; Metman, L.V.; Sani, S.; Rosenow, J.M.; Ciolino, J.D.; Kurani, A.S.; Corcos, D.M.; et al. Increased Subthalamic Nucleus Deep Brain Stimulation Amplitude Impairs Inhibitory Control of Eye Movements in Parkinson's Disease. *Neuromodul. Technol. Neural Interface* **2022**, *25*, 866–876. [CrossRef]
29. Przybyszewski, A.W.; Szlufik, S.; Dutkiewicz, J.; Habela, P.; Koziorowski, D.M. Machine learning on the video basis of slow pursuit eye movements can predict symptom development in Parkinson's patients. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Bali, Indonesia, 23–25 March 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 268–276.
30. Śledzianowski, A.; Szymański, A.; Szlufik, S.; Koziorowski, D. Rough set data mining algorithms and pursuit eye movement measurements help to predict symptom development in Parkinson's disease. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Dong Hoi City, Vietnam, 19–21 March 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 428–435.
31. Frei, K. Abnormalities of smooth pursuit in Parkinson's disease: A systematic review. *Clin. Park. Relat. Disord.* **2020**, *4*, 100085. [CrossRef]
32. MacAskill, M.R.; Graham, C.F.; Pitcher, T.L.; Myall, D.J.; Livingston, L.; van Stockum, S.; Dalrymple-Alford, J.C.; Anderson, T.J. The influence of motor and cognitive impairment upon visually-guided saccades in Parkinson's disease. *Neuropsychologia* **2012**, *50*, 3338–3347. [CrossRef]
33. Farashi, S. Analysis of vertical eye movements in Parkinson's disease and its potential for diagnosis. *Appl. Intell.* **2021**, *51*, 8260–8270. [CrossRef]
34. Tabashum, T.; Zaffer, A.; Yousefzai, R.; Colletta, K.; Jost, M.B.; Park, Y.; Chawla, J.; Gaynes, B.; Albert, M.V.; Xiao, T. Detection of Parkinson's Disease Through Automated Pupil Tracking of the Post-illumination Pupillary Response. *Front. Med.* **2021**, *8*, 645293. [CrossRef]
35. Tsitsi, P.; Benfatto, M.N.; Seimyr, G.; Larsson, O.; Svenningsson, P.; Markaki, I. Fixation Duration and Pupil Size as Diagnostic Tools in Parkinson's Disease. *J. Park. Dis.* **2021**, *11*, 865–875.
36. Bonnet, C.T.; Delval, A.; Singh, T.; Defebvre, L. Parkinson's disease-related changes in the behavioral synergy between eye movements and postural movements. *Eur. J. Neurosci.* **2021**, *54*, 5161–5172. [CrossRef]
37. Zhang, J.; Zhang, B.; Ren, Q.; Zhong, Q.; Li, Y.; Liu, G.; Ma, X.; Zhao, C. Eye movement especially vertical oculomotor impairment as an aid to assess Parkinson's disease. *Neurol. Sci.* **2020**, *42*, 2337–2345. [CrossRef]
38. Perkins, J.E.; Janzen, A.; Bernhard, F.P.; Wilhelm, K.; Brien, D.C.; Huang, J.; Coe, B.C.; Vadasz, D.; Mayer, G.; Munoz, D.P.; et al. Saccade, Pupil, and Blink Responses in Rapid Eye Movement Sleep Behavior Disorder. *Mov. Disord.* **2021**, *36*, 1720–1726. [CrossRef]
39. Chudzik, A.; Szymański, A.; Nowacki, J.P.; Przybyszewski, A.W. Eye-tracking and machine learning significance in Parkinson's disease symptoms prediction. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Phuket, Thailand, 23–26 March 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 537–547.
40. 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. Informaticae* **2020**, *176*, 167–181. [CrossRef]
41. 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 Proceedings of the Asian Conference on Intelligent Information and Database Systems, Phuket, Thailand, 23–26 March 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 548–559.
42. Przybyszewski, A.W.; Nowacki, J.P.; Drabik, A.; Szlufik, S.; Habela, P.; Koziorowski, D.M. Granular computing (GC) demonstrates interactions between depression and symptoms development in Parkinson's disease patients. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Yogyakarta, Indonesia, 8–11 April 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 591–601.

43. Yang, Q.; Wang, T.; Su, N.; Xiao, S.; Kapoula, Z. Specific saccade deficits in patients with Alzheimer's disease at mild to moderate stage and in patients with amnestic mild cognitive impairment. *Age* **2012**, *35*, 1287–1298. [CrossRef]
44. Wilcockson, T.D.; Mardanbegi, D.; Xia, B.; Taylor, S.; Sawyer, P.; Gellersen, H.W.; Leroi, I.; Killick, R.; Crawford, T.J. Abnormalities of saccadic eye movements in dementia due to Alzheimer's disease and mild cognitive impairment. *Aging* **2019**, *11*, 5389–5398. [CrossRef]
45. Petersen, R.C.; Doody, R.; Kurz, A.; Mohs, R.; Morris, J.; Rabins, P.; Ritchie, K.; Rossor, M.; Thal, L.; Winblad, B. Current concepts in mild cognitive impairment. *Arch. Neurol.* **2001**, *58*, 1985–1992.
46. Pereira, M.L.G.F.; Villa, M.; Koh, D.; Camargo, M.Z.A.; Belan, A.; Radanovic, M.; Pomplun, M.; Forlenza, O. Saccadic eye movements associated with executive function decline in mild cognitive impairment and Alzheimer's disease: Biomarkers (non-neuroimaging)/novel biomarkers. *Alzheimer's Dement.* **2020**, *16*, e040036. [CrossRef]
47. Boxer, A.L.; Garbutt, S.; Seeley, W.; Jafari, A.; Heuer, H.; Mirsky, J.; Hellmuth, J.; Trojanowski, J.; Huang, E.; DeArmond, S.; et al. Saccade abnormalities in autopsy-confirmed frontotemporal lobar degeneration and Alzheimer disease. *Arch. Neurol.* **2012**, *69*, 509–517.
48. Hutton, J.T.; Nagel, J.A.; Loewenson, R.B. Eye tracking dysfunction in Alzheimer-type dementia. *Neurology* **1984**, *34*, 99. [CrossRef]
49. Fletcher, W.A.; Sharpe, J.A. Smooth pursuit dysfunction in Alzheimer's disease. *Neurology* **1988**, *38*, 272. [CrossRef] [PubMed]
50. Kuskowski, M.A.; Malone, S.; Mortimer, J.; Dysken, M. Smooth pursuit eye movements in dementia of the Alzheimer-type. *Alzheimer Dis. Assoc. Disord.* **1989**, *3*, 157–171. [CrossRef] [PubMed]
51. Milgram, P.; Kishino, F. A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* **1994**, *77*, 1321–1329.
52. De Roeck, E.E.; De Deyn, P.P.; Dierckx, E.; Engelborghs, S. Brief cognitive screening instruments for early detection of Alzheimer's disease: A systematic review. *Alzheimer's Res. Ther.* **2019**, *11*, 21. [CrossRef] [PubMed]
53. Salimi, S.; Irish, M.; Foxe, D.; Hodges, J.; Piguet, O.; Burrell, J. Can visuospatial measures improve the diagnosis of Alzheimer's disease? *Alzheimer's Dement. Diagn. Assess. Dis. Monit.* **2018**, *10*, 66–74. [CrossRef]
54. Persky, S.; McBride, C.M. Immersive Virtual Environment Technology: A Promising Tool for Future Social and Behavioral Genomics Research and Practice. *Health Commun.* **2009**, *24*, 677–682. [CrossRef]
55. Reason, J.T. Motion sickness adaptation: A neural mismatch model. *J. R. Soc. Med.* **1978**, *71*, 819–829. [CrossRef]
56. Brooks, J.O.; Goodenough, R.R.; Crisler, M.C.; Klein, N.D.; Alley, R.L.; Koon, B.L.; Logan, W.C.; Ogle, J.H.; Tyrrell, R.A.; Wills, R.F. Simulator sickness during driving simulation studies. *Accid. Anal. Prev.* **2010**, *42*, 788–796. [CrossRef]
57. Park, G.D.; Allen, R.; Fiorentino, D.; Rosenthal, T.; Cook, M. Simulator sickness scores according to symptom susceptibility, age, and gender for an older driver assessment study. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2006**, *50*, 2702–2706. [CrossRef]
58. Kennedy, R.S.; Frank, L.H. A review of motion sickness with special reference to simulator sickness. In Proceedings of the National Academy of Science, Workshop on Simulator Sickness, Monterey, CA, USA, 26–28 September 1983.
59. Curry, C.; Li, R.; Peterson, N.; Stoffregen, T.A. Cybersickness in Virtual Reality Head-Mounted Displays: Examining the Influence of Sex Differences and Vehicle Control. *Int. J. Hum. Comput. Interact.* **2020**, *36*, 1161–1167. [CrossRef]
60. Caserman, P.; Garcia-Agundez, A.; Zerban, A.G.; Göbel, S. Cybersickness in current-generation virtual reality head-mounted displays: Systematic review and outlook. *Virtual Real.* **2021**, *25*, 1153–1170. [CrossRef]
61. Kartolo, A.; Methot-Curtis, E. A discussion of the use of virtual reality in dementia. *Virtual Real. Psychol. Med. Pedagog. Appl.* **2012**, 123–136. [CrossRef]
62. Flynn, D.; van Schaik, P.; Blackman, T.; Femcott, C.; Hobbs, B.; Calderon, C. Developing a Virtual Reality–Based Methodology for People with Dementia: A Feasibility Study. *CyberPsychology Behav.* **2003**, *6*, 591–611. [CrossRef]
63. Bek, J.; Poliakoff, E.; Lander, K. Measuring emotion recognition by people with Parkinson's disease using eye-tracking with dynamic facial expressions. *J. Neurosci. Methods* **2020**, *331*, 108524. [CrossRef]
64. Howett, D.; Castegnaro, A.; Krzywicka, K.; Hagman, J.; Marchment, D.; Henson, R.; Rio, M.; King, J.; Burgess, N.; Chan, D. Differentiation of mild cognitive impairment using an entorhinal cortex-based test of virtual reality navigation. *Brain* **2019**, *142*, 1751–1766. [CrossRef]
65. Allison, S.L.; Fagan, A.M.; Morris, J.C.; Head, D. Spatial Navigation in Preclinical Alzheimer's Disease. *J. Alzheimer's Dis.* **2016**, *52*, 77–90. [CrossRef]
66. Mitolo, M.; Gardini, S.; Caffarra, P.; Ronconi, L.; Venneri, A.; Pazzaglia, F. Relationship between spatial ability, visuospatial working memory and self-assessed spatial orientation ability: A study in older adults. *Cogn. Process.* **2015**, *16*, 165–176. [CrossRef]
67. Coutrot, A.; Schmidt, S.; Coutrot, L.; Pittman, J.; Hong, L.; Wiener, J.M.; Hölscher, C.; Dalton, R.C.; Hornberger, M.; Spiers, H.J. Virtual navigation tested on a mobile app is predictive of real-world wayfinding navigation performance. *PLoS ONE* **2019**, *14*, e0213272. [CrossRef]
68. Daffner, K.R.; Scinto, L.; Weintraub, S.; Guinessey, J.E.; Mesulam, M.M. Diminished curiosity in patients with probable Alzheimer's disease as measured by exploratory eye movements. *Neurology* **1992**, *42*, 320. [CrossRef]
69. Daffner, K.R.; Scinto, L.F.M.; Weintraub, S.; Guinessey, J.; Mesulam, M.-M. The Impact of Aging on Curiosity as Measured by Exploratory Eye Movements. *Arch. Neurol.* **1994**, *51*, 368–376. [CrossRef]
70. Daffner, K.R.; Mesulam, M.M.; Cohen, L.G.; Scinto, L.F. Mechanisms underlying diminished novelty-seeking behavior in patients with probable Alzheimer's disease. *Neuropsychiatry Neuropsychol. Behav. Neurol.* **1999**, *12*, 58–66. [PubMed]
71. Schacter, D.L.; Cooper, L.; Valdiserri, M. Implicit and explicit memory for novel visual objects in older and younger adults. *Psychol. Aging* **1992**, *7*, 299. [CrossRef] [PubMed]

72. Manera, V.; Chapoulie, E.; Bourgeois, J.; Guerchouche, R.; David, R.; Ondrej, J.; Drettakis, G.; Robert, P. A Feasibility Study with Image-Based Rendered Virtual Reality in Patients with Mild Cognitive Impairment and Dementia. *PLoS ONE* **2016**, *11*, e0151487. [CrossRef] [PubMed]
73. Kashif, M.; Ahmad, A.; Bandpei, M.A.M.; Farooq, M.; Iram, H.; e Fatima, R. Systematic review of the application of virtual reality to improve balance, gait and motor function in patients with Parkinson's disease. *Medicine* **2022**, *101*, e29212. [CrossRef]
74. Rottach, K.G.; Riley, D.; DiScenna, A.; Zivotofsky, A.; Leigh, R. Dynamic properties of horizontal and vertical eye movements in parkinsonian syndromes. *Ann. Neurol. Off. J. Am. Neurol. Assoc. Child Neurol. Soc.* **1996**, *39*, 368–377. [CrossRef]
75. Lueck, C.; Tanyeri, S.; Crawford, T.; Henderson, L.; Kennard, C. Anti-saccades and remembered saccades in Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **1990**, *53*, 284–288. [CrossRef]
76. Mosimann, U.P.; Müri, R.M.; Burn, D.; Felblinger, J.; O'Brien, J.; McKeith, I.G. Saccadic eye movement changes in Parkinson's disease dementia and dementia with Lewy bodies. *Brain* **2005**, *128*, 1267–1276. [CrossRef]
77. Pretegiani, E.; Optican, L.M. Eye Movements in Parkinson's Disease and Inherited Parkinsonian Syndromes. *Front. Neurol.* **2017**, *8*, 592. [CrossRef]
78. Molitor, R.J.; Ko, P.; Ally, B. Eye movements in Alzheimer's disease. *J. Alzheimer's Dis.* **2015**, *44*, 1–12. [CrossRef]
79. Beltrán, J.; García-Vázquez, M.S.; Benois-Pineau, J.; Gutierrez-Robledo, L.M.; Dartigues, J.-F. Computational Techniques for Eye Movements Analysis towards Supporting Early Diagnosis of Alzheimer's Disease: A Review. *Comput. Math. Methods Med.* **2018**, *2018*, 2676409. [CrossRef]
80. Lagun, D.; Manzanares, C.; Zola, S.M.; Buffalo, E.A.; Agichtein, E. Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *J. Neurosci. Methods* **2011**, *201*, 196–203. [CrossRef]
81. Lage, C.; López-García, S.; Bejanin, A.; Kazimierczak, M.; Aracil-Bolaños, I.; Calvo-Córdoba, A.; Pozueta, A.; García-Martínez, M.; Fernández-Rodríguez, A.; Bravo-González, M.; et al. Distinctive Oculomotor Behaviors in Alzheimer's Disease and Frontotemporal Dementia. *Front. Aging Neurosci.* **2021**, *12*, 525. [CrossRef]
82. Bauman, J.; Gibbons, L.; Moore, M.; Mukherjee, S.; McCurry, S.; McCormick, W.; Bowen, J.; Trittschuh, E.; Glymour, M.; Mez, J.; et al. Associations between depression, traumatic brain injury, and cognitively-defined late-onset Alzheimer's disease subgroups. *J. Alzheimer's Dis.* **2019**, *70*, 611–619. [CrossRef]
83. Buchman, A.S.; Bennett, D.A. Loss of motor function in preclinical Alzheimer's disease. *Expert Rev. Neurother.* **2011**, *11*, 665–676. [CrossRef]
84. Buracchio, T.; Dodge, H.H.; Howieson, D.B.; Wasserman, D.; Kaye, J. The Trajectory of Gait Speed Preceding Mild Cognitive Impairment. *Arch. Neurol.* **2010**, *67*, 980–986. [CrossRef]
85. Watson, N.; Rosano, C.; Boudreau, R.; Simonsick, E.; Ferrucci, L.; Hardy, S.; Atkinson, H.; Yaffe, K.; Satterfield, S.; Harris, T.B.; et al. Executive function, memory, and gait speed decline in well-functioning older adults. *J. Gerontol. Ser. A Biomed. Sci. Med. Sci.* **2010**, *65*, 1093–1100. [CrossRef]
86. Albert, M.; Zhu, Y.; Moghekar, A.; Mori, S.; Miller, M.I.; Soldan, A.; Pettigrew, C.; Selnes, O.; Li, S.; Wang, M.-C. Predicting progression from normal cognition to mild cognitive impairment for individuals at 5 years. *Brain* **2018**, *141*, 877–887. [CrossRef]
87. Przybyszewski, A.W.; BIOCARD Study Team. *AI Classifications Applied to Neuropsychological Trials in Normal Individuals That Predict Progression to Cognitive Decline*; Groen, D., de Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M.A., Eds.; ICCS 2022, LNCS 13352; Springer: Cham, Switzerland, 2022; pp. 150–156. [CrossRef]

Fundamenta Informaticae XXI (2001) 1001–1016
DOI 10.3233/FI-2016-0000
IOS Press

Comparison of different data mining methods to determine disease progression in dissimilar groups of Parkinson's patients

Andrzej W. Przybyszewski
Polish–Japanese Academy of Information Technology
Koszykowa 86, 00-097 Warszawa, Poland
Dept. Neurology, UMass Medical School
55 Lake Av. Worcester, MA 02135, USA
przy@pjwstk.edu.pl

Artur Chudzik
Polish–Japanese Academy of Information Technology
Koszykowa 86
00-097 Warszawa, Poland
artur.chudzik@pjwstk.edu.pl

Stanislaw Szlufik
Department of Neurology, Faculty of Health Science,
Medical University
03-242 Warszawa, Poland
stanislaw.szlufik@gmail.com

Piotr Habela
Polish–Japanese Academy of Information Technology
Koszykowa 86
00-097 Warszawa, Poland
piotr.habela@pjwstk.edu.pl

Dariusz M. Koziorowski
Department of Neurology, Faculty of Health Science,
Medical University
Kondratowicza 8
03-242 Warszawa, Poland
dkoziorowski@esculap.pl

Abstract. Parkinson's disease (PD) is the second after Alzheimer's most popular neurodegenerative disease (ND). Cures for both NDs are currently unavailable. OBJECTIVE: The purpose of our study was to predict the results of different PD patients' treatments in order to find an optimal one. METHODS: We have compared rough sets (RS) and others, in short, machine learning (ML) models to describe and predict disease progression expressed as UPDRS values (Unified Parkinson's Disease Rating Scale) in three groups of Parkinson's patients: 23 BMT (Best Medical Treatment) patients on medication; 24 DBS patients on medication and on DBS therapy (Deep Brain Stimulation) after surgery performed during our study; and 15 POP (Postoperative) patients

Address for correspondence: Andrzej Przybyszewski, PJIIT, Koszykowa 86, 00-097 Warszawa, Poland

who had had surgery earlier (before the beginning of our research). Every PD patient had three visits approximately every six months. The first visit for DBS patients was before surgery. On the basis of the following condition attributes: disease duration, saccadic eye movement parameters, and neuropsychological tests: PDQ39 (Parkinson's Disease Questionnaire - disease-specific health-related quality-of-life questionnaire), and Epworth Sleepiness Scale tests we have estimated UPDRS changes (as the decision attribute). RESULTS: By means of RS rules obtained for the first visit of BMT/DBS/POP patients, we have predicted UPDRS values in the following year (two visits) with global accuracy of 70% for both BMT visits; 56% for DBS, and 67, 79% for POP second and third visits. The accuracy obtained by ML models was generally in the same range, but it was calculated separately for different sessions (MedOFF/MedON). We have used RS rules obtained in BMT patients to predict UPDRS of DBS patients; for the first session DBSW1: global accuracy was 64%, for the second DBSW2: 85% and the third DBSW3: 74% but only for DBS patients during stimulation-ON. ML models gave better accuracy for DBSW1/W2 session S1(MedOFF): 88%, but inferior results for session S3 (MedON): 58% and 54%. Both RS and ML could not predict UPDRS in DBS patients during stimulation-OFF visits because of differences in UPDRS. By using RS rules from BMT or DBS patients we could not predict UPDRS of POP group, but with certain limitations (only for MedON), we derived such predictions for the POP group from results of DBS patients by using ML models (60%). SIGNIFICANCE: Thanks to our RS and ML methods, we were able to predict Parkinson's disease (PD) progression in dissimilar groups of patients with different treatments. It might lead, in the future, to the discovery of universal rules of PD progression and optimise the treatment.

Keywords: Neurodegenerative disease, rough set, decision rules, granularity.

1. Introduction

This publication is an extension of the original communication given at PReMI 2017 and published in LNCS [1]. Only very experienced PD neurologists are successful in implementing individually adjusted therapy. In general doctors have very limited time for each patient and they have different academic backgrounds which may lead to introduction of variable treatments of patients and it may lead to confusions and ineffective therapy. We propose to improve doctor's approach by adding more automatic measurements and intelligence symptom classification [2, 3] that is similar to that was found in the visual system for the complex objects recognition [4].

It is important to estimate patient's disease stage as it determines different sets of therapies. Typical neurological standards are based on the Hoehn and Yahr and the UPDRS (Unified Parkinson's Disease Rating) scales. The last one is more precise and it will be mainly used in this study. We would like to estimate disease progression in different groups of patients that were tested during three visits every half-year.

This study is using several different AI methods in order to predict disease progression of Parkinson's patients with different treatments and in different disease stages. The ultimate purpose of our distinctive data mining methods is to find direction(s) to the optimal therapy for different patients with Parkinson's disease (PD). It is not easy as 'each PD is different' as effect of many compensatory individual processes in response to dying brain cells.

Our method may lead to the introduction of a more precise follow-up and even can be extended to telemedicine by using results of neuropsychological tests with parameters of the reflexive eye movements in order to predict UPDRS (disease progression) [1-3]. The primary Parkinson's disease symptoms are motor tested by neurologists as UPDRS III (Unified Parkinson's Disease Rating Scale, motor part III). Generally, it is difficult to measure them in a doctor-independent, automatic way. However, measurements of the gait as well as eye movements (EM) can be automatised [5]. Our fast (saccadic) EM tests are natural. When we notice a new object that appears in our visual field, we automatically, with a delay, look at it (perform a saccade) or sometimes look in the opposite direction (perform an antisaccade). Parameters of these fast EM are important PD biomarkers. Our antagonistic oculomotor actions are mainly suppressed by signals that substantia nigra (SN) sends to superior colliculus [6]. As SN is affected by Parkinson's disease and in less extent by ageing, the reflexive saccades delay is an important parameters related to the PD progression as we have demonstrated previously [2, 3]. The new aspect of this study is related to an extension of our previous results for patients with similar treatments [2, 3]. We would like to predict, using data mining methods, disease progression of patients with different therapies: on only medication, with short term DBS (deep brain stimulation), or more advanced patients with long term DBS. In the original communication [1], we have used only rough set theory methods and our predictions were not possible for all groups of patients. In this work, we have also used other ML models that helped to improve our estimates. Contents of each section are the following: the Methods section has two subsections: Rough Set and Machine Learning; the Results section has the following subsections: Comparing longitudinal UPDRS changes, BMT patients disease progression prediction, DBS and POP patients: rules for estimation of disease progression, Disease progression of DBS patients estimated by rules from BMT patients, Disease progression of POP patients estimated by rules from DBS patients; the following sections: Discussion and Conclusions have no subsections.

2. Methods

All 62 PD patients were divided into three groups: BMT patients (medication only), and patients on medication and with implanted electrodes in the STN (subthalamic nucleus [3]) during our study: DBS group or before our study: POP group.

The Deep Brain Stimulation (DBS) surgery was performed in the Institute of Neurology and Psychiatry, Warsaw Medical University. PD patients were tested in the following sessions: MedON/MedOFF sessions (sessions with or without medication). The other groups: DBS and POP patients were also tested in StimON/StimOFF session were DBS stimulation was switched ON or OFF. All combinations gave four sessions: 1) MedOFFStimOFF; 2) MedOFFStimON; 3) MedONStimOFF; 4) MedONStimON. Details of these procedures were described earlier [3]. Neurologists from WMU performed the UPDRS and neuropsychological tests. Fast eye movements - reflexive saccades were recorded as described in detail before [2, 3]. The following parameters of saccades were measured: the delay (latency) related to time difference between the beginning of the light spot movements and the beginning of the eye movement; saccade's amplitude in comparison to the light spot amplitude; max velocity of the eye movement; duration of saccade defined as the time from the beginning to the end of the saccade.

2.1. Rough Set

Our data mining analysis follows rough set (RS) theory after Zdzislaw Pawlak [7] because RS gave previously the best results in PD symptoms classifications in comparison to other methodologies [3]. Our data are represent as a decision table where rows represented different measurements (may be obtained from the same or different patients) and columns are related to different attributes. An information system [7] is as a pair *S = (U, A)*, where *U, A* are finite sets: *U* is the universe of objects; and *A* is the set of attributes. The value *a(u)* constitutes a unique element of *V* (where *V* is a value set) for $a \in A$ and $u \in U$.

A decision table for S is the triplet: *S = (U, C, D)* where: *C, D* are condition and decision attributes [8]. Each row of the information table gives a particular rule that connects condition and decision attributes for a single measurement of a particular patient. As there are many rows related to different patients and sessions, they gave many particular rules. Rough set approach allows generalising these rules into universal hypotheses that may determine optimal treatment options for an individual PD patient. Different rules' granularities (abstraction) are similar to complex objects recognition [4] and may simulate association processes of the "Golden Neurologist".
In the present study, we are trying to use data from different groups of patients for training and testing. The purpose was to find limits of rules that may predict symptoms development of patients with different treatments at different disease stages.
We have used the RSES 2.2 (Rough System Exploration Program) [9] with implementation of RS rules to process our data. By means of RSES we have generated rules using four different methods: exhaustive algorithm, genetic algorithm, covering algorithm, or LEM2 algorithm. In each case, we selected a particular algorithm that gave the shortest set of rules.

In rough set approach our discretisation was based on the cut values. We have replaced original attributes with new, binary attributes which indicate whether actual attribute value for an object is greater or lower than c (see [8]), we define c as a cut. This algorithm is implemented in RSES 2.2.

In RSES classification we have used 6-fold method, which means that the population was divided into 6 random subgroups and 5 were used for training and one for testing six times by changing in every trial-testing group. Global accuracy and coverage were the means of accuracy and coverage from all individual tests.

We have used standard RSES parameters: type of classifier: Decision rules (with 6 folds as described above), exhaustive algorithm of rules computation with shortening ratio: 1.0, and conflicts resolved by standard voting. These parameters gave the best global accuracy.

2.2. Machine Learning

For the evaluation of the results acquired by RSES we have also used additional methods of a supervised learning process. Because our data contains a set of N training samples of the form (x_1, y_1), ..., (x_n, y_n) such that x_i is the feature vector of the i_{th} sample and y_i is its class, it is possible to use a supervised learning algorithm which seeks for a function $g : X \rightarrow Y$, where the X is the input space, and the Y is the output space. The g function is an element of some space of possible functions

G, known as the hypothesis space. There is a slight difference in this approach in the comparison to RSES because supervised learning does not allow us to create the "none of the above" labels per se. In the circumstances of dealing with medical data, it is often precisely the class that is under-represented in the data, the disease or potential fault, that the network should detect [10], and it would be introduced in future research. Furthermore, we cannot generalise rules into universal hypotheses that may determine optimal treatment options for an individual PD patient, so they had been prepared for each group individually.

We have prepared an automated framework, written in Python and based on the scikit-learn [11], which was responsible for the evaluation of different supervised learning algorithms and provided the details of the result of the most accurate algorithm for a given dataset. The challenging task was to create decision classes, which were comparable with RSES. First of all, some values were missing, so we applied a strategy of imputation using the statistics (specifically: mean values). Then, to achieve similarity with the number of classes produced by RSES, we have applied unsupervised learning for clustering of the training set to obtain normalised classes (considered as different UPDRS ranges). We have observed that different strategies provide better results when considering different subsets of patients. Thus, we used two popular clustering approaches. For the DBS and POP patients' the best results we have achieved by K-Means [12] parametrised with a number of clusters = 4 and with a 'k-means++' method for initialisation including 10 runs and 300 maximum iterations. For BMT patients, better (closer to results of RSES for the sake of comparison) clustering was achieved by the Mean Shift [13] with an estimated value of bandwidth used in the RBF kernel based on standard quantile = 0.30.

When training data and labels were prepared for each set, we were able to apply a wide range of supervised algorithms and choose the best one on the basis of the highest accuracy score (in scikit-learn estimators have a 'score' method providing a default evaluation criterion for the problem they are designed to solve). The framework challenged every two pairs of the training set (e.g. POPW1) against validation dataset (e.g. BMTW1) and attempted to predict the UPDRS class of every patient. All scikit-learn classifiers are capable of multiclass classification, so the selection of presented algorithms has been based on the strategy. We have chosen well-known models that apply: inherently multiclass behavior (e.g. GradBoost), multiclass as One-Vs-One (e.g. SVC), multiclass as One-Vs-All (e.g. logit), multilabel support (e.g. Decision Tree), and with multiclass-multioutput support (e.g. Random Forest). Among others, we have evaluated the algorithms listed below; the selection was made both empirically and based on previous research [3].

- Logistic Regression (logit) - applied with 'multinomial' option with the 'lbfgs' solver and 'balanced' weights associated with classes,
- Linear Supported Vector Classification (LinearSVC) - with a 'crammer_singer' multi-class strategy;
- Gradient Boosting Classifier (GradBootst) - chosen loss function was 'deviance' with learning rate = 0.1;
- Gaussian Naive Bayes (GaussianNB) - used without prior probabilities of the classes;

- Decision Tree - with Gini impurity metric;
- K-Neighbors Classifier (KNC) - with a number of neighbours to get = 5 and uniform weights (all points in each neighbourhood were weighted equally);
- Random Forest - applied with 10 trees in the forest and Gini impurity metric;
- C-Support Vector Classification (SVC) - with the 'rbf' kernel type, 'ovo' (one-versus-one) multi-class strategy, penalty parameter $C = 1.0$ and $\gamma = 0.001$.

As the Results section shows, there was no most accurate model and most of them were found to be optimal only for a specific subset.

3. Results

All 62 PD patients were divided into three groups: BMT patients (medication only), and patients on medication and with implanted electrodes in the STN (subthalamic nucleus [2, 5]) during our study: DBS group or before our study: POP group.

In 23 patients of BMT group the mean age was 57.8+/- 13 (SD) years; disease duration was 7.1+/- 3.5 years, UPDRS was 36.1+/-19.2. In 24 patients of DBS group the mean age of 53.7+/- 9.3 years, disease duration was 10.25+/- 3.9 years (stat. diff. than BMT-group: $p < 0.025$), UPDRS was 62.1+/- 16.1 (stat. diff. than BMT-group: $p < 0.0001$). In 15 patients of POP group the mean age was 56.2+/- 11.3 (SD) years and disease duration was 13.5+/- 3.6 years (stat. diff. than DBS-group: $p < 0.015$), UPDRS was 59.2+/-24.5 (stat. diff. than BMT-group: $p < 0.0001$).

These statistical data are related to the data obtained during the first visit for each group: so-called BMTW1 (visit one), DBSW1 (visit one) and POPW1 (visit one).

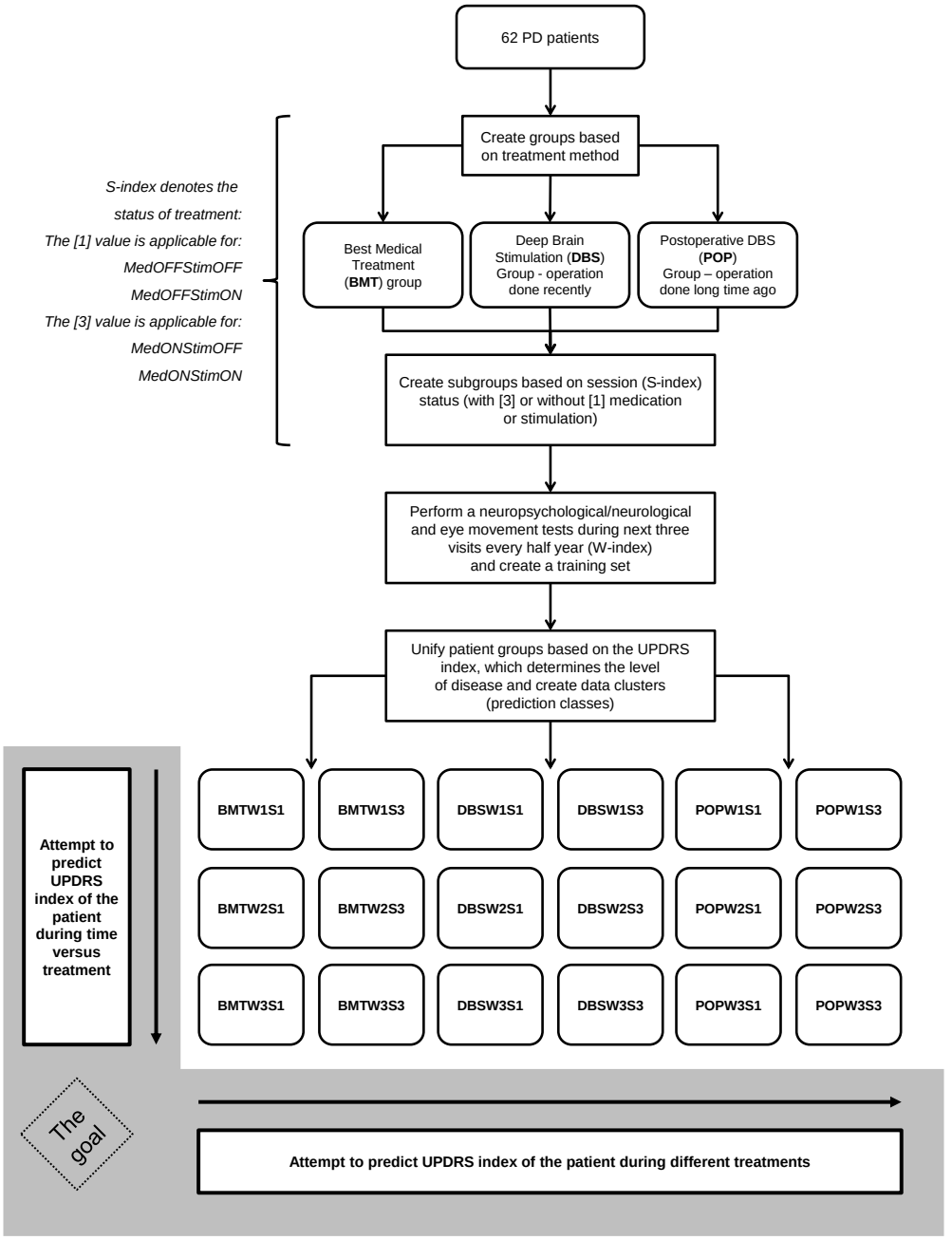

Figure 1: The diagram shows a flow model of the data mining process presented in the article. From 62 PD patients, we distinguished three groups based on treatment (BMT / DBS / POP). Each group was split further into subgroups based on the session (S) status and visit index (W). We aimed to predict the UPDRS index in two ways: during single treatment and among different treatments.

3.1. Comparing longitudinal UPDRS changes

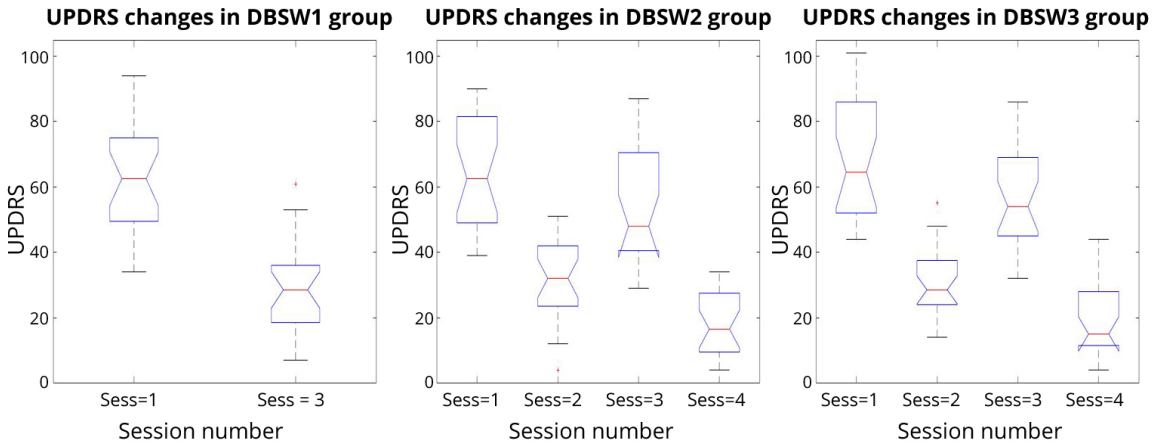

Figure 2: The boxplot shows that the median of UPDRS in different sessions and visits; W1, W2, W3 of the DBS group. There are minimum and maximum values in each session. Since the notches in the box plot do not overlap, you can conclude, with 95% confidence that the true medians do differ.

The first plot from the right presents UPDRS of DBSW1 group. There are only two sessions as patients were before the surgery. In the first session (described as Ses1) mean UPDRS was 62.2+/-16.1, in the second (described as Ses3) was 29.9+/-13.3 strongly ($p < 0.0001$) different from Ses1 and it represents the effect of medication. Plot in the middle of Fig. 2 represents DBSW2 after the surgery and UPDRS in Ses1 is larger than that before the surgery 65.3+/- 17.6 but there are no statistically significant differences, however UPDRS in Ses1 of DBSW3 is 68.7+/- 17.7 and statistically significant ($p < 0.03$) than in W2. Effects of different therapies (session numbers) are significantly different in W1, W2, and W3, but not different between the same session numbers in different visit (with the exception of the Ses3 in W1 as after the surgery the dosage of medication is reduced).

In POP-group UPDRS values are similar. There is an increase of the UPDRS Ses1 from W1: 63.1 +/- 18.2 to W2: 68.9+/-20.3 to W3: 74,2+/- 18.4 but there were smaller differences for Ses4 (MedOnDBSOn) W1: 21 +/- 11.3 to W2: 23.3+/-9.5 to W3: 23,8+/- 10.7. Therefore, we have assumed that groups DBS and POP are similar.

In BMT group UPDRS in Ses1 was W1: 48.3+/-17.9; W2: 57.3+/-16.8 ($p < 0.0005$ different than W1); W3: 62.2+/-18.2 ($p < 0.05$ different than W2). In Ses3 UPDRS was W1: 23.6+/- 10.3; W2: 27.8+/-10.8; W3: 25+/-11.6 (no statisticall difference between visits for Ses3).

3.2. BMT patients disease progression prediction

In these sections we have compared two different approaches: in 3.2.1 we have found RSES rules describing attributes for the first visit BMTW1 and have predicted on their basis disease progression in the same patients half (BMTW2) and one year later (BMTW3). In section 3.2.2 we have predicted disease progression in the same time periods, but using Gaussian Nave Bayes.

3.2.1. BMT patients: RS rules for the disease progression

The BMT patients (only on medication) were tested in two sessions (session 1: without, and session 3: with medication) three times every half-year.

We have used ML and rough set theory [9] in order to obtain rules determining decision and condition attributed for the first visit BMTW1. We have obtained these rules using Exhaustive algorithm and additionally rules filter that removes rules with support 1 (a single case rules) [9]. In the most cases Exhaustive algorithm gave the shortest set of rules and the largest support (for more cases).

On the basis of these rules we have predicted the UPDRS values obtained during the second (half - year later W2 - BMTW2) and the third (one year later BMTW3) visits. UPDRS was optimally divided by RSES into 4 ranges: $"(-Inf, 24.0)"$, $"(24.0, 36.0)"$, $"(36.0, 45.0)"$, $"(45.0, Inf)"$ for both visits (W2 and W3) the global coverage was 1.0 and the global accuracy was 0.7. Example of rules from BMTW1:

$$(Ses = 3)\&(PDQ39 = "(-Inf, 50.5)") \Rightarrow (UPDRS = "(-Inf, 33.5)"[12])12 \quad (1)$$

$$(dur = "(-Inf, 5.65)")\&(Ses = 3)\&(Epworth = "(-Inf, 14.0)") \Rightarrow (UPDRS = "(-Inf, 33.5)"[7])7 \quad (2)$$

$$(dur = "(5.65, Inf)")\&(Ses = 3)\&(Epworth = "(14.0, Inf)") \Rightarrow (UPDRS = "(-Inf, 33.5)"[4])4 \quad (3)$$

In the first rule (1) if the session number 3 and *PDQ39 = "(-Inf, 50.5)"* then UPDRS was *(-Inf, 33.5)* in 12 cases. The second rule (2) was fulfilled in 7 cases and the third one (3) in 4 cases. There were altogether 70 rules.

3.2.2. BMT patients: ML models for the disease progression

As mentioned above, every ML model was independently trained for every session separately, to enhance its prediction results.

For session 1, the UPDRS ranges of BMTW1 were optimally divided into four classes with the middle intervals in points: $C_0 = 26.0, C_1 = 45.5, C_2 = 58.67, C_3 = 104.0$. The scopes of 4 ranges were: $"(0, 35.0)"$, $"(36.0, 52.0)"$, $"(53.0, 81.0)"$, $"(82.0, 114.0)"$. On the basis of these scopes, we have predicted the UPDRS values obtained during the second (half -year later W2 - BMTW2) and the third (one year later BMTW3) visits. In this session, the Gaussian Naive Bayes (GaussianNB) achieved the best results. The prediction on BMTW2 set had shown the accuracy of 0.61 with the coverage 1.0. For BMTW3 the accuracy was 0.57 with the coverage = 1.0. For the session 3, as the basis for learning model, we had chosen BMTW2 to predict classes of UPDRS in BMTW3. UPDRS was normalised and clustered into new scopes: $"(0.0, 19.0)"$, $"(20.0, 28.0)"$, $"(29.0, 40.0)"$, $"(41, 114.0)"$ with centroids: $C_0 = 13.87, C_1 = 25.00, C_2 = 32.67, C_3 = 47.50$. On the basis of those rules, we were able to predict the UPDRS classes of BMTW3 with the accuracy = 0.61 (Table 1).

Table 1: Confusion matrix for UPDRS of BMTW3 (Session 3) by model obtained from BMTW2 predicted by Decision Tree. TPR: True positive rates for decision classes; ACC: Accuracy for decision classes: the global coverage was 1 and the global accuracy was 0.61.

		Predicted				
Actual		”(20.0, 28.0)”	”(0.0, 19.0)”	”(29.0, 40.0)”	”(41.0, 114.0)”	ACC
	”(20.0, 28.0)”	5	1	1	0	0.71
	”(0.0, 19.0)”	1	7	2	0	0.70
	”(29.0, 40.0)”	0	1	1	2	0.25
	”(41.0, 114.0)”	1	0	0	1	0.50
	TPR	0.71	0.78	0.25	0.33	

3.3. DBS and POP patients: rules for estimation of disease progression

In these sections we have compared two different approaches: in 3.3.1 we have found RSES rules describing attributes for the second visit DBSW2 and have predicted on their basis disease progression in the same patients half year later (DBSW3). We have performed similar predictions for POP patients. On the basis of the first visit POPW1 we have predicted disease progression in two later visits: POPW2 and POPW3. I section 3.3.2 we have used the Logic Regression algorithm in order to predict disease progression in POPW3 on the basis of POPW2 group.

3.3.1. DBS patients: RS rules for estimation of disease progression

As DBSW1 had only 2 sessions (before surgery) we could only predict session DBSW3 on the basis of DBSW2 (half a year earlier) (Tab.2). We have predicted UPDRS for visits POPW2 and POPW3 on the basis of visit POPW1 with total accuracy: 0.667 and 0.793 with a coverage: 1 and 0.967.

Table 2: Confusion matrix for UPDRS of DBSW3 by rules obtained from DBSW2. TPR: True positive rates for decision classes; ACC: Accuracy for decision classes: the global coverage was 1 and the global accuracy was 0.562.

		Predicted					
Actual		”(46.0, 72.0)”	”(38.0, 46.0)”	”(19.5, 38.0)”	”(72.0, Inf)”	”(-Inf, 19.5)”	ACC
	”(46.0, 72.0)”	12	5	2	5	1	0.48
	”(38.0, 46.0)”	2	5	1	2	2	0.42
	”(19.5, 38.0)”	0	4	13	3	7	0.48
	”(72.0, Inf)”	4	0	0	12	0	0.75
	”(-Inf, 19.5)”	0	0	4	0	12	0.75
	TPR	0.67	0.4	0.65	0.55	0.6	

3.3.2. POP patients: ML models for estimation of disease progression

We were only able to predict POPW3 in Session 3 by model obtained from POPW2 with the usage of Logistic Regression, which achieved the accuracy of 0.6. UPDRS was normalised and clustered into scopes: $"(0.0, 15.0)", "(16.0, 22.0)", "(23.0, 32.0)", "(33.0, 114.0)"$ with centroids: $C_0 = 12.33, C_1 = 18.00, C_2 = 27.00, C_3 = 38.00$ (Tab. 3).

Table 3: Confusion matrix for UPDRS of POPW3 (Session 3) by model obtained from POPW2 by Logistic Regression. TPR: True positive rates for decision classes; ACC: Accuracy for decision classes: the global coverage was 1 and the global accuracy was 0.60.

		Predicted				
		”(16.0, 22.0)”	”(34.0, 114.0)”	”(0.0, 15.0)”	”(23.0, 32.0)”	ACC
Actual	”(16.0, 22.0)”	1	2	1	0	0.25
	”(34.0, 114.0)”	2	2	1	0	0.40
	”(0.0, 15.0)”	0	0	4	0	1.00
	”(23.0, 32.0)”	0	0	0	2	1.00
	TPR	0.33	0.50	0.67	1.00	

3.4. Disease progression of DBS patients estimated by rules from BMT patients

In these sections we have compared different approaches: in 3.4.1 by using RSES we have found rules describing attributes for the first visit BMTW1 and have predicted on their basis disease progression in different patients with different therapy in three visits: DBSW1, DBSW2, and DBSW3. In session 3.4.2 in order to predictions of DBS patients disease progression in the BMT group we have used several different algorithms: Gradient Boosting Classifier, C-Support Vector Classification, and LinearSVC algorithms.

3.4.1. Disease progression of DBS patients estimated by RS rules from BMT patients

As BMT patients had only two sessions (S1 - MedOff, and S3 - MedON) and DBS patients four sessions (see Methods) we have divided them into two sets: one with StimON set-up and another one with StimOFF set-up. We were not successful in the prediction of StimOFF sessions as DBS patients were in a more advanced stage than BMTW1 group. Our UPDRS predictions for DBSW1 had global accuracy 0.64 (coverage 0.5); for DBSW2 - global accuracy was 0.85 (coverage 0.3); for DBSW3 - global accuracy was 0.74 (coverage 0.6).

3.4.2. Disease progression of DBS patients estimated by ML models from BMT patients

For S1, we were able to predict the UPDRS of DBSW1 by model obtained from BMTW3 based on Gradient Boosting Classifier with the accuracy 0.54. C-Support Vector Classification shown the accuracy of prediction as high as 0.88 for DBSW2 and DBSW3 when it was trained on the basis of

BMTW3 (coverage 1.0). However, this classification should be treated with caution because there is a significant difference between UDPRS scores distribution between the classes, and their scopes and centroids are shifted (for example: the first calculated class of model based on BMTW3 has a range which is very extensive: "$(0.0, 55.0)$" and every UPDRS score from DBSW2 is lower than 51.0).

For S3, the prediction of UPDRS scores in DBSW2 has an accuracy of 0.58 and 0.54 for DBSW3 when trained on BMTW2 by LinearSVC (coverage 1.0). We were unable to predict classes of DBSW1 based on BMT.

3.5. Disease progression of POP patients estimated by rules from DBS patients

In these sections we have compared different approaches: in 3.5.1 we have no success to predict POP patient disease progression on the basis of DBS patients. In 3.5.2 by using Random Forest algorithm we could predict POPW1 disease progression from DBSW2 patients.

3.5.1. Disease progression of POP patients estimated by RS rules from DBS patients

We could not predict UPDRS of POP patients from rules obtained from DBS patients probably because many years of DBS might influence some brain synaptic connections. In case of POP patients responses to MedON/OFF are not consistent with responses in DBS patients.

3.5.2. Disease progression of POP patients estimated by ML models from DBS patients

Table 4: Confusion matrix for UPDRS of POPW1 (Session 3) by model obtained from DBSW2 by Random Forest. TPR: True positive rates for decision classes; ACC: Accuracy for decision classes: the global coverage was 1 and the global accuracy was 0.60.

		Predicted				
		"(12.0, 18.0)"	"(0.0, 11.0)"	"(27.0, 114.0)"	"(19.0, 26.0)"	ACC
Actual	"(12.0, 18.0)"	4	1	1	0	0.67
	"(0.0, 11.0)"	0	2	0	0	1.00
	"(27.0, 114.0)"	3	0	2	0	0.40
	"(19.0, 26.0)"	0	0	1	1	0.50
	TPR	0.57	0.67	0.50	1.00	

However, in this section we are looking for prediction of UPDRS only in one session (MedON). The UPDRS data in subset of DBSW2 (Session 3 - MedON) was optimally divided with centroids: $C_0 = 6.71, C_1 = 16.12, C_2 = 21.60, C_3 = 32.17$ and with ranges of: "$(0.0, 11.0)$", "$(12.0, 18.0)$", "$(19.0, 26.0)$", "$(27.0, 114.0)$". On the basis of the model from DBSW2 - session 3, we were able to predict the UPDRS classes of POPW1 session 3 with the accuracy of 0.60 (Tab.4). It is worth noting that it was true only for particular datasets: DBSW2 and POPW1 and the particular session - MedON.

Table 5: Performance comparison among different ML models.

Reference	Training Set	Test Set	RS Accuracy	RS Coverage	ML Accuracy	ML Coverage	ML Model
3.2.1	BMTW1S{1,3}	BMTW2S{1,3}	0.70	1.00			
3.2.1	BMTW1S{1,3}	BMTW3S{1,3}	0.70	1.00			
3.2.2	BMTW1S1	BMTW2S1			0.61	1.00	Gaussian Naive Bayes
3.2.2	BMTW1S1	BMTW3S1			0.57	1.00	Gaussian Naive Bayes
3.2.2	BMTW2S3	BMTW3S3			0.61	1.00	Decision Tree
3.3.1	DBSW2S{1,3}	DBSW3S{1,3}	0.56	1.00			
3.3.2	POPW1S{1,3}	POPW2S{1,3}	0.67	1.00			
3.3.2	POPW1S{1,3}	POPW3S{1,3}	0.79	0.97			
3.3.3	POPW2S3	POPW3S3			0.60	1.00	Logistic Regression
3.4.1	BMTW{1,2,3}S{1,3}	DBSW1S{1,3}	0.64	0.50			
3.4.1	BMTW{1,2,3}S{1,3}	DBSW2S{1,3}	0.85	0.30			
3.4.1	BMTW{1,2,3}S{1,3}	DBSW3S{1,3}	0.74	0.60			
3.4.2	BMTW3S1	DBSW1S1			0.54	1.00	Gradient Boosting
3.4.2	BMTW3S1	DBSW2S1			0.88	1.00	C-Support Vector
3.4.2	BMTW3S1	DBSW3S1			0.88	1.00	C-Support Vector
3.4.2	BMTW2S3	DBSW2S3			0.58	1.00	LinearSVC
3.4.2	BMTW2S3	DBSW3S3			0.54	1.00	LinearSVC
3.5.2	DBSW2S3	POPW1S3			0.60	1.00	Random Forest

4. Discussion

There are novel technologies and data constantly improving PD patients' treatments, but there are also still doubts if the actual procedures are optimal for a particular case. Our long time purpose is to use data mining and machine learning in order to compare different neurological protocols and their effectiveness. We believe that the best future approach will be to perform all tests automatically at home, process them with intelligent algorithms and to submit results to the doctor for his/her decision (compare with telemedicine based methods [14]). Another, more advanced approach that we were testing in this work, would be to create standard treatment for each new case on the basis of already successfully treated patients and correct treatment as symptoms are developing in time (see technology in PD [15]). We have demonstrated that it is relatively easy to estimate symptoms and their time development in populations treated in different ways (e.g. only medication treatment) (compare to motor automaticity [16]). This result may give the basic (locally optimal) follow-up PD symptoms. If the patient is doing significantly worse than others (rules), their treatment is not optimal and should be changed. In the next step, we may use rules obtained from different clinics to make them even more universal and optimal. Our new approach is related not only to longitudinal study but also to test different patient population with different treatments. Could our results lead to finding optimal procedures for different cases and diverse cares? It seems to be a good direction, but there are many particular issues, for example when patients are in different stages of the disease. In our case, the second group of patients (DBS group) were in a more advanced stage of disease so it was not possible to get 100% coverage like in the first case. The second group with longitudinal study had a new treatment (brain stimulation) that started from the second visit. We have tested if the same treatment in different populations gives similar results. Patients got two treatments: medication (medication ON and OFF) and electric brain stimulation (ON and OFF). We have analysed these treatments as two different sets: 1) StimOFF: medication ON and OFF; 2) StimON: medication ON and OFF. As a result,

it was not possible to obtain sufficient accuracy in the first situation (1) as patients with DBS (DBS and POP groups) were more advanced in their disease stages and had different UPDRS ranges: mean UPDRS in BMT group was 36 and in DBS/POP groups was 54/56 respectively (see Results section). Therefore, we can interpret brain stimulation (DBS) as resetting symptoms in time of disease - patients with 10 years disease duration have symptoms similar to patients with 7 years disease duration and we could predict their symptoms with reasonable accuracy.

But this assumption was not exact for patients with long period of brain stimulation (DBS) - POP group (with over 13 years disease duration). This group was different than the other two as we did not succeed in forming a good prediction by rules obtained by other groups BMT or DBS. It might be related to the longer period of brain stimulation (DBS) that has changed some central mechanisms [17]. It is an important negative result that needs further study. In the near future, we may look for additional condition attributes in order to improve the global accuracy. In the present paper in comparison to the previous one [1] we have introduced additional to RS machine learning (ML) models [10-11].

We have compared intelligent methods based on rough set theory with several different machine learning algorithms: Gaussian Nave Bayes, Decision Tree, Logistic Regression, Gradient Boosting, C-Support Vector, LinearSVC, and Random Forest. Generally rough set method gave better accuracy but not as good coverage as other algorithms (see Table 5).

In most cases these new methods gave similar results to our RS rules. However, in contrast to RS rules, ML models could predict results of one session (session 3 - MedON) in POPW1 patients on the basis of DBSW2 session 3 results. It is a very limited case, but showing some common mechanisms between DBS and POP groups.

Furthermore, as shown in the Table 5, Rough Set-based approach allows creating more general rules without the necessity of additional data splitting (in this case: into different sessions), which was required in the ML modelling. The principal conclusion which comes from that comparison is an observation that RS is much more universal method when considering medical data, which constitutes a confirmation of the findings from [3].

In summary, we have demonstrated that the DBS (electric STN stimulation) procedure revoked and improved rules that became similar to rules of early stage Parkinson's disease patients.

5. Conclusions

This work is a continuation of our previous findings [1, 3], comparing classical approach used by most neurologists and based on their partly subjective experience and intuitions with the intelligent data processing (machine learning, data mining) classifications. This paper represents a new technological and data mining trend in treatments of the neurodegenerative diseases [18, 19, 20]. We have furthermore demonstrated that such additional attributes as parameters of eye movements and neuropsychological data are significant in predicting longitudinal symptom developments in different therapy related groups of PD patients. We have confirmed that a multidisciplinary approach incorporating neurologists, automatic measurements and intelligent data mining such as granular computing is an example of open-science, collaborative research that is the future of complex healthcare. In the near future, we are planning to involve also other clinics in our project as well as other neurodegen-

erative diseases such as Alzheimer's disease. Our direction is to increase the number of attributes in order to find significant attributes not only for each disease or treatment, but also for each individual patient. It would lead to a patient-centred and patient-oriented intelligent telemedicine.

We have demonstrated that on the basis of the first group of early PD patients (only on medications), we could predict symptoms of patients in more advanced stages and with different treatments. It means that our rules have some universal properties, which means that in the future we may suggest an optimal treatment for each individual patient.

Acknowledgments

Authors would like to thank English Lecturer Michal Zawistowski for proofreading the article.

Authors' contributions

The first author AWP has written the paper and performed RST data analysis, the second author AC has written part of the paper and performed ML analysis, the third SS and fifth DMK coauthors have performed all tests with PD patients, the fourth coauthor has created the database.

References

1. Przybyszewski, A.W., Szlufik, S., Habela, P., Koziorowski, D.M.: Rough Set Rules Determine Disease Progressions in Different Groups of Parkinson's Patients. B.U. Shankar et al. (Eds.): PReMI 2017, LNCS 10597, pp. 270-275, (2017)

2. Przybyszewski, A.W., Kon, M., Szlufik, et al.: Data Mining and Machine Learning on the Basis from Reflexive Eye Movements Can Predict Symptom Development in Individual Parkinson's Patients. In Nature-Inspired Computation and Machine Learning; Eds. Gelbukh et al. Springer pp. 499-509 (2014)

3. Przybyszewski, A.W., Kon, M., Szlufik, S., Szymanski, A., Koziorowski, D.M.: Multimodal Learning and Intelligent Prediction of Symptom Development in Individual Parkinson's Patients. Sensors, 16(9), 1498; doi:10.3390/s16091498 (2016)

4. Przybyszewski, A.W.:The Neurophysiological Bases of Cognitive Computation Using Rough Set Theory. *Transactions on Rough Sets IX* (J.F. Peters et al. Eds.), LNCS 5390, 287-317, (2008)

5. Przybyszewski, A.W.: Applying Data Mining and Machine Learning Algorithms to predict symptoms development in PD. Annales Academiae Medicae Silesiensis 68(5), 332 - 349 (2014)

6. Yasuda, M., Hikosaka, O.: To Wait or Not to Wait - Separate Mechanisms in the Oculomotor Circuit of Basal Ganglia. Front Neuroanat. 11: 35 (2017)

7. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer, Dordrecht, (1991)

8. Bazan, J., Son, Nguyen, H., Trung, T., Nguyen, Skowron A., Stepaniuk, J.: Desion rules synthesis for object classification. In: E. Orlowska (ed.), Incomplete Information: Rough Set Analysis, Physica - Verlag, Heidelberg, pp. 23-57 (1998)

9. Bazan, J., Szczuka, M.: RSES and RSESlib - A Collection of Tools for Rough Set Computations. W. Ziarko and Y. Yao (Eds.): RSCTC 2000, LNAI 2005, pp. 106-113 (2001)

10. Marsland, S.: Novelty detection in learning systems: Neural computing surveys, 3(2), pp. 157-195 (2003)

11. Pedregosa et al.: Scikit-learn: Machine Learning in Python: JMLR 12, pp. 2825-2830 (2011)

12. MacQueen, J.: Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. 1, 14. (1967)

13. Comaniciu D., Meer P.: Mean Shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence. pp. 603-619 (2002)

14. Lindauer, A., Croff, R. Mincks, N. Mattek, S.J. Shofner, N. Bouranis, L. Teri R.: It Took the Stress out of Getting Help: The STAR-C-Telemedicine Mixed Methods Pilot. Care Wkly. 2: 7-14. doi: 10.14283/cw.2018.4 (2018)

15. Espay, A.J., Bonato, P., Nahab FB et al. from Movement Disorders Society Task Force on Technology.: Technology in Parkinson disease: Challenges and Opportunities. Mov Disord. 31(9): 1272-1282, (2016)

16. Wu, T., Hallett, M., Chan, P.: Motor automaticity in Parkinson's disease. Neurobiol Dis. 82: 226-234, (2015)

17. Umemura, A., Oyama, G., Shimo, Y., et al.: Current Topics in Deep Brain Stimulation for Parkinson Disease. Neurol Med Chir (Tokyo). 56(10): 613-625 (2016)

18. Tucker, C., Han, Y., Nembhard. H.B. et al.: A data mining methodology for predicting early stage Parkinson's disease using non-invasive, high-dimensional gait sensor data. Healthcare (Basel) Jun; 6(2): 54 (2018)

19. Dinov, I.D., Heavner, B., Tang, M. et. al.: Predictive Big Data Analytics: A Study of Parkinson's Disease Using Large, Complex, Heterogeneous, Incongruent, Multi-Source and Incomplete Observations. PLoS One. 2016; 11(8): e0157077. (2016)

20. Gao, C., Sun, H., Tuo Wang, T. et al.: Model-based and Model-free Machine Learning Techniques for Diagnostic Prediction and Classification of Clinical Outcomes in Parkinson's Disease. Sci Rep. 8: 7129. doi: 10.1038/s41598-018-24783-4 (2018)

Eye-Tracking and Machine Learning Significance in Parkinson's Disease Symptoms Prediction

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

[1] Polish-Japanese Academy of Information Technology, Koszykowa 86 Street, 02-008 Warsaw, Poland
{artur.chudzik,artur.szymanski,nowacki,przy}@pjwstk.edu.pl
[2] Department of Neurology, University of Massachusetts Medical School, 65 Lake Avenue, Worcester, MA 01655, USA
andrzej.przybyszewski@umassmed.edu

Abstract. Parkinson's disease (PD) is a progressive, neurodegenerative disorder characterized by resting tremor, rigidity, bradykinesia, and postural instability. The standard measure of the PD progression is Unified Parkinson's Disease Rating (UPDRS). Our goal was to predict patients' UPDRS development based on the various groups of patients in the different stages of the disease. We used standard neurological and neuropsychological tests, aligned with eye movements on a dedicated computer system. For predictions, we have applied various machine learning models with different parameters embedded in our dedicated data science framework written in Python and based on the Scikit Learn and Pandas libraries. Models proposed by us reached 75% and 70% of accuracy while predicting subclasses of UPDRS for patients in advanced stages of the disease who respond to treatment, with a global 57% accuracy score for all classes. We have demonstrated that it is possible to use eye movements as a biomarker for the assessment of symptom progression in PD.

Keywords: Eye-tracking · Saccades · Parkinson's Disease · Machine learning

1 Introduction

In Parkinson's disease, we can distinguish multiple therapies, which could be combined. The gold-standard treatment for PD is a pharmacological treatment with Levo-dopa (L-dopa) [1, 2]. Nevertheless, L-dopa associates with long-term disturbances. We can distinguish hypo-hyperkinetic phenomena and psychosis as examples of motor and mood side effects. [3]. The practical and safe procedure, which is lacking these effects and remains the preferred surgical treatment for advanced Parkinson's disease, is Deep Brain Stimulation of the subthalamic nucleus (STN) [4]. However, there is still a necessity of the support for the neurologists in the field of optimum treatment parameters, because due to the huge diversity of cases, even the most experienced doctors could not be sure how the therapy would influent on the patient.

N. T. Nguyen et al. (Eds.): ACIIDS 2020, LNAI 12034, pp. 537–547, 2020.
https://doi.org/10.1007/978-3-030-42058-1_45

2 Methods

2.1 The Subject of the Study

The research group was composed of 62 patients who have Parkinson's disease and who are under the supervision of the Warsaw Medical University (Warsaw, Poland) neurologists. We differentiate the patients into three groups. The first one, we named the Best Medical Treatment (BMT). In this group, we placed patients who were treated only by the medication. The second one, the Deep Brain Stimulation (DBS), was a group where patients, who had implanted electrodes in the STN during our study. The last group, which was named Post-Operative Patients (POP) aggregates individuals who had had surgery earlier (before the beginning of our research). Every PD patient had three visits, which were underdone approximately every six months. Every patient from the DBS group had his/her visit before surgery.

In order to obtain the countable value of the disease, it is crucial to provide a precise neurological tool which could objectively measure all of the symptoms and determine a score. For the metric of Parkinson's disease advancement, there are two common neurological standards: the Hoehn and Yahr scale and the Unified Parkinson's Disease Rating (UPDRS). The UPD rating scale is the most commonly used in the clinical study of Parkinson's disease [5], and we also decided to adopt it in this study. Altogether with UPDRS, every patients' disease metric was combined with the disease duration, the result of Parkinson's Disease Questionnaire PDQ39 (which is a disease-specific health-related quality-of-life outline), the result of Epworth Sleepiness Scale (which is intended to measure daytime sleepiness), and the parameters of saccadic eye movements, described further.

The mean age of patients was 51.1 $\pm$ 10.2 (standard deviation) years. The mean duration of the disease was 11.6 $\pm$ 4.3 years. The mean of UPDRS score (for all symptoms) was 33.8 $\pm$ 19.4. The mean of PDQ39 score was 50.5 $\pm$ 26.0, and the mean of Epworth score was 8.7 $\pm$ 4.6.

2.2 Eye Movements

The often diagnosed impairment of automatic behavioral responses accompanies the slowness of initiation of voluntary movements in individuals with PD [6]. An individual set of behavioral tasks may provide insight into the neural control of response suppression with the usage of motor impairments analysis based on saccadic eye movements [7, 8]. Saccades are a quick, simultaneous movement of both eyes between two or more phases of fixation in the same direction, and can be measured quickly and precisely.

We choose this marker because of the considerable understanding of the neural circuitry controlling the planning and execution of saccadic eye movements [9].

During this study, we have used a head-mounted saccadometer JAZZ-pursuit (oberconsulting.com), which was able to measure the reflexive saccades (RS) in the high frequency (1000 Hz). We have chosen this device because it is optimized for easy set-up and provides minimal intrusiveness while can keep stable 1 kHz frequency of measurement. During the experiment, we created a task of the horizontal reflexive saccades

analysis. Used hardware allowed us to obtain high accuracy and precision in eye tracking and the compensation of possible subjects' head movements relative to the monitor. Thus subjects did not need to be positioned in an unnatural chinrest, which has a positive influence on the ergonomy of the experiment. However, we asked patients to use a headrest in order to minimize the head motion because they could have a significant influence on the accuracy of the high-frequency measurements. Each patient was seated in front of the monitor at a distance of 60–70 cm.

The patients' task was to fix their eyes on the spot placed in the middle of the screen (0°). Then, the spot changed color into one of the possible variations and shift horizontally to one of the possible directions: 10° to the left, or 10° to the right, after arbitrary time ranging between 0.5–1.5 s. During that task, we measured the fast eye movements of the patient, according to the spot color transition.

When the transition was from white to green, it was a signal for the execution of RS. We also prepared an additional protocol for antisaccades (AS) measurement. In this task, the individual was asked to make a saccade in the direction away from the stimulus. A signal for that was when the spot changed color from white to red. After that, the central marker was hidden, and one of the two peripheral targets was shown. The selection was made randomly, with the same probability.

According to the task, each patient looked at the spots and followed them as they moved (in the RS task) or made opposite direction eye movement (in the AS task). After that, the target remained still for 0.1 s before the next experiment initialization.

In each test, the subject had to perform twenty saccades and antisaccades in a row-twice. The recording of the first session (marked as S1) was with the patient who has temporarily disabled treatment (without medicine/with the disabled neurostimulator). In the next session (marked as S3), the patient took medication and had a break for one half to one hour, and then the same experiments were performed (with L-dopa/with enabled neurostimulator).

In this experiment, we have investigated only RS data using the following population parameters averaged for both eyes: mean latency (*RSLat* $\pm$ SD), mean amplitude (*RSAmp* $\pm$ SD), mean of the maximum velocity (*RSPVel* $\pm$ SD), and mean duration of the saccades (*RSDur* $\pm$ SD).

2.3 Dataset

We implemented a dedicated database for measurements storage that was designed and maintained by Polish-Japanese Academy of Information Technology (Warsaw, Poland). For operations described further, we flattened the data, placing every experiment in each row, with the results and metadata in the separated columns. Therefore it could be represented by a single table represented by comma-separated-values, which is a universal format for computational engines. The basic structure of the dataset contains 374 observations, and each has 13 variables, which are: *Duration* - the duration of the disease; *UPDRS, PDQ39, Epworth* - the score for each test; *RSLat, RSDur, RSAmp, RSPVel* - the parameters of recorded saccades; *Session, Visit* - the indexes of the experiment; and *BMT, DBS, POP* - boolean variables which describe the kind of patients' therapy.

2.4 Computational Learning Theory

Our data contains a set of N training samples of the form $(x_1, y_1), \ldots, (x_n, y_n)$ such that x_i is the feature vector of the $i - th$ sample with the class denoted by y_i. Thus, it is possible to use a supervised learning algorithm which seeks for a function $g : X \rightarrow Y$, where the X is the input space, and the Y is the output space. The g function is an element of some space of possible functions G, known as the hypothesis space.

The task itself could is a multiclass classification. Our goal is to predict a level of the disease measured as UPDRS value binned into intervals, for different groups of patients. For the predictions, we have created a dedicated machine learning framework, written in Python and based on two libraries: Scikit Learn [11] and Pandas, which are high-quality, well-documented collection of canonical tools for data processing and machine learning. We have chosen well-known models that implement different multiclass strategies, such as K Neighbors Classifier, Support Vector Classifier, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier. The framework allowed us to find an optimal solution by the examination of multiple algorithms with different parameters.

2.5 Hypothesis

Our goal was to predict Parkinson's disease progression in the advanced stage, based on the data obtained from the patients in the different treatment and stage of this disease. This task is non-trivial because there are significant differences between symptom developments and the effects of different treatments in individual PD patients.

As a training dataset, we used patients from the BMT group (3rd visit), DBS (3rd visit), and POP (1st visit). The independent test set consisted of the POP group from the second visit.

3 Results

Our framework was responsible for every step of data processing in order to evaluate the best model based on given data. Therefore, we implemented procedures which were responsible for the creation of the Profiling Report, Correlation Matrices, Missing Data Imputation, Data Discretization, One-Hot Encoding or Data Normalisation of selected variables and Machine Learning Algorithms Evaluation for various parameters.

3.1 Profiling

Our dataset consists of 374 observations, where each has 13 variables. First, we generated a Pearson correlation coefficient matrix, where each cell in the table shows the correlation between two variables. It is a useful tool that proves if a correlation between the parameters from measurements and the symptoms suggest that some close relations exist (Fig. 1).

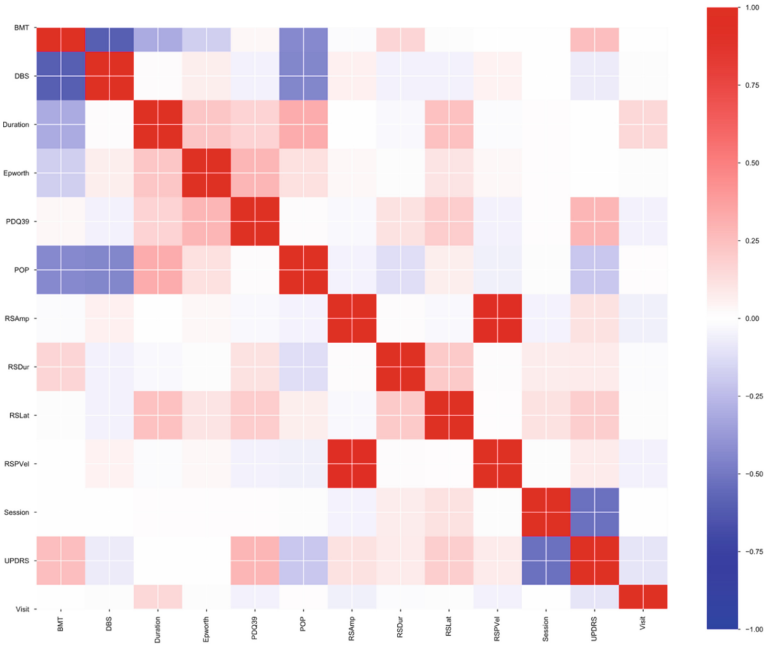

Fig. 1. Correlation matrix. The similarity between all paired parameters is correlated with the color of the cell in the matrix (red shade represents high similarity; blue stands contrariwise). (Color figure online)

3.2 Pre-processing

In the report, we noticed that some of the values are empty. *Epworth* column had 5 (1.3%) zeros; *RSAmp*, *RSDur*, and *RSLat* columns had 6 (1.6%) zeros. Because of the Epworth scale ranges between 0–24, we decided not to apply data imputation on that column. When we analyzed records related to a single patient, we noticed, that on an early stage of the disease (duration about 7 ± 0.3 years) two of them indeed reported no problems with the daytime sleepiness, yet visits in the following years revealed a linear increase in the results. For missing parameters of the saccades, we applied the imputation transformer for completing missing values which replaced missing values using the mean along each column.

For neurological and neuropsychological tests results (Fig. 2), we applied k-bins discretization, which provides a way to partition continuous features into discrete values. Thus those features are a one-hot encoded allowing the model to be more expressive while maintaining interpretability.

For neurological and neuropsychological tests results, we applied k-bins discretization, which provides a way to partition continuous features into discrete values. Thus those features are a one-hot encoded allowing the model to be more expressive while maintaining interpretability.

$$Epworth = (-\infty, 6.00), [6.00, 11.00), [11.00, +\infty)$$

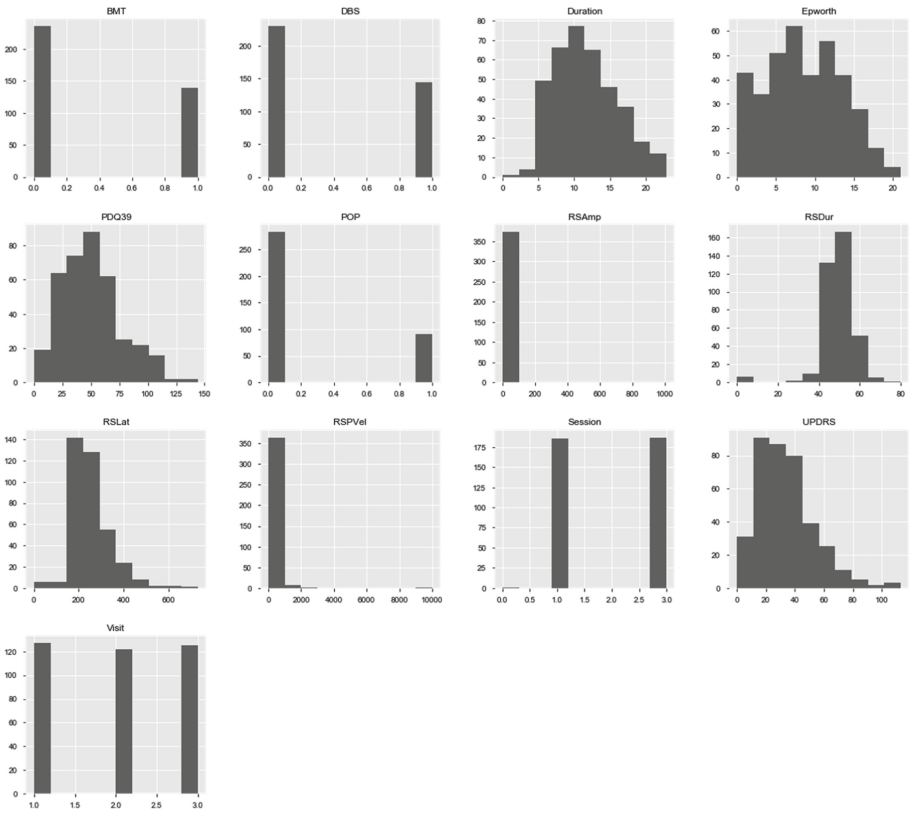

Fig. 2. A histogram, which is a representation of the distribution of data for every measurable property in the data set. Y-Axis represents a number of samples. The X-Axis presents a parameter value.

$$PDQ39 = (-\infty, 25.0), [25.0, 38.0), [38.0, 47.5), [47.5, 58.0), [58.0, 74.0), [74.0, +\infty)$$

Other columns (*Duration, RSLat, RSDur, RSAmp, RSPVel*) were standardize by removing the mean and scaling to unit variance to keep the subtle representation of the eye movement signal. The calculation of the standard score of X sample is defined by:

$$z = \frac{x - u}{s}$$

where u is the mean of the training samples, and s is the standard deviation of the training samples.

Target value, the *UPDRS* score was optimally divided into four ranges

$$UPDRS = (-\infty, 19.25), [19.25, 30.50), [30.50, 44.00), [44.00, +\infty)$$

This split ensured approximately the same data set size for each class.

3.3 Machine Learning

We decided to evaluate multiple algorithms in order to determine the best approach in the meaning of accuracy of the predictions. In the previous research [10], Random Forrest Classifier achieved the high prediction level, being second to Rough Set. This article covers the evaluation of a few machine learning models which were not used in the previous research, such as K Neighbors Classifier, Support Vector Classifier, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier. The following sections present each algorithm with evaluated parameters and scores obtained from our machine learning framework.

K Neighbors Classifier
Classifier implementing the k-nearest neighbors' vote with uniform weights achieved maximum accuracy score (0.50) with a parameter defining the number of neighbors on the level of 12 and 13 (Fig. 3).

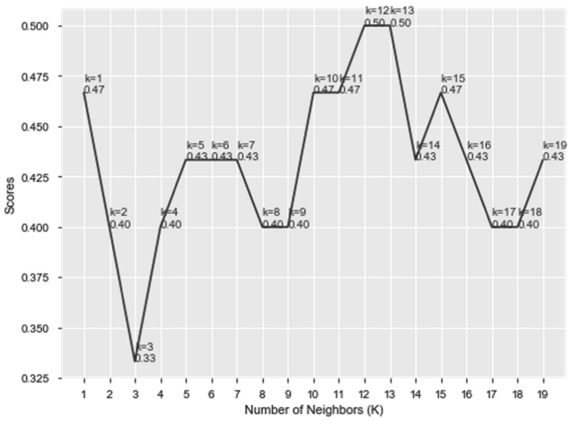

Fig. 3. K Neighbors Classifier accuracy scores for different K values.

Support Vector Classifier
We evaluated the C-Support Vector Classification against different kernels (linear, polynomial, radial-basis, and sigmoid). The "linear" kernel achieved the best accuracy score (0.40) under the penalty parameter $C = 0.3$ (Fig. 4).

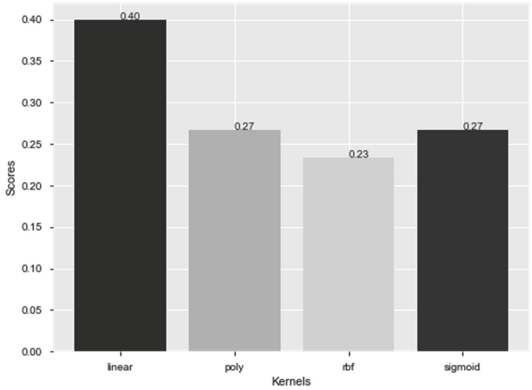

Fig. 4. Support Vector Classifier accuracy scores for different kernels.

Decision Tree Classifier

A decision tree classifier could provide different results when we change the number of features to consider when looking for the best split. Varying them between 1 and the size of all columns of the learning set, the best accuracy (0.50) was for 6, 14, and 15 features (Fig. 5).

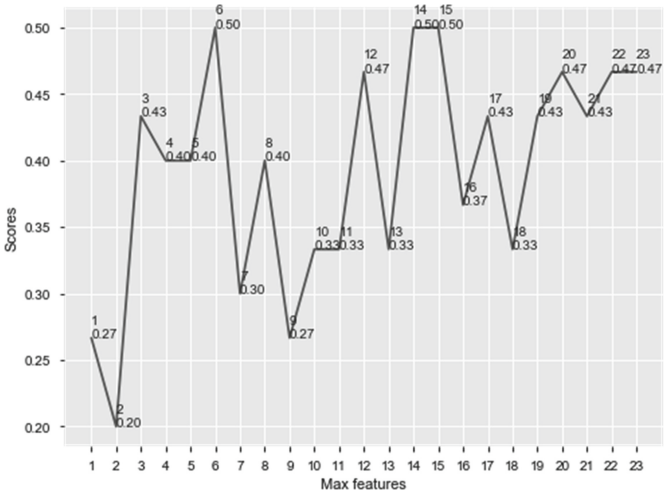

Fig. 5. Decision Tree Classifier accuracy scores for different number of maximum features.

Gradient Boosting Classifier

Gradient Boosting it allows for the optimization of arbitrary differentiable loss functions. We challenged a different number of boosting stages to perform. Gradient boosting is relatively robust to over-fitting, so a large number usually results in better performance.

The scope was between 10 and 1000 estimators, and the highest score (0.43) reveals in the 200 boosting stages (Fig. 6).

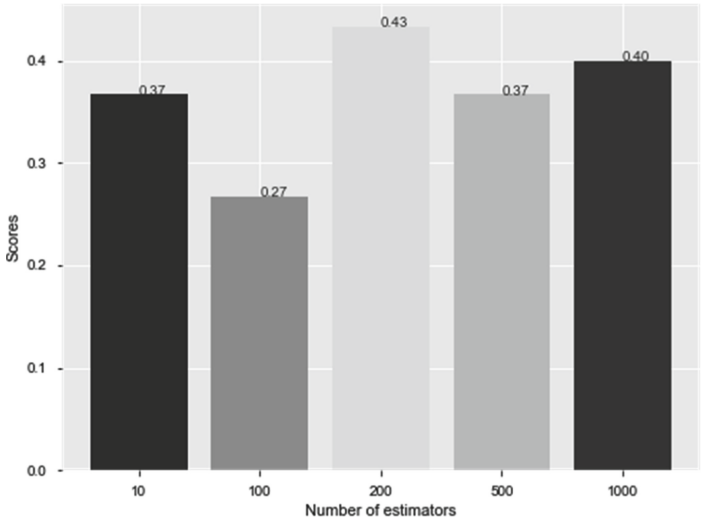

Fig. 6. Gradient Boosting Classifier accuracy scores for different number of estimators.

Random Forest Classifier

In this classifier, we evaluated a parameter which defines the number of trees in the forest. We used Gini impurity as a criterion of the quality of a split. Random Forest Classifier achieved the highest overall accuracy score (0.57) among other machine learning algorithms when the number of estimators exceeded 120 (Table 1, Fig. 7).

Table 1. Confusion matrix based on Random Forest Classifier result.

		Predicted label				
		(−Inf, 19.25)	[19.25, 30.5)	[30.5, 44.)	[44., +Inf)	Accuracy
True label	(−Inf, 19.25)	6	1	1	0	**0.75**
	[19.25, 30.5)	1	7	2	0	**0.70**
	[30.5, 44.)	3	2	3	0	**0.38**
	[44., +Inf)	1	1	1	1	**0.25**

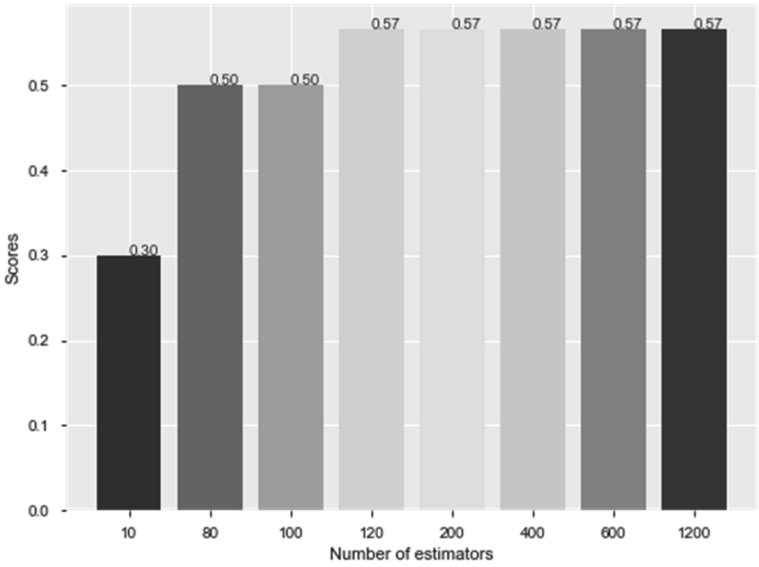

Fig. 7. Random Forest Classifier scores for different number of estimators.

4 Discussion

Examination of the variation of the saccades is a non-invasive method of evaluating the neural networks involved in the control of eye movements. The examples above demonstrated that it is possible to use eye movements as a biomarker for the assessment of symptom progression in PD.

We live in the "age of implementation" of the machine learning models. Commercial companies are acquiring data on a large scale in order to present the content which is best-fitted to the end-user, which influences the businesses. However, there is still an emerging necessity of medical data extension because it is a crucial factor in the context of machine learning. Paradoxically, we have access to much fewer data of the medical records (even about ourselves) than most of the managers of the commercial websites who aggregates about our shopping behavior. The modeled data sample, presented in this article, is relatively abundant in the neuroscience scale. Based on the examinations conducted by neurologists, we were able to create predictions based on the different patient population with different treatments for the most advanced stages of the disease. This results supported with further extended and in-depth research could lead to a new approach in the development of a follow-up tool for PD symptoms. As an outcome, this automated mechanism could provide to a doctor an objective opinion about applied therapy symptoms. Hence we can conclude that when the patient is doing significantly worse than others, their treatment is not optimal and should be changed. However, there is still a large field for the model improvements that should lead to more accurate results. The proposed protocol allowed us to evaluate multiple machine learning models in a relatively agile process of data aggregation. Consequently, more complex variations of saccadic tasks can give insight into higher-order eye movement control [12]. Our work is

an evaluation of well-known models that implement different multiclass strategies, such as K Neighbors Classifier, Support Vector Classifier, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifier in the context of saccades research. In this trial, Random Forest Classifier achieved the highest overall accuracy score, which could lead to a direction in further discoveries in the field of bioinformatics.

5 Conclusions

We believe that the multidisciplinary cooperation between neurologists and information engineering is essential to achieve significant results in the open-science approach. We attempted to predict the longitudinal symptom developments during different treatments based on the neuropsychological data aligned with the parameters of the eye movements. As a training dataset, we used patients from the BMT group (3rd visit), DBS (3rd visit), and POP (1st visit). The independent test set consisted of the POP group from the second visit. For predictions, we used a machine learning framework, written in Python. The best classifier - Random Forest - reached 75% and 70% of accuracy while predicting subclasses of UPDRS for patients in advanced stages of the disease who respond to treatment, with a global 57% accuracy score for all classes. Thanks to collaborative research, we have presented a comparison of different machine learning models that could be useful in the context of bioinformatics. Our direction is to create a new research ecosystem, that would significantly increase (by a factor of 10) the number of attributes and measurements in order to implement deep learning methods.

References

1. Connolly, B.S., Lang, A.E.: Pharmacological treatment of Parkinson disease: a review. JAMA **311**(16), 1670–1683 (2014)
2. Goldenberg, M.M.: Medical management of Parkinson's disease. Pharm. Therapeutics **33**(10), 590 (2008)
3. Thanvi, B.R., Lo, T.C.N.: Long term motor complications of levodopa: clinical features, mechanisms, and management strategies. Postgrad. Med. J. **80**(946), 452–458 (2004)
4. Benabid, A.L., et al.: Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease. Lancet Neurol. **8**(1), 67–81 (2009)
5. Ramaker, C., et al.: Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Mov. Disord. Off. J. Mov. Disord. Soc. **17**(5), 867–876 (2002)
6. Henik, A., et al.: Disinhibition of automatic word reading in Parkinson's disease. Cortex **29**(4), 589–599 (1993)
7. Jones, G.M., DeJong, J.D.: Dynamic characteristics of saccadic eye movements in Parkinson's disease. Exp. Neurol. **31**(1), 17–31 (1971)
8. White, O.B., et al.: Ocular motor deficits in Parkinson's disease: II. Control of the saccadic and smooth pursuit systems. Brain **106**(3), 571–587 (1983)
9. Chan, F., et al.: Deficits in saccadic eye-movement control in Parkinson's disease. Neuropsychologia **43**(5), 784–796 (2005)
10. Przybyszewski, A., et al.: Multimodal learning and intelligent prediction of symptom development in individual Parkinson's patients. Sensors **16**(9), 1498 (2016)
11. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
12. Nij Bijvank, J.A., et al.: A standardized protocol for quantification of saccadic eye movements: DEMoNS. PLoS ONE **13**(7), e0200695 (2018)

sensors

MDPI

Review

Machine Learning and Digital Biomarkers Can Detect Early Stages of Neurodegenerative Diseases

Artur Chudzik [1], Albert Śledzianowski [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; artur.chudzik@pjwstk.edu.pl (A.C.); albert.sledzianowski@pjwstk.edu.pl (A.Ś.)
[2] UMass Chan Medical School, Department of Neurology, 65 Lake Avenue, Worcester, MA 01655, USA
* Correspondence: przy@pjwstk.edu.pl

Abstract: Neurodegenerative diseases (NDs) such as Alzheimer's Disease (AD) and Parkinson's Disease (PD) are devastating conditions that can develop without noticeable symptoms, causing irreversible damage to neurons before any signs become clinically evident. NDs are a major cause of disability and mortality worldwide. Currently, there are no cures or treatments to halt their progression. Therefore, the development of early detection methods is urgently needed to delay neuronal loss as soon as possible. Despite advancements in Medtech, the early diagnosis of NDs remains a challenge at the intersection of medical, IT, and regulatory fields. Thus, this review explores "digital biomarkers" (tools designed for remote neurocognitive data collection and AI analysis) as a potential solution. The review summarizes that recent studies combining AI with digital biomarkers suggest the possibility of identifying pre-symptomatic indicators of NDs. For instance, research utilizing convolutional neural networks for eye tracking has achieved significant diagnostic accuracies. ROC-AUC scores reached up to 0.88, indicating high model performance in differentiating between PD patients and healthy controls. Similarly, advancements in facial expression analysis through tools have demonstrated significant potential in detecting emotional changes in ND patients, with some models reaching an accuracy of 0.89 and a precision of 0.85. This review follows a structured approach to article selection, starting with a comprehensive database search and culminating in a rigorous quality assessment and meaning for NDs of the different methods. The process is visualized in 10 tables with 54 parameters describing different approaches and their consequences for understanding various mechanisms in ND changes. However, these methods also face challenges related to data accuracy and privacy concerns. To address these issues, this review proposes strategies that emphasize the need for rigorous validation and rapid integration into clinical practice. Such integration could transform ND diagnostics, making early detection tools more cost-effective and globally accessible. In conclusion, this review underscores the urgent need to incorporate validated digital health tools into mainstream medical practice. This integration could indicate a new era in the early diagnosis of neurodegenerative diseases, potentially altering the trajectory of these conditions for millions worldwide. Thus, by highlighting specific and statistically significant findings, this review demonstrates the current progress in this field and the potential impact of these advancements on the global management of NDs.

Keywords: neurodegenerative diseases; Alzheimer's disease; Parkinson's disease; digital endpoints; online cognitive testing; eye-tracking; machine learning; early detection; digital phenotyping

Citation: Chudzik, A.; Śledzianowski, A.; Przybyszewski, A.W. Machine Learning and Digital Biomarkers Can Detect Early Stages of Neurodegenerative Diseases. *Sensors* **2024**, *24*, 1572. https://doi.org/10.3390/s24051572

Academic Editor: Urban M. Fietzek

Received: 18 January 2024
Revised: 16 February 2024
Accepted: 27 February 2024
Published: 29 February 2024

1. Introduction

Aging is a significant risk factor for neurodegenerative diseases (NDs) such as Alzheimer's (AD) and Parkinson's (PD), despite advancements in technology that have improved our quality of life and longevity [1–3]. Unfortunately, the complexity of the disease process, involving various contributing factors, presents a challenge in identifying effective remedies [4].

The complexity of NDs lies in a spectrum of disorders characterized by a primary loss of cells, leading to secondary cell loss in other brain regions [5]. Processes correlated with AD begin over 30 years, whereas cognitive changes begin over about 15–11 years, before the first AD symptoms [6,7]. Unfortunately, the prevalence of Alzheimer's Disease-related dementia is fast increasing due to our aging population [8].

Sadly, the prevalence worldwide is estimated to be as high as 24 million; by 2050, the AD number could rise to 139 million worldwide [9]. Currently (Q1'24), there is no cure for AD, as during the first clinical symptoms and neurological diagnosis many parts of the brain are already affected without the possibility to recover.

The second (after AD) most common neurodegenerative disease is Parkinson's Disease (PD). This disease is characterized mainly by motor but also by cognitive disorders [10]. The prevalence of Parkinson's Disease is expected to increase significantly by 2050, with estimates suggesting a doubling of the current number of affected individuals. This is due to a combination of factors, including an aging population, declining smoking rates, and increasing industrialization [11,12]. The economic burden of the disease is also projected to rise, with the cost of medical expenses and indirect costs such as reduced employment expected to increase substantially [11]. These projections highlight the urgent need for innovative treatments and a coordinated global response to address the growing impact of Parkinson's Disease.

It is noteworthy that both Alzheimer's Disease and Parkinson's Disease are neurodegenerative diseases characterized by substantial and irreversible neuronal loss, however in different regions of the brain. Commonly, neurodegeneration begins two to three decades before observed symptoms. Hence, the best chance to fight NDs is to estimate the beginning period of the ND-related brain changes [13].

Recent research on the beginnings of NDs has focused on the role of molecular biomarkers, such as miRNAs (small gene regulators) and exosomes (tiny cellular messengers). It has shown promise in detecting neuronal dysfunction in the presymptomatic stage of NDs [14]. These biomarkers, along with other laboratory and biochemical markers, are being explored for their potential in early diagnosis and disease progression assessment [15]. However, the use of analog biomarkers in Alzheimer's Disease diagnosis has limitations, including the need for biological samples and hospitalization [16].

Despite these challenges, the development of bioassays (sensitive biological detection tests) and the identification of biological markers in blood, plasma, and serum have shown promise in overcoming these limitations [17]. However, the validation of the clinical usefulness of these biomarkers is still incomplete, and further research is needed to standardize their readout and evaluate their performance in detecting early disease [18].

Therefore, the search for an ideal biomarker for Alzheimer's and Parkinson's Disease continues, with the goal of achieving a reliable and accurate diagnosis at the earliest clinical stages. Analog biomarkers, despite their strengths, hold limitations as well. Conversely, digital technologies, which provide objective, high-frequency data, are being considered as a solution to the current subjective measures of NDs [19].

These tools, including AI and remote sensing technologies, are promising avenues for early detection and monitoring of neurodegenerative diseases that can be deployed non-invasively and potentially at scale [20,21]. Yet, there is an ongoing discussion regarding the connection between digital and analog biomarkers and their correlation. This is a gap in the research that still calls for interdisciplinary validation, hence posing a challenge between the medical, IT, legal, and ethical fields.

For that reason, this review aims to provide an interdisciplinary perspective and highlight the challenges in the intersection between technology and medicine, based on recent findings.

2. Biological Definition and Hidden Nature of NDs

Digital biomarkers emerged from analog biomarkers [22]. Hence, to properly interpret the outcomes of digital tools, it is important to understand first the biological definitions

of both diseases, their measurement parameters, their relevance to neurodegenerative diseases, and how AI can enhance the interpretation and application of these biomarkers.

Therefore, we define NDs as a group of diseases that cause the progressive loss of structure or function of neurons, including death of neurons. Despite the fact that Alzheimer's and Parkinson's are in the same cluster of disorders, they exhibit distinct biological characteristics [23].

For example, in Parkinson's Disease, by the time symptoms like *bradykinesia* or tremor became apparent, between 30 and 70% of the substantia nigra (SN—an area of the midbrain) is already irreparably damaged, limiting the actionable time window. The range of 30–70% is commonly cited, but this can vary based on individual cases and the methods used for measurement. As a consequence of this damage, we lose SN neurons, the so-called dopaminergic neurons of the midbrain, which are the main source of dopamine (DA) in the central nervous system (CNS) [24].

Moreover, the impact of neurodegeneration can be noticeable in eye movements (EMs). This is because the substantia nigra influences eye movements among its many roles in movement regulation. The SN helps regulate the action of the superior colliculus (SC) by providing inhibitory input. The SC initiates reflex orienting by sending control signals to the gaze centers in the midbrain, and this process is impaired by the loss of neurons in the SN.

In the context of Alzheimer's Disease, the main affected part of the brain is a region responsible for memory and spatial navigation. AD onset is primarily linked to changes in the hippocampus, which rapidly loses tissue and functional connectivity with other brain regions, leading to atrophy and cognitive impairment [25]. Changes in the hippocampus, including altered neurogenesis, are early events in AD and may worsen memory impairment [26]. Spatial disorientation, a common symptom of AD, is linked to changes in the medial temporal and parietal brain regions [27]. These changes in the hippocampus and associated brain regions are associated with altered gene expression, synaptic excitability, and plasticity, which contribute to memory loss in AD [28].

However, thanks to internal compensation mechanisms, the neurodegeneration that occurs in the meantime is hidden from the outside. This hidden progression of the disease makes early detection a challenge in research, which can hinder the development of prevention strategies [29,30].

3. Biological Definition Determines Parameters for Digital Measurements

There are three kinds of symptoms in NDs, related to different structures in the brain being affected by the disease: cognitive (primarily in AD, secondarily in PD), motor (primarily in PD, less evident in AD), and emotional (observed in both, but characteristic for late-onset AD). Sadly, research has identified a range of non-motor symptoms (NMSs) as well, including cognitive problems, apathy, depression, anxiety, hallucinations, and psychosis, as well as sleep disorders, fatigue, autonomic dysfunction, sensory problems, and pain [31]. Importantly, symptoms can occur in the pre-motor phase of the disease and are not fully addressed by current treatments [32].

To better sense the hidden and ongoing changes in the brain, the intersection of technology and medicine stands out as a perspective direction, especially with the opportunity for wide application that can help in prevention strategies. The importance of those detectors, sensitive enough to catch invisible signs of the disease, is repeatedly underlined in the literature [20,33].

Fortunately, as we present, researchers analyzing NDs have become interested in changes in emotional, cognitive, and EM patterns, deeply examining their parameters for building AI models for disease classification.

Given the narrative nature of our review, our aim is to provide a broad overview rather than a detailed technical analysis. However, recognizing the importance of precision in terminology, we have reviewed the manuscript to ensure that each mention of AI and

ML is contextually accurate, clarifying the distinction between the broader field of AI and the specific application of ML as a subset of AI techniques.

4. Methods

This narrative review synthesizes existing research on the integration of machine learning and digital biomarkers in the diagnosis and monitoring of neurodegenerative diseases. We explore the advancements, challenges, and future directions in this interdisciplinary field, emphasizing the potential of these technologies to support early detection and patient care. The scope encompasses studies related to eye tracking, facial expression analysis, cognitive testing, and other digital phenotyping tools in the context of diseases such as Alzheimer's and Parkinson's.

To capture a comprehensive body of literature, we conducted searches in major scientific databases, including Google Scholar, IEEE Xplore, and PubMed. Our search was supplemented by the use of research tools like Litmaps and Mendeley, which helped in mapping the literature connections and identifying key contributions.

We utilized a combination of keywords and phrases such as "eye tracking", "facial expressions", "cognitive testing", "machine learning," "digital biomarkers," "digital phenotyping," "neurodegenerative diseases," "Alzheimer's Disease," and "Parkinson's Disease." Boolean operators (AND, OR) were used to refine the search, ensuring a focused retrieval of relevant studies.

Our selection process included a diverse range of studies that provide insight into the development and application of digital biomarkers and machine learning in the context of neurodegenerative diseases. We cite articles published between 1937 and 2023, written in English, and focusing on original research. The selection was guided by the relevance of each article to the review's themes, the innovative use of technology in neurodegenerative disease diagnostics, and the potential impact on clinical practices and patient outcomes.

The literature was synthesized to highlight key findings, identify thematic trends, and discuss the implications of integrating digital biomarkers with machine learning for neurodegenerative disease diagnostics. We analyzed the selected studies in the context of their contribution to advancing technology, addressing challenges such as data accuracy, privacy concerns, and the need for validation, as well as their potential for clinical application. This thematic analysis allowed us to draw insights into the current state of the field, identify gaps in the literature, and propose directions for future research.

The review integrates findings from the selected literature into a cohesive narrative assessment, discussing the evolution of digital biomarkers and machine learning in the medical field, with a particular focus on neurodegenerative diseases. By examining the studies through a thematic lens, it provides a comprehensive overview of the landscape, including the technological advancements, methodological challenges, and the ethical considerations of implementing these technologies in clinical settings.

5. Remote Oculomotor and Facial Expression Analysis Indicate Progression of NDs

Eye tracking, an established technique for measuring eye movements, plays a crucial role in understanding neurodegenerative diseases. This technique is employed to record the paths of eye movements, often under controlled conditions such as the "follow the green dot on the screen" task. Camera-based eye trackers are the most common form, but other methods like electrooculography (EOG) are available as well [34]. Eye movement disorders offer a window into early changes in brain computation, especially as they are affected early in neurodegenerative diseases [35]. By studying these movements, researchers can gain insights into how these diseases impact the brain's functionality.

Additionally, when combined with facial expression analysis, eye tracking becomes even more powerful [36]. These methodologies are being studied in various contexts of neurodegenerative diseases to identify early signs and progression markers. The combination of eye tracking and facial expression analysis offers a comprehensive approach to studying NDs, providing a more nuanced understanding of the diseases.

6. Eye Tracking Helps to Determine the Disease Probability

Eye tracking technology has been utilized to differentiate between Parkinson's Disease patients and healthy controls, with promising results. For example, in a recent (2023) study, Brien et al. collected video-based eye tracking data on an interleaved pro/anti-saccade task of 104 PD patients and 106 healthy controls [37]. They used features of saccades, pupil, and blink behavior to predict confidence scores for PD/PD-MCI/PDD diagnosis with the Linear Mixed Model to determine the disease probability by different eye tracking biomarkers. This classifier reached a sensitivity of 0.83 and a specificity of 0.78 and the Receiver Operator Characteristic Area Under the Curve (ROC-AUC) of the classifier was 0.88. The predicted confidence scores were indicative of PD motor and cognitive performance scores. The study's findings demonstrate that eye tracking biomarkers can reliably indicate PD motor and cognitive performance scores, offering a non-invasive method for early disease detection.

Similarly, other studies have explored eye tracking for disease detection. Bejani et al. (2022) also utilized video-based eye tracking to analyze smooth pursuit eye movements in PD patients and controls [38]. A collection of parameters was obtained, including complexity measures based on entropy/regularity, describing the system's dynamic and features for assessing self-similarity. The Support Vector Machine (SVM) was used for classification and for the PD and control groups an accuracy of 0.74 was obtained (sensitivity 0.73, specificity 0.74). This research further supports the idea that eye tracking can provide valuable data in distinguishing PD patients from healthy individuals, emphasizing the role of dynamic system features in diagnosis.

Continuing this trend, additional studies have used similar methodologies with high accuracy. Prashanth et al. (2016) used SVM on a dataset from the Parkinson Progression Markers Initiative. This method applies data mining techniques to attributes considered to be early symptoms of PD, such as cognitive disorders, rapid eye movements, sleep behavior disorders, and cerebrospinal fluid measurements [39]. Researchers achieved 0.96 accuracy, 0.97 sensitivity, and 0.95 specificity in distinguishing early PD patients from a control group. This high level of accuracy underlines the effectiveness of data mining techniques in early PD detection, providing a strong case for the use of eye tracking in clinical assessments.

Beyond PD, eye tracking has shown promise in other neurodegenerative conditions as well. In another research study by Vodrahalli et al. (2022), infrared oculography was used with visuo-motor tasks involving rapid reading of 40 single-digit numbers [40]. They used convolutional neural networks (CNNs) as a classifier with widow-based analysis of recording which includes fixations and saccades. The results were discriminated among diseases impacting EM (like PD), diseases associated with vision loss, and healthy controls (81% accuracy compared with the baseline of 33%). This approach showcases further the versatility of eye-tracking technology in diagnosing a range of conditions, highlighting its diagnostic potential.

Similarly, recent research has focused on combining eye tracking with other diagnostic methods. Belan et al. (2023) investigated the effectiveness of an eye-tracking-assisted visual inference language task in differentiating individuals with mild cognitive impairment (MCI) or Alzheimer's Disease dementia from cognitively unimpaired older adults [41]. The research involved 95 participants, including 49 with MCI, 18 with mild AD dementia, and 28 controls. The authors used a non-parametric repeated measures ANOVA model for verbal answers and a linear mixed model (LMM) or its generalized version for the analysis of eye tracking variables. The results showed significant differences in verbal answers across all diagnostic groups, and eye-tracking parameters successfully discriminated AD from MCI and controls. Analyzing oculomotor behavior alongside language assessments demonstrated increased sensitivity for detecting subtle deficits in the MCI-AD continuum, making it a valuable diagnostic tool. These findings collectively indicate the growing importance of eye tracking in diagnosing NDs, offering a more nuanced and sensitive approach to identifying early stages of diseases like PD and AD (Table 1).

Table 1. List of sensors and their respective domains and metrics.

Sensor	Metrics	Sense—Domain	Results	Reference
Video-based eye tracking	Saccades, pupil, blink behavior	PD detection	Sensitivity: 0.83, Specificity: 0.78, ROC-AUC: 0.88	[37]
Video-based eye tracking	Smooth pursuit eye movements, entropy/regularity, self-similarity	PD detection	Accuracy: 0.74, Sensitivity: 0.73, Specificity: 0.74	[38]
Video-based eye tracking	Cognitive disorders, rapid eye movements, sleep behavior disorders, cerebrospinal fluid measurements	Early PD detection	Accuracy: 0.96, Sensitivity: 0.97, Specificity: 0.95	[39]
Infrared oculography	Fixations, saccades	Distinguishing neurodegenerative diseases	Accuracy: 81% (compared to 33% baseline)	[40]
Eye tracking with visual inference language task	Oculomotor behavior, verbal answer analysis	Differentiating MCI/AD dementia from controls	Significant discrimination between AD, MCI, and controls	[41]

However, eye-tracking has traditionally been considered an expensive research method due to the high cost of commercial eye tracking systems [42]. Fortunately, as presented in the next section, recent advancements have made it possible to develop low-cost remote eye tracking systems that maintain clinically significant parameters.

7. Convoluted Neural Networks Allow Cost Optimization of Eye Tracking

The development of a model for classification using disease biomarkers, such as those obtained from eye tracking, is a crucial step in neurodegenerative disease research. Embedding these models into practical environments, particularly through web-based eye-tracking measurements, facilitates their use in healthcare settings.

Interestingly, while web-based eye-tracking has shown effectiveness, it typically exhibits marginally reduced accuracy and increased data variance compared to laboratory-based devices. This limitation highlights the necessity for model enhancement and methodological refinement to ensure practical applicability and accuracy in real-world settings. Addressing this need, recent research has turned to convolutional neural networks (CNNs) for enhancing eye-tracking methods. Thus, the integration of CNNs in eye tracking technology represents a significant advancement, aiming to improve the precision and reliability of these diagnostic tools, thus making them more suitable for clinical applications in neurodegenerative disease diagnostics.

For example, CNNs have been used to enhance the accuracy of gaze estimation on mobile devices by focusing on facial features. Akinyelu et al. (2022) utilized the face component, gaze features were extracted from the eyes, and the shape and location of the eyes were encoded into the network through a 39-point facial landmark component and a Visual Geometry Group (VGG) convolutional neural network [43]. Different experiments were performed, and the experimental result revealed that 39-point facial landmarks can be used to improve the performance of CNN-based gaze estimation models. The researchers achieved the highest eye tracking accuracy of 0.96 and Mean Square Error (MSE) of 0.01. This approach demonstrates the potential of using facial landmarks to refine CNN-based gaze estimation models, significantly improving performance.

Similarly, another study explored the use of CNNs with webcams for eye tracking. Meng et al. (2017) presented a paper where CNNs were employed in conjunction with

webcams for an eye-tracking approach that relies on the detection of key eye features, including the inner and outer corners, the center of the upper and lower eyelids, and the center of the iris [44]. These feature points were then used to construct six types of original time-varying eye movement signals, reducing reliance on the iris center, especially in low-quality videos. The final step involved training a Behaviors-CNN using these signals to recognize diverse eye movement patterns. This strategy helped to mitigate errors stemming from basic eye movement-type detection and artificial eye movement feature construction. The researchers conducted experiments with their application across various activities and, to assess performance, a visual activity dataset was created using a webcam, comprising nearly 0.5 million frames gathered from 38 subjects. They achieved the highest results in the "reading" category, with a precision of 0.87 and a recall of 0.89. By training a Behaviors-CNN with these signals, the study mitigated errors from basic eye movement-type detection and feature construction, showcasing the adaptability of CNNs in different settings.

Continuing the trend of CNN integration, Gunawardena et al. (2022) have also compared various CNN models for mobile eye tracking. They compared four modern lightweight CNN models (LeNet-5, AlexNet, MobileNet, and ShuffleNet) in search of optimal performance in real-time eye tracking based on video oculometry [45]. The researchers used four lightweight CNN models (LeNet-5, AlexNet, MobileNet, and ShuffleNet) to assess the performance of gaze estimation on mobile devices using the Gaze Capture dataset. To analyze the feasibility of various inference modes—on-device, edge-based, and cloud-based—they conducted an empirical measurement study, quantifying inference time, communication time, and resource consumption. The MobileNet-V3 in this study outperformed in terms of model accuracy with the lowest training MSE after 60 epochs, which was 0.5 ± 0.01, and also providing the lowest response inference time, 17.4 milliseconds, of all evaluated network architectures. The findings also revealed that while cloud-based inference yields faster predictions, the communication time introduces significant latency, eliminating real-time eye tracking based on cloud hosting. On the other hand, the researchers point out that on-device inference is limited by energy and memory consumption, leaving edge-based solutions as the best solution with reasonable response time, memory usage, and energy consumption for eye-tracking applications on mobile devices.

Finally, Rakhmatulin and Duchowski (2020) provide an in-depth analysis of contemporary techniques for webcam-based gaze tracking, offering practical implementations of popular methods [46]. The focus is on exploring various deep neural network models for online gaze monitoring. A novel eye-tracking approach is introduced, enhancing the effectiveness of deep learning methods. The system employs a dual-coordinate system, determining the position of the face relative to the computer through infrared LED detection and the gaze position is obtained through a CNN and a method involving three objects (left, right, and center) for accurate gaze tracking. The implementation demonstrates practical applications by enabling computer interaction control based on gaze.

These findings highlight the effectiveness of edge solutions by balancing response time, memory utilization, and power consumption in eye tracking applications on mobile devices, representing a significant step in optimizing eye tracking technologies using CNNs (Table 2). Hence, we note that the development of digital diagnostic applications in real-time environments is crucial, especially in telemedicine.

Importantly, the nuances introduced by different digital environments can significantly impact diagnostic reliability. Moreover, the approach of not storing registration data enhances privacy and reduces data management requirements, aligning with the expectations of medical experts and patients for efficient and secure medical services.

Therefore, it is vital to consider these digital environments when developing and implementing digital diagnostic tools. The immediate results are not only a convenience but a necessity in telemedicine, underlining the importance of real-time processing capabilities.

Table 2. List of sensors and their respective domains and metrics.

Sensor	Metrics	Sense—Domain	Results	Reference
Facial features and gaze estimation	Facial landmarks, gaze features	Eye tracking accuracy enhancement	Highest accuracy: 0.96, MSE: 0.01	[43]
Webcam-based eye tracking	Key eye features (corners, eyelids, iris center), eye movement signals	Eye movement pattern recognition	Precision: 0.87, Recall: 0.89	[44]
Mobile eye tracking	Lightweight CNN models (LeNet-5, AlexNet, MobileNet, ShuffleNet)	Gaze estimation on mobile devices	Best model accuracy (MobileNet-V3): Training MSE: 0.5 ± 0.01, Response time: 17.4 ms	[45]
Webcam-based gaze tracking	Dual-coordinate system, deep neural network models	Online gaze monitoring	Enhanced accuracy and practicality for computer interaction control	[46]

8. Digital Environments Impact Diagnostics Speed and Reliability

Immediate results from automated diagnostic processes are expected by both medical experts and patients. This aspect is obviously important in telemedicine. When using automated diagnostic processes, both medical experts and patients expect immediate results. Additionally, thanks to this approach, there can be no need to store the data of the registration itself. Moreover, there is growing interest in non-invasive predictors of Alzheimer's and Parkinson's Disease, as seen in the use of webcam-based eye-tracking data for classification.

Therefore, experiments with web-based applications were conducted to demonstrate the feasibility of client-side solutions. For example, Yang et al. in 2021 developed the "WebGazer", a web-based eye-tracking application that was integrated into a widely used JavaScript library (jsPsych) for behavioral research [47]. The procedure and code were modified to minimize calibration/validation efforts and enhance temporal resolution. Testing this approach with a decision-making study on Amazon MTurk, the researchers successfully replicated in-lab findings on the connection between gaze and choice. Notably, there was minimal degradation in spatial or temporal resolution, demonstrating the feasibility of online web-based eye-tracking in behavioral research.

Śledzianowski et al. (2023) experimented with disconnecting the software from the hardware capabilities to compensate for the lack of professional equipment in the patient's household [48,49]. The methodology was executed in an online system utilizing readily available home-grade equipment, yielding results comparable to those obtained with an infrared 60 Hz eye-tracker but with fewer artifacts. The findings indicated that the disparity in EM latency, a crucial parameter for distinguishing patients with Parkinson's Disease, was 16 ms when compared to a laboratory-grade 1000 Hz eye-tracker. It is expected that this approach will play a role in advancing analytic tools for NDs (especially for PD) within computational health, consequently expediting the development of new preventive measures for such conditions.

A similar off-eye-tracker approach is presented by Harisinghani et al. (2023) [50]. The study addresses the growing interest in non-invasive predictors of Alzheimer's Disease by exploring the use of webcam-based eye-tracking data for classification. Their previous successful attempts utilized high-end eye trackers during picture narration and reading tasks. In contrast, this study employs a deep-learning approach to build classifiers using eye-tracking data collected with a webcam. While the webcam gaze classifier does not match the performance of the high-end eye-tracking classifier, it outperforms the majority-class baseline classifier in terms of the AU-ROC. The findings suggest that predictive signals

can be extracted from webcam gaze tracking, offering a promising proof of concept for the potential use of this technology as an affordable alternative to high-end eye trackers in AD detection, despite the need for further exploration.

Moreover, eye tracking in Virtual Reality (VR) technology is being actively explored for remote diagnosis of neurodegenerative diseases [51]. For example, Orlosky et al. (2017) deliberate the need to conduct research in a laboratory environment [52]. This research addresses the challenges in early and accurate diagnosis of neurodegenerative conditions, particularly Parkinson's Disease, stating that current evaluation methods are time consuming, require travel to specialized centers, and may lead to misdiagnosis. The authors present a cost-effective Virtual Reality interface designed for the evaluation and diagnosis of neurodegenerative diseases. Utilizing a VR display with an integrated infrared camera, they created a 3D virtual environment mimicking tasks used in patient evaluations. The virtual tasks were designed to elicit eye movements associated with neurodegenerative diseases. The study involved nine Parkinson's Disease patients and seven healthy controls, testing the system's ability to emulate clinical tasks. Eye tracking algorithms and image enhancement were applied to the recorded eye movements, and evaluation by physicians confirmed three out of four abnormalities. The VR interface demonstrated potential for clinical diagnosis, with physicians rating visualizations as potentially useful.

Beyond VR, there is also a focus on using pupil and oculomotor measures as biomarkers in neurodegeneration. O'Callaghan et al. (2022) focus on using pupil and oculomotor measures as biomarkers to detect changes in the locus coeruleus (LC) [53]. The LC role is synthesizing norepinephrine (noradrenaline) and it is involved in physiological responses to stress and panic. It is also involved in various neural processes including attention, memory, and cognitive functions. The LC's degeneration or dysfunction is associated with several neurodegenerative diseases, including Alzheimer's and Parkinson's. Hence, the study involved Parkinson's Disease patients who underwent a pharmacological challenge with the noradrenergic reuptake inhibitor atomoxetine. Ultra-high field 7T MRI characterized the locus coeruleus, and patients were tested on and off atomoxetine using oculomotor eye-tracking tasks and a learning task with pupillometry. The results indicate that atomoxetine improves cognitive performance and saccadic reaction times, with larger pupil responses correlated with locus coeruleus integrity.

Hence, the findings suggest that pupil and eye tracking measures serve as effective biomarkers for this system and are sensitive to pharmacological interventions, offering potential implications for the early detection and monitoring of subcortical changes in Alzheimer's Disease (Table 3).

Table 3. List of sensors and their respective domains and metrics.

Sensor	Metrics	Sense—Domain	Results	Reference
Web-based eye tracking (WebGazer)	Spatial/temporal resolution	Behavioral research	Minimal degradation in resolution, replication of in-lab findings	[47]
Online system with home-grade equipment	EM latency	Parkinson's Disease diagnostics	Comparable results to infrared eye-tracker with less artifacts, latency disparity: 16 ms	[48,49]
Webcam-based eye tracking	AU-ROC	Alzheimer's Disease classification	Outperforms majority-class baseline classifier, indicates potential for AD detection	[50]

Table 3. *Cont.*

Sensor	Metrics	Sense—Domain	Results	Reference
VR with integrated eye tracking	Eye movement analysis	Evaluation of neurodegenerative diseases	Emulates clinical tasks, confirmed abnormalities, potential for clinical diagnosis	[52]
Pupillometry and oculomotor tasks	Locus coeruleus integrity, cognitive performance	Biomarkers for neurodegeneration	Improved cognitive performance and saccadic reaction times with atomoxetine	[53]

Eye movement measurement offers valuable insights. But it is not the only key to understanding neurodegenerative diseases. Here we can expand our perspective to include facial expressions. This aspect, often affected in NDs, plays a key role in interpreting patients' emotional states, correlates strongly with eye movements, and allows tracking the disease progression.

9. Emotional States Estimations by Facial Expressions Can Indicate Hypomimia

An interesting extension of the EM approach is the analysis of the patient's emotions. The analysis of emotional states in PD with ML methods is not yet common. However, the current findings emphasize a gap in real-world implementations. Emotions can change during neurodegeneration, creating a clear trace of the disease visible, for example in different facial expressions. Here, scientists often mix bio-signals coming from different sources, for example, they study the relationships between the properties of eye movements in various emotional states expressed by facial expressions.

Various studies have explored this multifaceted approach, offering insights into the diagnostic potential of these methods. Importantly, mature open-source projects like OpenFace provide state-of-the-art results for selected facial action unit (AU) recognition [54]. OpenFace utilizes the Facial Action Coding System (FACS), which objectively measures and categorizes facial expressions by breaking them down into distinct movements known as Action Units [55]. It provides two machine learning models: one for determining the presence of an AU and another for describing its intensity on a 5-point scale. This system allows for a detailed analysis of facial expressions, such as identifying the combination of AUs that comprise the expression of "happiness" or "sadness", providing meaningful support for research. The contribution of open-source projects like OpenFace is invaluable in research on neurodegenerative diseases.

Using this tool, Śledzianowski et al. (2021) studied facial emotions (FEs) through the muscle activity model (FACS) and chaos parameters in EMs [56]. The research revealed that parameters associated with chaos exhibited a strong positive correlation with happiness, while linear and noise components were mostly negatively correlated with this emotion. The model proposed simultaneous emotion recognition based on facial landmarks and EMs, providing a significant modification in emotion estimation models. Many nonlinear models have been tested, with the best results achieved by using the K-Nearest Neighbors algorithm with an accuracy of 0.89 with an ROC-AUC score of 0.88, F1 score of 0.89, Precision of 0.85, Sensitivity/Recall of 0.93, and Specificity of 0.82. This study highlights the nuanced relationship between facial expressions and EMs in emotional states, providing a significant modification in emotion estimation models.

Interestingly, further validation was conducted using the Affectiva-MIT Facial Expression Dataset (AM-FED) which included both human and algorithmic classifications [57]. It confirmed that EM chaos emerged as a biomarker of happiness, and it can play a more crucial role than the intensity of specific muscle activity (AU12). It means that true happiness can be detected even in situations when the lower part of a face is covered, which often

happened, e.g., during the COVID pandemic. The model is suggested as an extension for estimating happiness based on facial muscle activity, with potential applications in medical analysis of diseases like PD where facial expressions may be affected by hypomimia. This approach demonstrates the potential of using facial tracking for diagnostic purposes in neurodegenerative diseases.

Others also examined the relationship between facial expressions and PD symptoms. Pegolo et al. (2022), addressing hypomimia in PD, aimed to create a quantitative index, termed the Face Mobility Index (FMI), utilizing a face tracking algorithm based on the Facial Action Coding System [58]. The researchers used OpenFace to detect facial landmarks' positions and extract features. The software considers distances between pairs of facial features to distinguish between healthy individuals and those with PD. The index was also evaluated for emotion classification. The results indicated that FMI effectively quantifies impairment in PD, showing statistically significant differences for all emotions when distances between features are considered and for happiness and anger when FMI is applied. kNN was found to be the optimal technique in the classification with FMI and the results based on the AUC yielded values ranging between 88.9 and 88.4 and F1 scores were between 70.1 and 73. This research contributes to understanding the quantitative aspects of facial impairment in PD, offering new avenues for clinical evaluation.

Moreover, Almutiry et al. (2016) suggest the potential for the automated measurement of day-to-day variations in PD symptoms based on FEs. The research introduces a method to assess facial expressivity for enhancing PD clinical evaluations [59]. Given controversial evidence about PD-related facial impairment, the research aims to explore discriminative and quantitative measures of PD through facial expression analysis. Video clips of eight subjects (four healthy controls, and four PD patients) were recorded over several weeks, focusing on emotion variation. A statistical shape model tracked facial expressions, measuring the amount of expressivity for each subject. The results indicated that movement measures during happiness, disgust, and anger expressions were the most discriminative, with PD patients exhibiting less movement than controls, indeed confirming hypomimia in PD patients.

The impact of hypomimia on social interaction and quality of life has also been addressed in clinical studies. A comprehensive review examined the role of computational analysis techniques in measuring emotional facial expressions in PD patients [60]. Clinical studies often use the Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS), with item 3.2 assessing facial expression. Traditional observer-based scales can be time consuming, prompting the exploration of computational analysis techniques for facial expressions. The paper provides a comprehensive review of these techniques for measuring emotional facial expression in people with PD, emphasizing their clinical applications. Additionally, a pilot experimental work on masked face detection in PD is presented, utilizing a deep learning-based model trained on an NVIDIA GeForce 920M GPU, achieving 85% accuracy on testing images. These findings further emphasize the importance of innovative approaches in detecting and understanding the nuances of facial expressions in PD.

In another study, the authors identified a correlation between impaired facial emotion recognition, detectable through online tools, and neuroanatomical changes that necessitate laboratory examinations [61]. The study is focusing on structural changes in the orbitofrontal cortex (OFC) and the amygdala. Using the Iowa Gambling Task (IGT) and Ekman 60 faces test, 24 early PD patients and 24 controls were assessed. Voxel-based morphometry (VBM) analysis of high-resolution structural magnetic resonance images revealed significant gray matter loss in the right amygdala and bilaterally in the OFC in PD patients. Volumetric analyses did not yield significant differences. Correlations between OFC gray matter volume and test performance suggest that OFC and amygdala degeneration are associated with neuropsychological deficits in early-stage PD. This correlation highlights the importance of structural brain changes in understanding the neuropsychological deficits in early-stage PD.

Notably, the recognition of emotions in speech and graphic representation has also been explored. Lu et al. (2022) investigated the recognition of emotions in speech and the graphic representation of expressions gathered during the learning process [62]. They employed an unsupervised Adversarial Autoencoder for feature extraction and utilized convolution neural networks with Bi-directional Long Short-Term Memory (CNN-BiLSTM), achieving an accuracy of 0.99 in emotion classifications. This approach underlines the potential of advanced machine learning techniques in classifying emotions, offering insights into neurodegenerative disease progression.

Moreover, research on EEG features in cross-subject emotion recognition highlights the significance of various EEG parameters in understanding emotions. Li et al. (2018) examined the significance of 18 types of linear and non-linear EEG features in cross-subject emotion recognition, encompassing various channels, brain regions, rhythms, and feature types [63]. Employing SVM with automatic feature selection methods, they verified the potential of exploring robust EEG features in cross-subject emotion. For the dataset containing physiological signals, they attained a mean recognition accuracy of 0.83 (AUC = 0.9). This research demonstrates the potential of EEG features in providing robust biomarkers for emotional states in neurodegenerative diseases.

An interesting direction of research is also the analysis of emotions based on text (sentiment), which in the case of neurodegenerative diseases rather concerns the early phase of the disease or tests before the medical diagnosis of the disease. Such an example is the study presented in "Deep learning approach to text analysis for human emotion detection from big data", where the authors introduced a Deep Learning-Assisted Semantic Text Analysis (DLSTA) for identifying human emotions in text documents using big data [64]. This method involved an RNN/CNN on text content to generate word embeddings, extract feature vectors, and perform final classification with SVM. For different emotions, they achieved a mean accuracy for prediction of 0.83 and detection between 0.92, with a mean recall of 0.85 and an F-Measure of 0.92. This method's success in detecting emotions from text highlights the potential for early diagnosis and monitoring of neurodegenerative diseases through linguistic analysis.

The research leveraging machine learning and facial expression analysis, including tools like OpenFace and datasets such as AM-FED, demonstrates significant potential in diagnosing and monitoring neurodegenerative diseases by identifying emotional states and facial mobility impairments (Table 4). This leads us to another important area: digital tools that help to spot signs of early changes in the brain.

Table 4. List of sensors and their respective domains and metrics.

Sensor	Metrics	Sense—Domain	Results	Reference
OpenFace	Facial Action Coding System (FACS)	Facial expressions analysis	State-of-the-art results for facial action unit recognition. Provides machine learning models for AU presence and intensity.	[54,55]
Facial Landmarks and EM	Chaos parameters, k-Nearest Neighbors (KNN) algorithm	Emotion estimation models	Strong correlation of chaos parameters with happiness. High accuracy (0.89) and ROC-AUC score (0.88) for emotion recognition.	[56]
AM-FED Dataset	Machine learning, chaos as a biomarker	Happiness estimation	Confirmed EM chaos as a biomarker for happiness, crucial even when the lower face is covered.	[57]

Table 4. *Cont.*

Sensor	Metrics	Sense—Domain	Results	Reference
Face Mobility Index (FMI)	Face tracking, kNN	Facial impairment in PD	Statistically significant differences in facial impairment between healthy individuals and PD patients. AUC values between 88.9 and 88.4, F1 scores between 70.1 and 73.	[58]
Video Clips	Statistical shape model	Day-to-day variations in PD symptoms	Highlighted hypomimia in PD patients through decreased movement in expressions of happiness, disgust, and anger.	[59]
Computational Analysis	Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS)	Emotional facial expressions in PD	Reviewed computational techniques for measuring emotional facial expressions, with a deep learning model achieving 85% accuracy in masked face detection.	[60]
Iowa Gambling Task and Ekman 60 Faces Test	Voxel-based morphometry (VBM)	Neuropsychological deficits in PD	Correlation of OFC and amygdala degeneration with neuropsychological deficits in PD patients.	[61]
Adversarial Autoencoder and CNN-BiLSTM	Emotion recognition in speech and graphic representations	Emotion recognition	Achieved an accuracy of 0.99 in emotion classifications, demonstrating the effectiveness of advanced machine learning techniques.	[62]
EEG Features	SVM with automatic feature selection	Cross-subject emotion recognition	Mean recognition accuracy of 0.83 (AUC = 0.9), highlighting the potential of EEG features in emotional state biomarkers.	[63]
Text Analysis (DLSTA)	Deep Learning-Assisted Semantic Text Analysis	Emotion detection from text	Mean accuracy for emotion prediction of 0.83, with detection accuracy up to 0.92 and mean recall of 0.85.	[64]

10. Digital Tools Allow Cognitive Decline Detection

New evidence shows that AI and digital tools are good ways to find signs of brain diseases early. They might help slow down the diseases [65,66]. These tools can spot diseases that cause dementia long before people start showing clear signs of memory or thinking problems. They can help guide changes in how people live and decide who should join medical studies [65,67]. Furthermore, these tools might even help find these diseases before any symptoms show up, and this could lead to new ways to treat the diseases [68].

In mild cognitive impairment, AI and virtual reality can be used to create predictive models based on digital biomarkers, enabling early detection and interventions [69,70]. Moreover, AI shows the potential to assist with the detection of early-stage dementia, offering cost-effective and objective methods [71]. Furthermore, an increasing array of Medtech tools are being developed to monitor daily routines and behaviors, capturing early cognitive changes indicative of NDs [72].

However, it became evident that the field introduced multiple viewpoints of digital biomarkers that might appear unrelated to or inconsistent with each other, obscuring the clear picture of research perspectives. This critique follows Motahari-Nezhad (2022), who emphasizes the need for high-quality studies and the consideration of methodological criteria and evidence quality [73].

As well, Sobolev (2021) highlights the potential of digital biomarkers in predicting and influencing health conditions but also underscores the need for technological integration and the challenges in this area [74]. Similarly, Babrak (2019) discusses the basic differences and similarities between traditional and digital biomarkers and points out the necessity for synchronization and unique feature definition claiming, therefore, that the unclear definition of digital biomarkers, population groups, and their intersection with traditional biomarkers hinders their discovery and validation [22]. Finally, and recently, Alonso et al. (2023) even concluded that there is no consensus about what this emerging term *means* [75].

Additionally, it's interesting to note that the approach to diagnosing Alzheimer's Disease has evolved over the years. Previously, AD was diagnosed when the patient exhibited both cognitive changes and symptoms related to dementia. However, the US National Institute on Aging and the Alzheimer's Association (NIA-AA) now defines AD as a diagnosis based solely on biological biomarkers, even if clinical symptoms are not present yet. This means that AD can be diagnosed much earlier than before. The FDA has developed a staging system for AD, which ranges from stage 1 (normal cognition and biomarker evidence of AD) to stages 4–6 (mild, moderate, and severe dementia).

Importantly, these findings collectively underscore the necessity for a coherent definition of those approaches, at least, to reveal the potential of digital biomarkers and AI methods in Medtech applications.

Hence, to characterize the tandem of digital biomarkers and AI tools, the literature suggests using the term "digital phenotyping", where the assumption is that an individual's health experience is reflected in the digital traces that they leave behind [76]. These traces are further translated as data collected from everyday interactions with technology like smartphones and wearable devices. This approach is notably transforming the way we understand and detect NDs, leveraging everyday technology to gather valuable health data [77].

However, the effectiveness of these digital tools as reliable screening mechanisms remains underdeveloped despite their potential and advancements in the AI field. Those challenges were presented in several studies and projects that explored the use of digital technology in detecting and monitoring neurodegenerative diseases, though with varying degrees of success [78]. Here, we must emphasize the importance of good user experience (UX) in cognitive screening tools, not only for engagement but also for the accuracy and reliability of the collected data.

11. User Experience Impacts Data Reliability

The necessity of a good user experience in cognitive screening tools is crucial, not only for user engagement but also for the accuracy and reliability of the data collected. One notable project in this domain is the Altoida initiative, which focuses on identifying digital biomarkers for mild cognitive impairment [79].

The Altoida model, based on app data, demonstrated high accuracy (AUC = 0.92) in predicting the transition from MCI to dementia [80]. This accuracy was further validated in subjects with MCI due to AD, corroborated by positive beta-amyloid and imaging biomarkers [81]. This high level of accuracy is significant as it showcases the potential of mobile-based applications in the pre-symptomatic detection of NDs. Altoida's recognition by regulatory bodies like the FDA and CE further highlights its clinical relevance [82].

Similarly, another study focused on the Smart Aging Serious Game (SASG) demonstrated the effectiveness of digital platforms in cognitive health assessment [83,84]. The project included a virtual reality platform designed for the ecological assessment of mild neurocognitive disorders. With a focus on mild cognitive impairment and vascular cognitive impairment (VCI), conditions associated with a heightened risk of developing dementia, the research aimed to validate SASG's diagnostic capabilities [84].

Importantly, the SASG was successful in identifying the distinct cognitive profiles associated with MCI and VCI, aligning with traditional neuropsychological assessments. An ROC analysis indicated that SASG and MoCA both demonstrated high diagnostic sensitivity and specificity (AUC values greater than 0.80) for identifying VCI versus HC

and MCI versus HC. The classification accuracy for distinguishing these groups was high, with RF and LR analyses showing between 75% and 91% accuracy [84]. Hence, the SASG proved to be an effective digital phenotyping tool that aligns with traditional neuropsychological evaluations, capable of early and potentially self-administered assessment of cognitive impairments.

What is noteworthy in that context is a study by Illiadou et al. (2021) which was conducted with the administration of a self-administered test in a VR (the "Virtual Supermarket Test (VST)") environment combined with a cheap and commercially available wearable EEG. It is possible to register elevated EEG rhythms in the MCI group when solving tasks in virtual reality, which therefore may be associated with an overall cognitive decline [85].

Interestingly, the SASG example further supports the necessity of a good user experience in cognitive screening tools. Here, usability is crucial, not only for user engagement but also for the accuracy and reliability of the data collected. A recent study (2022) by Zygouris et al. evaluated the usability of the Virtual Supermarket Test [86]. Twenty-four older adults with subjective cognitive decline and thirty-three patients with MCI completed the VST and subsequently assessed its usability using the System Usability Scale (SUS). The results were notable, with an average SUS score of 83.11 out of 100 points (SD = 14.6), indicating a high level of user-friendliness. This score is particularly significant as it remained consistent regardless of the participant's age, educational background, familiarity with touch devices, or MCI diagnosis. Furthermore, there was a notable correlation between the SUS score and VST performance ($r = -0.496$, $p = 0.000$), suggesting that better usability is associated with more accurate cognitive assessment [86].

Therefore, the development of cognitive screening tools like the VST must prioritize UX design [87,88]. Good UX not only facilitates broader accessibility and usability among diverse user groups but also ensures the integrity and validity of the data gathered, which is essential for accurate diagnosis and effective monitoring of conditions like MCI.

Hence, but not surprisingly, user-centered design and evaluation are crucial in clarifying and adjusting the level of detail in web-based decision aids for individuals with MCI [89]. This is important because usability studies have shown that users with MCI require more time and help to complete tasks and that the speed of audio help can significantly impact performance [90].

However, these findings have further implications for tools that can detect (with submillisecond precision) slowness of reaction time, serving as potential biomarkers for bradykinesia (slowness of movement). Thus, while it is important to make web-based tools easy for people with MCI to use, these tools can also detect slow reaction times (Table 5). This slowness is a key sign of bradykinesia, which is often seen in neurodegenerative diseases.

Table 5. List of sensors and their respective domains and metrics.

Sensor	Metrics	Sense—Domain	Results	Reference
Altoida App	Digital biomarkers	Mild cognitive impairment (MCI) detection	AUC = 0.92	[79–81]
VR (Smart Aging Serious Game-SASG)	Cognitive health assessment	MCI and Vascular Cognitive Impairment (VCI) diagnosis	Diagnostic sensitivity and specificity (AUC > 0.80), Accuracy: 75–91%	[83,84]
VR (Virtual Supermarket Test) and Wearable EEG	EEG rhythms	Cognitive decline in MCI	Elevated EEG rhythms associated with cognitive tasks in VR	[85]
VR (Virtual Supermarket Test) and System Usability Scale (SUS)	Usability assessment	Usability and cognitive assessment accuracy	SUS score: 83.11/100, Correlation between SUS score and VST performance: r = −0.496	[86]

12. Digital Tools Can Detect Slowness of Reaction Time

Interestingly, slower task completion by users with mild cognitive impairment is not just a challenge but somehow obviously an important feature that can be leveraged for early detection.

These findings align with the study by Donoghue et al. (2012), which showed an association between functional mobility, as measured by the Timed Up-and-Go (TUG) test, and various cognitive functions. Their analysis revealed that slower TUG performance is independently associated with poorer global cognition, executive function, memory, and slower processing speed [91].

Importantly, simple reaction time (SRT) decline is visible not only in advanced stages of AD or PD but also in individuals with MCI (a risk factor for PD and AD), therefore being possible early digital biomarkers of those diseases. This is supported by a meta-analysis by Andriuta (2019), who analyzed seven selected studies with a total of 327 participants with MCI and 468 healthy controls (HCs); the mean SRT was significantly (p = 0.0217) longer in the MCI group (by 11%) than in the HC group [92]. This observation is crucial because slower response times, often consistent with cognitive impairment, can be precisely captured on web platforms.

Web platforms can measure reaction times and task completion speeds in milliseconds, turning into a part of digital phenotype. This finding is further supported by our recent study (2023) that involved collecting cognitive and behavioral data from PD patients and healthy controls [93]. Here, an online platform helped to collect the time between the screen appearing and the participant's first option selection (instrumental reaction time—IRT) and the time it takes for the participant to click the submit button (time to submit—TTS). The key finding was that IRT and TTS were significantly slower ($p < 0.001$) in the PD group compared to healthy controls, especially in tests like the MoCA and Epworth Sleepiness Scale and beyond the healthy age-related decline of the reaction time [93].

Moreover, research in the literature highlights the significant potential of online platforms and mobile devices in measuring reaction times and task completion speeds, key indicators of cognitive and behavioral changes. Already in 2011, Cinaz et al. explored wearable devices as tools for measuring reaction times, particularly in the context of everyday cognitive functioning, and concluded that they are feasible to measure changes in reaction times [94]. That is supported by Burke (2017), who demonstrated that devices like the Apple iPad and iTouch can accurately measure reaction times [95]. Bonnechère (2022) extended this work by using cognitive mobile games to assess the evolution of reaction times across different tasks [96]. These studies collectively underscore the versatility and practicality of digital devices in cognitive monitoring.

This finding aligns with the research by Roos J. Jutten et al. (2022), which investigated fluctuations in reaction time (RT) performance as a marker of early amyloid-related neurodegeneration in preclinical Alzheimer's Disease. Their study employed computerized cognitive testing to measure intraindividual variability in RT (IIV-RT) over some months, using tasks of varying complexity. They found that greater IIV-RT, especially in complex RT tasks, was associated with steeper cognitive decline in individuals with elevated AD biomarkers [97]. Therefore, incorporating these metrics into clinical assessments can provide valuable insights into the early stages of neurodegenerative diseases, enhancing diagnostic strategies and patient care.

However, Schatz (2015) emphasized the need for careful validation and calibration of these digital tools, particularly tablet-based devices, to ensure accurate reaction time assessments [98]. This highlights an essential consideration in the development and application of these technologies. Nevertheless, when introduced properly, even the casual card game Klondike Solitaire may measure time-related values (like time spent on thinking of a move) that can be useful in distinguishing games played by older people with MCI from their healthy peers (AUC > 0.877), as presented by Gielis (2021) [99].

Therefore, the slower performance seen in PD patients is not a drawback but a useful clue about the disease's progress, helping build a 'digital phenotype' that improves early di-

agnosis. In line with this, digital versions of tests like the MoCA, Epworth, or TMT can add time-related measurements, making digital tools like user-friendly mobile apps valuable for spotting early signs of diseases like PD or AD (Table 6). Moving to a broader context, web and mobile technologies stand out as affordable options for checking cognitive health.

Table 6. List of sensors and their respective domains and metrics.

Sensor	Metrics	Sense—Domain	Results	Reference
Timed Up-and-Go (TUG) test	Functional mobility, global cognition, executive function, memory, processing speed	Cognitive impairment detection	Slower TUG performance associated with poorer cognitive functions	[91]
Web Platforms	Simple reaction time (SRT)	Early detection of cognitive decline	MCI group showed 11% longer SRT than healthy controls ($p = 0.0217$)	[92]
Online Platforms	Instrumental reaction time (IRT), time to submit (TTS)	Parkinson's Disease cognitive and behavioral data	IRT and TTS significantly slower in PD group ($p < 0.001$)	[93]
Wearable Devices and Mobile Devices	Reaction times in daily cognitive functioning	Cognitive monitoring	Wearables and devices like iPad/iTouch feasible for measuring reaction times	[94,95]
Cognitive Mobile Games	Evolution of reaction times across tasks	Cognitive health monitoring	Mobile games used to assess and monitor reaction time changes	[96]
Computerized Cognitive Testing	Intraindividual variability in reaction time (IIV-RT)	Early amyloid-related neurodegeneration	Greater IIV-RT associated with steeper cognitive decline in preclinical AD	[97]
Tablet-based Devices	Reaction time assessment	Validation and calibration of digital tools	Need for careful validation to ensure accurate assessments	[98]
Digital Card Game (Klondike Solitaire)	Time spent on thinking of a move	Differentiating cognitive impairment	Can distinguish MCI from healthy peers (AUC > 0.877)	[99]

13. Web and Mobile Technology Are Affordable Tools for Screening Cognitive Deficits

Recent findings suggest that in the context of the early detection of cognitive decline, the potential of mobile applications is substantial. As Thabtah (2020) noted, some of these apps employ advanced techniques like machine learning and AI to enhance diagnostic accuracy [100]. This approach expands access to early cognitive health assessments, benefiting a broader population segment, including those in underserved areas.

Consequently, digital applications (utilizing smartphones and wearables) hold significant promise as accessible, affordable, and equitable tools for screening cognitive deficits. Chinner (2018) and Naslund (2017) highlight how these tools, by leveraging widely available technology, become practical even in resource-limited settings and offer an affordable alternative to traditional diagnostic methods [101,102]. These methods are particularly effective in supporting clinical care and promoting treatment adherence, especially in low- and middle-income countries.

Moreover, the web-based cognitive testing approach presented in "cCOG: A web-based cognitive test tool for detecting neurodegenerative disorders" explores the effectiveness

of a web-based cognitive tool in detecting mild cognitive impairment and dementia for standardized screening in neurodegenerative disorders [103]. The computerized test battery was based on the three classical cognitive tasks: a modification of the wordlist test, a simple reaction task, and the Trail Making Test. It was divided into seven tasks to complete in approximately 20 min to complete with a keyboard and mouse or a touchscreen. Analyzing clinical data from three European cohorts, including 306 cognitively normal, 120 MCI, and 69 dementia subjects, the study compared the global cognitive scores derived from standard neuropsychological tests. The tool demonstrates accuracies (ROC-AUC) ranging from 0.71 to 0.84 for MCI and 0.86 to 0.94 for dementia when administered both at the clinic and in the home environment. Hence, the results indicate this tool as a promising and cost-effective tool for MCI and dementia detection through home-based cognitive assessments.

Web and mobile technologies, through the use of machine learning and AI in apps and web-based tools, offer affordable, accessible screening options for cognitive deficits, showing promise in the early detection of neurodegenerative disorders with demonstrated effectiveness even in resource-limited settings (Table 7). Understandably, the growing use of mental health applications also brings to the fore challenges concerning evidence-based guidelines and transparency.

Table 7. List of sensors and their respective domains and metrics.

Sensor	Metrics	Sense—Domain	Results	Reference
Mobile applications	Machine learning and AI techniques	Early detection of cognitive decline	Enhances diagnostic accuracy, expands access	[100]
Smartphones and wearables	Accessibility and affordability	Screening cognitive deficits	Practical in resource-limited settings, supports clinical care	[101,102]
Web-based cognitive testing (cCOG)	Wordlist test, simple reaction task, Trail Making Test	Detecting MCI and dementia	ROC-AUC: 0.71–0.84 for MCI, 0.86–0.94 for dementia	[103]

14. Quality of Digital Health Can Be Enhanced through Standardization

The issue regarding guidelines and transparency is highlighted by Torous (2017), who emphasizes the need for transparency and trust in the evaluation and use of these applications, given the concerns about their proliferation and the lack of established guidelines [104]. That statement was further supported by de Francisco Carvalho (2019), who pointed out the necessity of clear ethical guidelines on how information can and should be used [105]. Thus, while digital applications offer immense potential in the early detection and diagnosis of cognitive deficits, ensuring their effective and ethical implementation requires addressing challenges such as evidence-based validation, transparency, and user trust.

These challenges also include the crucial aspects of data privacy and user engagement. Ensuring data privacy is vital for patient trust and legal compliance, while maintaining user engagement is essential for consistent and reliable data collection. Beyond these practical considerations, ethical implications present additional tension. As Ford, Milne, and Curlewis (2023) discuss, deploying digital biomarkers and AI at scale raises concerns regarding accuracy, bias, and equitable access [106]. This is important because to leverage digital tools effectively in clinical settings, those challenges need to be addressed urgently.

These concerns are critical, as they directly impact the validity and fairness of the diagnostic process. Hence, while digital applications offer immense possibilities in the early detection and monitoring of cognitive health, their successful implementation centers on addressing these multifaceted challenges, ensuring they can be effectively integrated into healthcare practice.

In response, recent studies suggest that a consensus on standards in digital health research could markedly improve the quality of these studies. The lack of consensus on digital outcome measures and their integration into clinical practice and research has been noted by other experts, prompting various proposals for standardization. This is underlined by Bejani et al. (2023), who, moreover, offer strategic insights for the inclusion of patient-centered digital measures in research [38].

Furthermore, the main critique by Jha et al. (2023), termed the "Single Digital Biomarker Hypothesis," revolves around the tendency of current digital measures to simplify the complexity of PD into a single severity score [107]. The authors argue that PD is a multi-dimensional, multi-etiologic syndrome that cannot be adequately represented by a single number. The nuances of individual patient experiences, which can vary widely in symptoms and progression, are likely to be overlooked by such a reductive approach.

Moreover, Espay et al. (2016) discuss the challenges and opportunities in utilizing technology for PD diagnosis and treatment [108]. This is an outcome of the vast amount of data collected and its limited clinical application. Additionally, challenges include the incompatibility of technology platforms, the need for widespread deployment among elderly patients, and the complexity of translating big data into clinically relevant insights. As a solution, Espay et al. (2019) formed The International Parkinson and Movement Disorders Society Task Force on Technology, which aims to address these issues by promoting the development of open-source and open-hardware platforms for multichannel data capture [109]. The goal is to create adaptable systems for individualized treatment delivery, encouraging early detection, tailored therapy, and subgroup targeting for testing disease-modifying treatments and identifying objective biomarkers for improved longitudinal tracking of PD.

Still, Assunção (2022) argues for the adoption of a more comprehensive framework. This framework would assess and communicate the benefits of early interventions, crucially bridging the gap between the development of disease-modifying therapies (DMTs) and their practical, clinical applications [30].

An additional critical observation within the literature echoed in practical applications is the disappearing patient interest in digital solutions over time. It is hypothesized that such tools may overly focus on the disease rather than the patient, leading to a perceived lack of value and progressing disengagement. As articulated by Bloem et al. (2020), the patient centricity of digital tools is imperative for sustained patient engagement [110].

Hence, while standardized, patient-centric approaches in digital health tools enhance interventions and improve outcomes in neurodegenerative diseases, this might be a good moment to take a step back and reconsider the algorithms we use. To further refine these tools, we should delve deeper into the brain's underlying mechanisms, drawing inspiration from how our minds work, which, as Turing's theories and recent psychophysical experiments suggest, might be more about simple rules and pattern recognition than complex, computational-heavy, AI-driven 'black box' methods.

15. Machine Learning Models Support Diagnosis and Monitoring of NDs

It is interesting to note that the brain's functioning is often compared to that of a digital computer or a Universal Turing Machine, which processes symbols [111]. However, psychophysical experiments and our ability to recognize complex objects, such as faces, in various contexts and lighting conditions suggest otherwise. This argues against symbolic representation and instead supports the idea that concept representation based on similarities may be a more appropriate model for how the brain works.

Hence, we propose to direct our eyes to Turning's lesser-known contribution to the field of developmental biology. Turing proposed that natural patterns like stripes, spots, and spirals can arise naturally from the interaction of two or more chemical substances, which he called "morphogens" (that is, the movement of "chemicals" between cells that causes cells to transform/morph into the next "state") [112]. This research explores how

complicated patterns, like those seen in zebra stripes, can emerge from relatively elementary biochemical processes.

Applying this concept to brain development, Turing's theory suggests that complex structures and patterns in the brain could emerge from simple, preprogrammed rules at the cellular level. This perspective contrasts with the view of the brain as a Universal Turing Machine, which implies a more fixed, predetermined computational process. Importantly, the shift from viewing the brain as a rigid, symbol-processing Universal Turing Machine to a more fluid, self-organizing system, as suggested by Turing's morphogenetic principles, allows for a more nuanced understanding of cognitive processes.

Interestingly, this approach resonates with the research of Levin et al. (2021) who created the first *living robots*, known as xenobots [113]. Levin's research explores how cells can self-organize into complex structures and forms using basic rules. Using xenobots, he presents how individual cells self-organize into complex tissues and morphologies. This is important in the context of NDs, because changes in cellular patterns and processes could be detectable before clinical symptoms arise, enabling earlier intervention and potentially more effective treatment. Additionally, understanding how cells communicate and organize themselves to regenerate tissue can inform strategies to promote neural regeneration in neurodegenerative diseases.

Therefore, we suggest that the idea of using logical rules for object classification resonates with Turing's reaction-diffusion theory and Levin's work on morphogenesis. Both emphasize the emergence of complex patterns and structures from simpler rules and interactions. In the brain, these 'simpler rules' could be the logical rules used in Rough-Set Theory for visual processing.

16. Rough Rules Can Implement Visual System Principia

Interestingly, the anatomical and neurophysiological basis of object shape classification and the computational properties of the brain can be described by Rough-Set Theory (RST). Introduced by Pawlak (1982), the RST offers a framework for understanding how the brain processes complex visual information, despite the imprecise nature of concepts representing objects' physical properties [114,115]. This suggests that concept representation based on similarities rather than symbols may be a more accurate model for how the brain works [116].

The application of the RST in visual classification involves specific neural interactions. The visual classification model is based on the receptive field properties of neurons in different visual areas and uses both feedforward and feedback interactions between them.

- The feedforward pathways use "driver logical rules" to combine properties extracted in each area into hypotheses related to possible objects.
- The feedback pathways use "modulator logical rules" to help change weak concepts of objects' physical properties into crisp classifications in psychophysical space [117].

Hence, this process approximates how the brain utilizes logical rules to transform blurred object concepts into clear categorizations, mirroring the principles of the RST (Figure 1).

In higher visual brain areas, the described processes utilize Granular Computing (GrC) to identify upper and lower approximations of the retinal image. These approximations are then compared with different objects' models (images) stored in the visual cortex. As object recognition or classification progresses, the lower visual areas are tuned to extract the properties of the selected model (modulator logical rules), and the gap between the upper and lower approximations narrows. If the border set becomes empty, the object is successfully recognized. Hence, this approach can be applied to propose various models that approximate the actual (future) cognitive state of tested subjects.

Figure 1. The receptive field properties of neurons fed into both feedforward and feedback pathways. Feedforward pathways use "driver logical rules" to create hypotheses of possible objects, while feedback pathways employ "modulator logical rules" to refine these hypotheses into crisp classifications. Higher visual brain areas then utilize Granular Computing (GrC) to compare upper and lower approximations of retinal images with stored object models, leading to object recognition as the gap between these approximations narrows.

Importantly, this theory is further supported by research on macaques. Przybyszewski et al. (2000) investigated the back-projection pathways from the striate cortex (V1) to the lateral geniculate nucleus (LGN) [118]. These are connections that go from one part of the brain back to an area that supplies input to it. Research shows that these back connections make neuron responses stronger, depending on the contrast of what is seen. These findings revealed that the way neurons in the LGN react to visual stimuli is greatly increased by these back-projection pathways. This happens especially when the contrast in the visuals is high. Hence, it shows that the connections from the striate cortex to the LGN play a big part in how well neurons respond to different levels of contrast and colors in what we see.

Therefore, these new insights fit well with the Rough-Set Theory-based models of visual processing. These predictive models, developed based on clinical studies, reveal various patterns of neurodegenerative diseases, analogous to the visual representation of complex objects in higher visual cortical areas.

However, given the variability in symptoms among patients, it is necessary to have adaptive mechanisms that can accommodate the approximate and flexible nature of these variations. Hence, a critical question remains: how can these mechanisms identify objects or their elements in new conditions, especially in the context of identifying diseases with unseen preliminary indications?

To address this challenge, a novel approach has been proposed, which extends the classical definition of the receptive field (RF) to a fuzzy detector. The properties of the

RF are further defined by the computational attributes of the bottom-up and top-down pathways, which compare the stimulus against multiple predictions.

17. Fuzzy Detectors as Possible Predictive Models of NDs

By using this new approach, it is possible to detect and recognize objects in different conditions, including those that are unseen. This approach can also be used to identify the presence of neurodegenerative diseases in patients, even when the symptoms vary between individuals. Hence, a fuzzy detector is a promising tool for developing more accurate and effective predictive models for neurodegenerative diseases, which can help to improve diagnosis and treatment.

That is an insightful observation because the utilization of AI methods to generalize and intellectualize patients' symptoms could be a game-changer in detecting preclinical indications of neurodegenerative diseases such as PD and AD. By creating models based on granular computing approaches, AI can use these models as references to classify potential preclinical indications of ND. This approach could open new possibilities for preventing or curing neurodegenerative brain pathologies, potentially leading to more effective treatments for these devastating conditions.

The aim is to utilize the insights of movement disorder specialists to identify significant attributes of NDs for AI-based classification. Therefore, a project by Przybyszewski et al. (2021) employing a combination of Theory of Mind (ToM) and supervised learning, based on granular computing (GrC) and Rough-Set Theory, seeks to emulate the expertise of movement disorder specialists. This system is based on a computing approach that uses Rough-Set Theory and abstract granules to represent the ToM of many movement disorder experts. It identifies similarities between granules obtained from one group of more advanced PD patients to estimate the disease progression of other patients. While it was found that the accuracy of prediction increased with disease progression, it is noted that divergent sets of granules characteristic of different parts of the brain might degenerate in different ways with disease progression [117]. The exploration of various AI methods further underscores this field's potential.

To challenge this method, researchers have compared different AI methods, including GrC as implemented in Rough-Set Theory (RST) and fuzzy Rough-Set Theory (FRST) [119]. These methods were compared with other classical machine learning techniques for predicting PD progression, including Nearest Neighbors, Decision Tree, Random Forest, Support Vector Classification, and Gradient Boosting. Dutta and Skowron (2021) have used complex granules (c-granules) to model longitudinal disease development, finding that the RST provides the best estimations of disease progression [120]. The study found that the RST gave the best estimations of disease progressions measured as accuracies, while the FRST gave smaller values of accuracies but better global coverage. Other AI methods gave similar results but only when looking for the disease progression without or with medication separately, indicating their limitations in looking for longitudinal PD progressions [121].

It is fascinating to see the potential of machine learning models, including Rough Sets, in predicting the optimal treatment for NDs. A further example is a study by Przybyszewski et al. (2020) analyzing patients under different treatments, which achieved varying degrees of accuracy, suggesting the potential to discover universal rules for PD progression [122]. The research presents an overall accuracy of 70% for medication visits, 56% for DBS (deep brain stimulation), and 67% and 79% for post-op second and third visits, respectively. This exploration highlights the exciting possibilities of AI and machine learning in advancing our understanding and management of neurodegenerative diseases, potentially leading to more tailored and effective treatments (Table 8).

Table 8. List of sensors and their respective domains and metrics.

Sensor	Metrics	Sense—Domain	Results	Reference
Theory of Mind (ToM) and Granular Computing (GrC)	Granules representing expertise	PD progression prediction	Accuracy increases with disease progression; identifies granule differences in brain degeneration	[117]
AI Methods Comparison (GrC, RST, FRST)	Disease progression modeling	PD progression prediction	RST showed best accuracy; FRST provided better global coverage; limitations in longitudinal progression with other AI methods	[120]
Machine Learning Models	Treatment outcome prediction	Optimal treatment for NDs	Accuracies: 70% for medication, 56% for DBS, 67% and 79% for post-op visits	[122]

18. Intelligent Granular Computing (IGrC) Can Predict Cognitive Patterns

Cognitive symptoms are more dominant in Alzheimer's Disease, whereas motor symptoms are more pronounced in Parkinson's Disease. A study using an IGrC approach examined the relationship between cognitive and motor symptoms in PD, revealing that cognitive changes are independent of motor symptom development [123].

The study involved 47 Parkinson's Disease patients who underwent eye movement, neurological, and neuropsychological tests in two sessions: S#1—without medications (MedOFF) and S#2 after taking medications (MedON). The patients were divided into two groups: Gr1 (less advanced) and Gr2 (more advanced). By applying different sets of rules to different visits, the researchers found that cognitive changes are independent of the motor symptoms' development.

It is interesting to note that there seems to be a link between depressive symptoms and neurodegenerative diseases. In the case of Parkinson's Disease, for example, there is a reduction in the level of the reward hormone dopamine, which can lead to a decrease in positive emotional experiences and an increase in depression. On the other hand, in older individuals (over 65 years of age) who experience depression, there is a higher risk of developing late-onset Alzheimer's Disease (LOAD). The research aims to evaluate this relationship's strength.

AI and machine learning, including Rough-Set Theory, have been employed to classify patients' symptoms in PD. The study involved testing 24 patients with Parkinson's Disease (PD) who were only receiving medical treatment (BMT-group) and 23 patients with PD who were receiving both medical treatment and deep brain stimulation (DBS-group), as well as 15 older patients who had been receiving DBS treatment for one and a half years (POP-group). The patients were tested every six months (W1, W2, W3) using a range of neurological (disease duration, Unified Parkinson's Disease Rating Scale), neuropsychological (depression-Beck test, PDQ39, Epworth), and eye movement (reflexive saccadic) tests. Using RST, the researchers were able to identify rules from the BMTW1 data (patients receiving only medical treatment during their first visit) that allowed them to predict UPDRS scores for BMTW2 and BMTW3 with accuracies of 0.765 (0.7 without Beck test of depression result) and 0.8 (0.7 without Beck result), respectively [123].

They were also able to use these rules to predict disease progression (UPDRS) in a group of patients in the DBSW1 group with an accuracy of 0.765. Using the rules generated from the DBSW2 data, the researchers were able to predict UPDRS scores for the DBSW3 (acc. = 0.625), POPW1 (acc. = 0.77), POPW2 (acc. = 0.5), and POPW3 (acc. = 0.33) groups [124]. By adding the depression attribute and using GrC, the researchers were able

to make more accurate predictions about disease progression in a range of patient groups compared to predictions made without this attribute (Table 9).

Table 9. List of sensors and their respective domains and metrics.

Sensor	Metrics	Sense—Domain	Results	Reference
IGrC Approach	Cognitive and motor symptoms	PD symptom classification	Cognitive changes independent of motor symptoms development	[123]
AI and Machine Learning (RST)	Neurological, neuropsychological, eye movement tests	PD symptom classification and depression's impact	UPDRS prediction accuracies: BMTW2 = 0.765, BMTW3 = 0.8, DBSW1 = 0.765, DBSW3 = 0.625, POPW1 = 0.77, POPW2 = 0.5, POPW3 = 0.33 (improved with depression attribute)	[123]

19. AI and Machine Learning Can Predict Symptoms and Progression of NDs

Recent findings suggest that AI methods predict cognitive patterns in normal subjects, indicating pre-dementia stages. For example, Przybyszewski et al. (2022) used granular computing rules to classify cognitive data from the BIOCARD study, which has been ongoing for over 20 years with 354 normal subjects. The study's findings suggest that AI methods can predict patterns in cognitive attributes of normal subjects that might indicate their pre-dementia stage, something that may not be visible to neuropsychologists [125].

Another study based on Biocard data provides a significant advancement in the detection and prediction of Alzheimer's Disease, utilizing AI methods to identify early cognitive changes. Over 20 years, subjects were evaluated annually to determine their cognitive status—normal, mild cognitive impairment, or dementia. The study used the Clinical Dementia Rating Sum of Boxes (CDRSUM) as a quantitative index for assessing mild dementia and developed rough set rules (RSR) for classification. Researchers classified patients of AD, MCI, and normal, based on their CDRSUM scores. They discovered that some subjects showed signs of potential cognitive impairment or mild dementia that were not evident to neuropsychologists. These findings highlight the capacity of AI methods to detect subtle cognitive changes that might indicate a pre-dementia stage [126].

This approach is a critical step forward in the early detection of AD. By identifying patterns in cognitive attributes among normal subjects, AI methods can reveal early signs of dementia, offering a window for intervention before the condition becomes clinically apparent.

Another BIOCARD study utilizes multi-granular computing to refine the process of classifying cognitive data related to Alzheimer's Disease, aiming for early detection [127]. Researchers modified the number of attributes used in the BIOCARD study, increasing the variety of granules from five to seven attributes, compared to the constant fourteen attributes used previously. This allowed for a more nuanced comparison of classification results. The focus was also on the interpretability of the rules obtained from different granular levels. By creating rules with varying granularity and algorithms, the researchers aimed to identify classifications that are both complete and consistent across different rule sets. The goal is to develop a more accurate and reliable system for early diagnosis, which is critical for effective intervention. Researchers are testing various models to determine the most effective ones for identifying different stages of diseases like Alzheimer's and Parkinson's, and they seek classifications that remain consistent irrespective of the algorithms used [127]. Therefore, the overarching aim is to develop a more precise and reliable diagnostic system for early intervention.

Overall, the above studies highlight the potential of digital biomarkers and AI in detecting the early stages of neurodegenerative diseases like Alzheimer's and Parkinson's (Table 10). The integration of digital tools into clinical practice could revolutionize the way we diagnose and treat these conditions, ultimately improving the quality of life for millions of people worldwide.

Table 10. List of sensors and their respective domains and metrics.

Sensor	Metrics	Sense—Domain	Results	Reference
Granular Computing Rules	Cognitive data classification	Pre-dementia stage detection	AI methods predicted pre-dementia stages in normal subjects from BIOCARD study	[125]
Clinical Dementia Rating Sum of Boxes (CDRSUM) and rough set rules (RSR)	Early cognitive changes	Alzheimer's Disease detection	Identified subjects with potential cognitive impairment not evident to neuropsychologists	[126]
Multi-Granular Computing	Cognitive data related to Alzheimer's Disease	Early detection of Alzheimer's Disease	Developed more accurate and reliable system for early diagnosis with varying granularity	[127]

20. Conclusions

AI and digital tools promise to facilitate the early detection of neurodegenerative diseases, potentially leading to earlier interventions that could slow disease progression. Studies are being conducted to validate the efficacy of digital biomarkers and AI-based predictive models in identifying early-stage neurodegenerative diseases. These technologies could lead to more personalized treatment plans based on individual data patterns.

This review covers the use of digital phenotyping technologies to capture behavioral information relevant to neurological diseases. Digital endpoints could enhance the precision of clinical trials, aiding in patient stratification and the detection of treatment effects. However, the review also points out the need for a standardized approach to study design to allow for meaningful interpretation of the data collected from various studies.

This standardization would help in comparing results across studies and improving the understanding of neurodegenerative disease trajectories. Interestingly, despite reported advancements in healthcare solutions, it was found that the current digital and mobile health (mHealth) applications are in urgent need of improved functionalities to assist in both patient care management and early diagnosis of NDs.

Nevertheless, the findings presented in this paper support the idea that telemedicine solutions could lead to earlier identification of at-risk individuals and may be more sensitive to disease progression, which is beneficial for discovering disease-modifying treatments. Within this framework, sensors emerge as a fundamental component, highlighting the effectiveness of digital phenotyping in enhancing disease characterization and monitoring. The integration of clinical scales, imaging, biosamples, and digital tools is suggested as the most effective approach for characterizing and monitoring disease [33].

However, the role of digital tools in improving care, research, and outcomes for patients with movement disorders suggests that their full potential is yet to be realized. The care model is perceived as being reactive rather than proactive, with an inadequate response to complex issues due to a lack of disease-specific expertise and underutilization of non-pharmacological treatments.

Moreover, treatment plans are usually more disease centric rather than patient centric, not fully considering the needs and preferences of the individuals affected by these

conditions. Consequently, patients may find themselves excluded from the clinical decision-making process, leaving a gap that must be addressed urgently to bring truly meaningful research outcomes.

Author Contributions: Conceptualization: A.W.P., A.C. and A.Ś.; methodology: A.C. and A.W.P.; validation: A.Ś. and A.W.P.; investigation: A.C., A.Ś. and A.W.P.; writing—original draft preparation: A.C.; writing—review and editing: A.C.; visualization, A.C.; supervision, A.W.P.; Every author has written different parts of this review. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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

References

1. Zabel, C.; Nguyen, H.P.; Hin, S.C.; Hartl, D.; Mao, L.; Klose, J. Proteasome and oxidative phoshorylation changes may explain why aging is a risk factor for neurodegenerative disorders. *J. Proteom.* **2010**, *73*, 2230–2238. [CrossRef] [PubMed]
2. Donmez, G. Aging and Neurodegeneration. *J. Mol. Genet. Med.* **2013**, *7*, 1000071. [CrossRef]
3. Azam, S.; Haque, E.; Balakrishnan, R.; Kim, I.-S.; Choi, D.-K. The Ageing Brain: Molecular and Cellular Basis of Neurodegeneration. *Front. Cell Dev. Biol.* **2021**, *9*, 683459. [CrossRef] [PubMed]
4. Magalingam, K.B.; Radhakrishnan, A.; Ping, N.S.; Haleagrahara, N. Current Concepts of Neurodegenerative Mechanisms in Alzheimer's Disease. *BioMed Res. Int.* **2018**, *2018*, 3740461. [CrossRef] [PubMed]
5. Rossor, M.N. Parkinson's disease and Alzheimer's disease as disorders of the isodendritic core. *BMJ* **1981**, *283*, 1588–1590. [CrossRef] [PubMed]
6. Lane-Donovan, C.; Herz, J. ApoE, ApoE Receptors, and the Synapse in Alzheimer's Disease. *Trends Endocrinol. Metab.* **2017**, *28*, 273–284. [CrossRef]
7. Beason-Held, L.L.; Goh, J.O.; An, Y.; Kraut, M.A.; O'Brien, R.J.; Ferrucci, L.; Resnick, S.M. Changes in Brain Function Occur Years before the Onset of Cognitive Impairment. *J. Neurosci.* **2013**, *33*, 18008–18014. [CrossRef] [PubMed]
8. Lewis, S.J.; Gangadharan, S.; Padmakumar, C.P. Parkinson's disease in the older patient. *Clin. Med.* **2016**, *16*, 376–378. [CrossRef]
9. Reitz, C.; Mayeux, R. Alzheimer disease: Epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochem. Pharmacol.* **2014**, *88*, 640–651. [CrossRef]
10. Aarsland, D.; Kurz, M.W. The Epidemiology of Dementia Associated with Parkinson's Disease. *Brain Pathol.* **2010**, *20*, 633–639. [CrossRef]
11. Kowal, S.L.; Dall, T.M.; Chakrabarti, R.; Storm, M.V.; Jain, A. The current and projected economic burden of Parkinson's disease in the United States. *Mov. Disord.* **2013**, *28*, 311–318. [CrossRef]
12. Dorsey, E.R.; Bloem, B.R. The Parkinson Pandemic—A Call to Action. *JAMA Neurol.* **2018**, *75*, 9–10. [CrossRef]
13. Al-Chalabi, A. Preventing neurodegenerative disease. *Brain* **2021**, *144*, 1279–1280. [CrossRef]
14. Doroszkiewicz, J.; Groblewska, M.; Mroczko, B. Molecular Biomarkers and Their Implications for the Early Diagnosis of Selected Neurodegenerative Diseases. *Int. J. Mol. Sci.* **2022**, *23*, 4610. [CrossRef]
15. Menéndez-González, M. Biomarkers in neurodegenerative disorders: Translating research into clinical practice. *Front. Aging Neurosci.* **2014**, *6*, 281. [CrossRef]
16. Mobed, A.; Hasanzadeh, M. Biosensing: The best alternative for conventional methods in detection of Alzheimer's disease biomarkers. *Int. J. Biol. Macromol.* **2020**, *161*, 59–71. [CrossRef]
17. Schneider, P.; Hampel, H.; Buerger, K. Biological Marker Candidates of Alzheimer's Disease in Blood, Plasma, and Serum. *CNS Neurosci. Ther.* **2009**, *15*, 358–374. [CrossRef]
18. Frisoni, G.B.; Boccardi, M.; Barkhof, F.; Blennow, K.; Cappa, S.; Chiotis, K.; Démonet, J.-F.; Garibotto, V.; Giannakopoulos, P.; Gietl, A.; et al. Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. *Lancet Neurol.* **2017**, *16*, 661–676. [CrossRef] [PubMed]
19. Dorsey, E.R.; Papapetropoulos, S.; Xiong, M.; Kieburtz, K. The First Frontier: Digital Biomarkers for Neurodegenerative Disorders. *Digit. Biomark.* **2017**, *1*, 6–13. [CrossRef] [PubMed]
20. Mari, Z.; Mestre, T.A. The Disease Modification Conundrum in Parkinson's Disease: Failures and Hopes. *Front. Aging Neurosci.* **2022**, *14*, 810860. [CrossRef] [PubMed]
21. Piendel, L.; Vališ, M.; Hort, J. An update on mobile applications collecting data among subjects with or at risk of Alzheimer's disease. *Front. Aging Neurosci.* **2023**, *15*, 1134096. [CrossRef]

22. Babrak, L.M.; Menetski, J.; Rebhan, M.; Nisato, G.; Zinggeler, M.; Brasier, N.; Baerenfaller, K.; Brenzikofer, T.; Baltzer, L.; Vogler, C.; et al. Traditional and Digital Biomarkers: Two Worlds Apart? *Digit. Biomark.* **2019**, *3*, 92–102. [CrossRef]
23. Pathak, N.; Vimal, S.K.; Tandon, I.; Agrawal, L.; Hongyi, C.; Bhattacharyya, S. Neurodegenerative Disorders of Alzheimer, Parkinsonism, Amyotrophic Lateral Sclerosis and Multiple Sclerosis: An Early Diagnostic Approach for Precision Treatment. *Metab. Brain Dis.* **2022**, *37*, 67–104. [CrossRef]
24. Zampese, E.; Surmeier, D.J. Calcium, Bioenergetics, and Parkinson's Disease. *Cells* **2020**, *9*, 2045. [CrossRef]
25. Rao, Y.L.; Ganaraja, B.; Murlimanju, B.V.; Joy, T.; Krishnamurthy, A.; Agrawal, A. Hippocampus and its involvement in Alzheimer's disease: A review. *3 Biotech* **2022**, *12*, 55. [CrossRef] [PubMed]
26. Mu, Y.; Gage, F.H. Adult hippocampal neurogenesis and its role in Alzheimer's disease. *Mol. Neurodegener.* **2011**, *6*, 85. [CrossRef] [PubMed]
27. Vlček, K.; Laczó, J. Neural correlates of spatial navigation changes in mild cognitive impairment and Alzheimer's disease. *Front. Behav. Neurosci.* **2014**, *8*, 81685. [CrossRef]
28. Saura, C.A.; Parra-Damas, A.; Enriquez-Barreto, L. Gene expression parallels synaptic excitability and plasticity changes in Alzheimer's disease. *Front. Cell. Neurosci.* **2015**, *9*, 318. [CrossRef] [PubMed]
29. Aisen, P.S.; Jimenez-Maggiora, G.A.; Rafii, M.S.; Walter, S.; Raman, R. Early-stage Alzheimer disease: Getting trial-ready. *Nat. Rev. Neurol.* **2022**, *18*, 389–399. [CrossRef]
30. Assunção, S.S.; Sperling, R.A.; Ritchie, C.; Kerwin, D.R.; Aisen, P.S.; Lansdall, C.; Atri, A.; Cummings, J. Meaningful benefits: A framework to assess disease-modifying therapies in preclinical and early Alz-heimer's disease. *Alzheimers Res. Ther.* **2022**, *14*, 54. [CrossRef] [PubMed]
31. Rektorova, I.; Aarsland, D.; Chaudhuri, K.R.; Strafella, A.P. Nonmotor Symptoms of Parkinson's Disease. *Park. Dis.* **2011**, *2011*, 351461. [CrossRef]
32. Zis, P.; Erro, R.; Walton, C.C.; Sauerbier, A.; Chaudhuri, K.R. The range and nature of non-motor symptoms in drug-naive Parkinson's disease patients: A state-of-the-art systematic review. *NPJ Park. Dis.* **2015**, *1*, 15013. [CrossRef]
33. Mammen, J.R.; Speck, R.M.; Stebbins, G.M.; Müller, M.L.; Yang, P.T.; Campbell, M.; Cosman, J.; Crawford, J.E.; Dam, T.; Hellsten, J.; et al. Mapping Relevance of Digital Measures to Meaningful Symptoms and Impacts in Early Parkinson's Disease. *J. Park. Dis.* **2023**, *13*, 589–607. [CrossRef]
34. Duchowski, A. Eye Tracking Techniques. In *Eye Tracking Methodology*; Springer: London, UK, 2003; pp. 51–59. [CrossRef]
35. Hamedani, A.G.; Gold, D.R. Eyelid Dysfunction in Neurodegenerative, Neurogenetic, and Neurometabolic Disease. *Front. Neurol.* **2017**, *8*, 329. [CrossRef]
36. Marandi, R.Z.; Gazerani, P. Aging and eye tracking: In the quest for objective biomarkers. *Futur. Neurol.* **2019**, *14*, FNL33. [CrossRef]
37. Brien, D.C.; Riek, H.C.; Yep, R.; Huang, J.; Coe, B.; Areshenkoff, C.; Grimes, D.; Jog, M.; Lang, A.; Marras, C.; et al. Classification and staging of Parkinson's disease using video-based eye tracking. *Park. Relat. Disord.* **2023**, *110*, 105316. [CrossRef]
38. Bejani, M.; Luque-Buzo, E.; Gomez-García, J.; Burlaka-Petrash, A.; Grandas, F.; Godino-Llorente, J. Detection of Parkinson's Disease by Analysis of Smooth Pursuit Eye Movements and Machine Learning. *Mov. Disord.* **2023**, *37* (Suppl. S2), S184–S185. Available online: https://www.mdsabstracts.org/abstract/detection-of-parkinsons-disease-by-analysis-of-smooth-pursuit-eye-movements-and-machine-learning/ (accessed on 21 December 2023).
39. Prashanth, R.; Roy, S.D.; Mandal, P.K.; Ghosh, S. High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning. *Int. J. Med Inform.* **2016**, *90*, 13–21. [CrossRef] [PubMed]
40. Vodrahalli, K.; Filipkowski, M.; Chen, T.; Zou, J.; Liao, Y.J. Predicting Visuo-Motor Diseases from Eye Tracking Data. In Proceedings of the Pacific Symposium on Biocomputing, Big Island, HI, USA, 3–7 January 2022; pp. 242–253.
41. Belan, A.F.R.; Pais, M.V.; Camargo, M.v.Z.d.A.; Sant'ana, L.C.F.G.; Radanovic, M.; Forlenza, O.V. Diagnostic Performance of an Eye-Tracking Assisted Visual Inference Language Test in the Assessment of Cognitive Decline due to Alzheimer's Disease. *J. Alzheimer's Dis.* **2023**, *94*, 1105–1119. [CrossRef] [PubMed]
42. Narcizo, F.B.; de Queiroz, J.E.R.; Gomes, H.M. Remote Eye Tracking Systems: Technologies and Applications. In Proceedings of the 2013 26th Conference on Graphics, Patterns and Images—Tutorials, Arequipa, Peru, 5–8 August 2013; pp. 15–22.
43. Akinyelu, A.A.; Blignaut, P. Convolutional Neural Network-Based Technique for Gaze Estimation on Mobile Devices. *Front. Artif. Intell.* **2022**, *4*, 796825. [CrossRef] [PubMed]
44. Meng, C.; Zhao, X. Webcam-Based Eye Movement Analysis Using CNN. *IEEE Access* **2017**, *5*, 19581–19587. [CrossRef]
45. Gunawardena, N.; Ginige, J.A.; Javadi, B.; Lui, G. Performance Analysis of CNN Models for Mobile Device Eye Tracking with Edge Computing. *Procedia Comput. Sci.* **2022**, *207*, 2291–2300. [CrossRef]
46. Rakhmatulin, I.; Duchowski, A.T. Deep Neural Networks for Low-Cost Eye Tracking. *Procedia Comput. Sci.* **2020**, *176*, 685–694. [CrossRef]
47. Yang, X.; Krajbich, I. Webcam-based online eye-tracking for behavioral research. *Judgm. Decis. Mak.* **2021**, *16*, 1485–1505. [CrossRef]
48. Śledzianowski, A.; Nowacki, J.P.; Sitarz, K.; Przybyszewski, A.W. Universal Machine-Learning Processing Pattern for Computing in the Video-Oculography. In Proceedings of the International Conference on Computational Science, Prague, Czech Republic, 3–5 July 2023; pp. 200–212. [CrossRef]

49. Śledzianowski, A.; Nowacki, J.P.; Sitarz, K.; Przybyszewski, A.W. Novel Machine Learning Pipeline for Real-Time Oculometry. In *Recent Challenges in Intelligent Information and Database Systems. ACIIDS 2023*; Communications in Computer and Information Science; Springer: Cham, Switzerland, 2023; Volume 1863, pp. 498–509. [CrossRef]
50. Harisinghani, A.; Sriram, H.; Conati, C.; Carenini, G.; Field, T.; Jang, H.; Murray, G. Classification of Alzheimer's using Deep-learning Methods on Webcam-based Gaze Data. *Proc. ACM Hum. Comput. Interact.* **2023**, *7*, 1–17. [CrossRef]
51. Przybyszewski, A.W.; Śledzianowski, A.; Chudzik, A.; Szlufik, S.; Koziorowski, D. Machine Learning and Eye Movements Give Insights into Neurodegenerative Disease Mechanisms. *Sensors* **2023**, *23*, 2145. [CrossRef]
52. Orlosky, J.; Itoh, Y.; Ranchet, M.; Kiyokawa, K.; Morgan, J.; Devos, H. Emulation of Physician Tasks in Eye-Tracked Virtual Reality for Remote Diagnosis of Neurodegenerative Disease. *IEEE Trans. Vis. Comput. Graph.* **2017**, *23*, 1302–1311. [CrossRef] [PubMed]
53. O'callaghan, C.; Hezemans, F.H.; Ye, R.; Orlando, I.; Passamonti, L.; Rowe, J.B. Pupil and eye tracking measures as a tool for detection and intervention in neurodegeneration. *Alzheimer's Dement.* **2022**, *18*, e065889. [CrossRef]
54. Baltrusaitis, T.; Zadeh, A.; Lim, Y.C.; Morency, L.P. Openface 2.0: Facial behavior analysis toolkit. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 59–66.
55. Allen, M. (Ed.) Facial Action Coding System. In *The SAGE Encyclopedia of Communication Research Methods*; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2017. [CrossRef]
56. Sledzianowski, A.; Urbanowicz, K.; Glac, W.; Slota, R.; Wojtowicz, M.; Nowak, M.; Przybyszewski, A. Face emotional responses correlate with chaotic dynamics of eye movements. *Procedia Comput. Sci.* **2021**, *192*, 2881–2892. [CrossRef]
57. Śledzianowski, A.; Nowacki, J.P.; Przybyszewski, A.W.; Urbanowicz, K. Detecting True and Declarative Facial Emotions by Changes in Nonlinear Dynamics of Eye Movements. In *Intelligent Information and Database Systems. ACIIDS 2022*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2022; Volume 13757, pp. 106–116. [CrossRef]
58. Pegolo, E.; Volpe, D.; Cucca, A.; Ricciardi, L.; Sawacha, Z. Quantitative Evaluation of Hypomimia in Parkinson's Disease: A Face Tracking Approach. *Sensors* **2022**, *22*, 1358. [CrossRef]
59. Almutiry, R.; Couth, S.; Poliakoff, E.; Kotz, S.; Silverdale, M.; Cootes, T. Facial behaviour analysis in Parkinson's disease. In *Medical Imaging and Augmented Reality. MIAR 2016*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2016; Volume 9805, pp. 329–339. [CrossRef]
60. Sonawane, B.; Sharma, P. Review of automated emotion-based quantification of facial expression in Parkinson's patients. *Vis. Comput.* **2021**, *37*, 1151–1167. [CrossRef]
61. Ibarretxe-Bilbao, N.; Junque, C.; Tolosa, E.; Marti, M.; Valldeoriola, F.; Bargallo, N.; Zarei, M. Neuroanatomical correlates of impaired decision-making and facial emotion recognition in early Parkinson's disease. *Eur. J. Neurosci.* **2009**, *30*, 1162–1171. [CrossRef]
62. Lu, X. Deep Learning Based Emotion Recognition and Visualization of Figural Representation. *Front. Psychol.* **2022**, *12*, 818833. [CrossRef]
63. Li, X.; Song, D.; Zhang, P.; Zhang, Y.; Hou, Y.; Hu, B. Exploring EEG Features in Cross-Subject Emotion Recognition. *Front. Neurosci.* **2018**, *12*, 162. [CrossRef]
64. Guo, J. Deep learning approach to text analysis for human emotion detection from big data. *J. Intell. Syst.* **2022**, *31*, 113–126. [CrossRef]
65. Frey, A.; Karran, M.; Jimenez, R. Harnessing the potential of digital technologies for the early detection of neurodegenerative diseases. *EDoN* **2019**. [CrossRef]
66. Sabbagh, M.N.; Boada, M.; Borson, S.; Doraiswamy, P.M.; Dubois, B.; Ingram, J.; Iwata, A.; Porsteinsson, A.P.; Possin, K.L.; Rabinovici, G.D.; et al. Early Detection of Mild Cognitive Impairment (MCI) in an At-Home Setting. *J. Prev. Alzheimer's Dis.* **2020**, *7*, 171–178. [CrossRef] [PubMed]
67. Gold, M.; Amatniek, J.; Carrillo, M.C.; Cedarbaum, J.M.; Hendrix, J.A.; Miller, B.B.; Robillard, J.M.; Rice, J.J.; Soares, H.; Tome, M.B.; et al. Digital technologies as biomarkers, clinical outcomes assessment, and recruitment tools in Alzheimer's disease clinical trials. *Alzheimer's Dementia: Transl. Res. Clin. Interv.* **2018**, *4*, 234–242. [CrossRef] [PubMed]
68. DeKosky, S.T.; Marek, K. Looking Backward to Move Forward: Early Detection of Neurodegenerative Disorders. *Science* **2003**, *302*, 830–834. [CrossRef] [PubMed]
69. Cavedoni, S.; Chirico, A.; Pedroli, E.; Cipresso, P.; Riva, G. Digital Biomarkers for the Early Detection of Mild Cognitive Impairment: Artificial Intelligence Meets Virtual Reality. *Front. Hum. Neurosci.* **2020**, *14*, 245. [CrossRef] [PubMed]
70. Piau, A.; Wild, K.; Mattek, N.; Kaye, J. Current State of Digital Biomarker Technologies for Real-Life, Home-Based Monitoring of Cognitive Function for Mild Cognitive Impairment to Mild Alzheimer Disease and Implications for Clinical Care: Systematic Review. *J. Med Internet Res.* **2019**, *21*, e12785. [CrossRef]
71. Li, R.; Wang, X.; Lawler, K.; Garg, S.; Bai, Q.; Alty, J. Applications of Artificial Intelligence to Aid Detection of Dementia: A Narrative Review on Current Capabilities and Future Directions. April 2021. Available online: https://arxiv.org/abs/2104.14073v1 (accessed on 3 December 2023).
72. Hackett, K.; Giovannetti, T. Capturing Cognitive Aging in Vivo: Application of a Neuropsychological Framework for Emerging Digital Tools. *JMIR Aging* **2022**, *5*, e38130. [CrossRef]

73. Motahari-Nezhad, H.; Al-Abdulkarim, H.; Fgaier, M.; Abid, M.M.; Péntek, M.; Gulácsi, L.; Zrubka, Z. Digital Biomarker–Based Interventions: Systematic Review of Systematic Reviews. *J. Med. Internet Res.* **2022**, *24*, e41042. [CrossRef]
74. Sobolev, M.; Gullapalli, B.T.; Rahman, T. Advancing the science of digital biomarkers. In Proceedings of the 2021 Workshop on Future of Digital Biomarkers, Virtual Event, 25 June 2021; ACM: New York, NY, USA, 2021; pp. 1–2. [CrossRef]
75. Alonso, A.K.M.; Hirt, J.; Woelfle, T.; Janiaud, P.; Hemkens, L.G. Definitions of digital biomarkers: A systematic mapping of the biomedical literature. *medRxiv* **2023**. [CrossRef]
76. Milne, R.; Costa, A.; Brenman, N. Digital phenotyping and the (data) shadow of Alzheimer's disease. *Big Data Soc.* **2022**, *9*. [CrossRef]
77. Andrea, A.; Agulia, A.; Serafini, G.; Amore, M. Digital biomarkers and digital phenotyping in mental health care and prevention. *Eur. J. Public Health* **2020**, *30*, 1080. [CrossRef]
78. Klimova, B.; Valis, M.; Kuca, K. Potential of mobile technologies and applications in the detection of mild cognitive impairment among older generation groups. *Soc. Work. Health Care* **2017**, *56*, 588–599. [CrossRef] [PubMed]
79. Tort-Merino, A.; Tarnanas, I.; Bügler, M.; Harms, R.; Fernandez-Villullas, G.; Juncà-Parella, J.; Bosch, B.; Lladó, A.; Sanchez-Valle, R.; Balasa, M. ALTOIDA-iADL for the diagnosis of Mild Cognitive Impairment and early Alzheimer's disease. *Alzheimer's Dement.* **2021**, *17*, e057982. [CrossRef]
80. Rai, L.; Boyle, R.; Brosnan, L.; Rice, H.; Farina, F.; Tarnanas, I. Digital Biomarkers Based Individualized Prognosis for People at Risk of Dementia: The AltoidaML Multi-site Ex-ternal Validation Study. In *GeNeDis 2018: Computational Biology and Bioinformatics*; Springer: Cham, Switzerland, 2020; pp. 157–171. [CrossRef]
81. Kourtis, L.C.; Regele, O.B.; Wright, J.M.; Jones, G.B. Digital biomarkers for Alzheimer's disease: The mobile/wearable devices opportunity. *NPJ Digit. Med.* **2019**, *2*, 9. [CrossRef] [PubMed]
82. Buegler, M.; Harms, R.L.; Balasa, M.; Meier, I.B.; Exarchos, T.; Rai, L.; Boyle, R.; Tort, A.; Kozori, M.; Lazarou, E.; et al. Digital biomarker-based individualized prognosis for people at risk of dementia. *Alzheimer's Dementia Diagn. Assess. Dis. Monit.* **2020**, *12*, e12073. [CrossRef] [PubMed]
83. Manera, V.; Petit, P.D.; Derreumaux, A.; Orvieto, I.; Romagnoli, M.; Lyttle, G.; David, R.; Robert, P.H. "Kitchen and cooking", a serious game for mild cognitive impairment and alzheimer's disease: A pilot study. *Front. Aging Neurosci.* **2015**, *7*, 134267. [CrossRef]
84. Isernia, S.; Cabinio, M.; Di Tella, S.; Pazzi, S.; Vannetti, F.; Gerli, F.; Mosca, I.E.; Lombardi, G.; Macchi, C.; Sorbi, S.; et al. Diagnostic Validity of the Smart Aging Serious Game: An Innovative Tool for Digital Phenotyping of Mild Neurocognitive Disorder. *J. Alzheimer's Dis.* **2021**, *83*, 1789–1801. [CrossRef]
85. Iliadou, P.; Paliokas, I.; Zygouris, S.; Lazarou, E.; Votis, K.; Tzovaras, D.; Tsolaki, M. A Comparison of Traditional and Serious Game-Based Digital Markers of Cognition in Older Adults with Mild Cognitive Impairment and Healthy Controls. *J. Alzheimer's Dis.* **2021**, *79*, 1747–1759. [CrossRef] [PubMed]
86. Zygouris, S.; Segkouli, S.; Triantafyllidis, A.; Giakoumis, D.; Tsolaki, M.; Votis, K.; Tzovaras, D. Usability of the Virtual Supermarket Test for Older Adults with and without Cognitive Impairment. *J. Alzheimer's Dis. Rep.* **2022**, *6*, 229–234. [CrossRef] [PubMed]
87. Zhuang, L.; Yang, Y.; Gao, J. Cognitive assessment tools for mild cognitive impairment screening. *J. Neurol.* **2021**, *268*, 1615–1622. [CrossRef]
88. Contreras-Somoza, L.M.; Irazoki, E.; Toribio-Guzmán, J.M.; de la Torre-Díez, I.; Diaz-Baquero, A.A.; Parra-Vidales, E.; Perea-Bartolomé, M.V.; Franco-Martín, M. Usability and User Experience of Cognitive Intervention Technologies for Elderly People With MCI or Dementia: A Systematic Review. *Front. Psychol.* **2021**, *12*, 636116. [CrossRef]
89. Bogza, L.-M.; Patry-Lebeau, C.; Farmanova, E.; O Witteman, H.; Elliott, J.; Stolee, P.; Hudon, C.; Giguere, A.M.C. User-Centered Design and Evaluation of a Web-Based Decision Aid for Older Adults Living With Mild Cognitive Impairment and Their Health Care Providers: Mixed Methods Study. *J. Med. Internet Res.* **2020**, *22*, e17406. [CrossRef]
90. Castilla, D.; Suso-Ribera, C.; Zaragoza, I.; Garcia-Palacios, A.; Botella, C. Designing ICTs for Users with Mild Cognitive Impairment: A Usability Study. *Int. J. Environ. Res. Public Health* **2020**, *17*, 5153. [CrossRef]
91. Donoghue, O.A.; Horgan, N.F.; Savva, G.M.; Cronin, H.; O'Regan, C.; Kenny, R.A. Association Between Timed Up-and-Go and Memory, Executive Function, and Processing Speed. *J. Am. Geriatr. Soc.* **2012**, *60*, 1681–1686. [CrossRef] [PubMed]
92. Andriuta, D.; Diouf, M.; Roussel, M.; Godefroy, O. Is Reaction Time Slowing an Early Sign of Alzheimer's Disease? A Meta-Analysis. *Dement. Geriatr. Cogn. Disord.* **2019**, *47*, 281–288. [CrossRef]
93. 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. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Phuket, Thailand, 24–26 July 2023; pp. 235–246. [CrossRef]
94. Cinaz, B.; Arnrich, B.; Tröster, G. Monitoring of Cognitive Functioning by Measuring Reaction Times with Wearable Devices. In Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare, Dublin, Ireland, 23–26 May 2011.
95. Burke, D.; Linder, S.; Hirsch, J.; Dey, T.; Kana, D.; Ringenbach, S.; Schindler, D.; Alberts, J. Characterizing Information Processing with a Mobile Device: Measurement of Simple and Choice Reaction Time. *Assessment* **2017**, *24*, 885–895. [CrossRef]
96. Bonnechère, B. Evaluation of Processing Speed of Different Cognitive Functions Across the Life Span Using Cognitive Mobile Games. *Games Health J.* **2022**, *11*, 132–140. [CrossRef]

97. Jutten, R.J.; Rentz, D.M.; Amariglio, R.E.; Properzi, M.J.; Maruff, P.; Johnson, K.A.; Sperling, R.A.; Papp, K.V. Fluctuations in reaction time performance as a marker of incipient amyloid-related cognitive decline in clinically unimpaired older adults. *Alzheimer's Dement.* **2022**, *18*, e066578. [CrossRef]
98. Schatz, P.; Ybarra, V.; Leitner, D. Validating the Accuracy of Reaction Time Assessment on Computer-Based Tablet Devices. *Assessment* **2015**, *22*, 405–410. [CrossRef] [PubMed]
99. Gielis, K.; Abeele, M.-E.V.; Verbert, K.; Tournoy, J.; De Vos, M.; Abeele, V.V. Detecting Mild Cognitive Impairment via Digital Biomarkers of Cognitive Performance Found in Klondike Solitaire: A Machine-Learning Study. *Digit. Biomark.* **2021**, *5*, 44–52. [CrossRef] [PubMed]
100. Thabtah, F.; Peebles, D.; Retzler, J.; Hathurusingha, C. Dementia medical screening using mobile applications: A systematic review with a new mapping model. *J. Biomed. Inform.* **2020**, *111*, 103573. [CrossRef] [PubMed]
101. Naslund, J.A.; Aschbrenner, K.A.; Araya, R.; Marsch, L.A.; Unützer, J.; Patel, V.; Bartels, S.J. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: A narrative review of the literature. *Lancet Psychiatry* **2017**, *4*, 486–500. [CrossRef] [PubMed]
102. Chinner, A.; Blane, J.; Lancaster, C.; Hinds, C.; Koychev, I. Digital technologies for the assessment of cognition: A clinical review. *Évid. Based Ment. Health* **2018**, *21*, 67–71. [CrossRef]
103. Rhodius-Meester, H.F.; Paajanen, T.; Koikkalainen, J.; Mahdiani, S.; Bruun, M.; Baroni, M.; Lemstra, A.W.; Scheltens, P.; Herukka, S.; Pikkarainen, M.; et al. cCOG: A web-based cognitive test tool for detecting neurodegenerative disorders. *Alzheimer's Dement. Diagn. Assess. Dis. Monit.* **2020**, *12*, e12083. [CrossRef] [PubMed]
104. Torous, J.; Roberts, L.W. Needed Innovation in Digital Health and Smartphone Applications for Mental Health. *JAMA Psychiatry* **2017**, *74*, 437–438. [CrossRef]
105. Carvalho, L.D.F.; Pianowski, G. Digital phenotyping and personality disorders: A necessary relationship in the digital age. *Psicol. Teor. Prática* **2019**, *21*. [CrossRef]
106. Ford, E.; Milne, R.; Curlewis, K. Ethical issues when using digital biomarkers and artificial intelligence for the early detection of dementia. *WIREs Data Min. Knowl. Discov.* **2023**, *13*, e1492. [CrossRef]
107. Jha, A.; Espay, A.J.; Lees, A.J. Digital Biomarkers in Parkinson's Disease: Missing the Forest for the Trees? *Mov. Disord. Clin. Pract.* **2023**, *10*, S68–S72. [CrossRef] [PubMed]
108. Espay, A.J.; Bonato, P.; Nahab, F.B.; Maetzler, W.; Dean, J.M.; Klucken, J.; Eskofier, B.M.; Merola, A.; Horak, F.; Lang, A.E.; et al. Technology in Parkinson's disease: Challenges and opportunities. *Mov. Disord.* **2016**, *31*, 1272–1282. [CrossRef] [PubMed]
109. Espay, A.J.; Hausdorff, J.M.; Sánchez-Ferro, Á.; Klucken, J.; Merola, A.; Bonato, P.; Paul, S.S.; Horak, F.B.; Vizcarra, J.A.; Mestre, T.A.; et al. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. *Mov. Disord.* **2019**, *34*, 657–663. [CrossRef]
110. Bloem, B.R.; Henderson, E.J.; Dorsey, E.R.; Okun, M.S.; Okubadejo, N.; Chan, P.; Andrejack, J.; Darweesh, S.K.L.; Munneke, M. Integrated and patient-centred management of Parkinson's disease: A network model for reshaping chronic neurological care. *Lancet Neurol.* **2020**, *19*, 623–634. [CrossRef]
111. Turing, A.M. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* **1937**, *s2-42*, 230–265. [CrossRef]
112. Turing, A.M. The chemical basis of morphogenesis. *Bull. Math. Biol.* **1990**, *52*, 153–197. [CrossRef]
113. Blackiston, D.; Lederer, E.; Kriegman, S.; Garnier, S.; Bongard, J.; Levin, M. A cellular platform for the development of synthetic living machines. *Sci. Robot.* **2021**, *6*, eabf1571. [CrossRef]
114. Pawlak, Z. *Rough Sets: Theoretical Aspects of Reasoning about Data*; Springer Science & Business Media: Berlin, Germany, 1991; Volume 9.
115. Pawlak, Z. Rough sets. *Int. J. Parallel Program.* **1982**, *11*, 341–356. [CrossRef]
116. Przybyszewski, A.W. *The Neurophysiological Bases of Cognitive Computation Using Rough Set Theory*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2008. [CrossRef]
117. Przybyszewski, A.W. *Theory of Mind Helps to Predict Neurodegenerative Processes in Parkinson's Disease*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2021. [CrossRef]
118. Przybyszewski, A.W.; Gaska, J.P.; Foote, W.; Pollen, D.A. Striate cortex increases contrast gain of macaque LGN neurons. *Vis. Neurosci.* **2000**, *17*, 485–494. [CrossRef] [PubMed]
119. Zadeh, L. From computing with numbers to computing with words. From manipulation of measurements to manipulation of perceptions. *IEEE Trans. Circuits Syst. I Regul. Pap.* **1999**, *46*, 105–119. [CrossRef]
120. Dutta, S.; Skowron, A. *Toward a Computing Model Dealing with Complex Phenomena: Interactive Granular Computing*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2021. [CrossRef]
121. Przybyszewski, A.W.; Śledzianowski, A. Parkinson's disease development prediction by c-granule computing compared to different AI methods. *J. Inf. Telecommun.* **2020**, *4*, 425–439. [CrossRef]
122. 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. Informaticae* **2020**, *176*, 167–181. [CrossRef]

123. 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*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2020. [CrossRef]
124. Przybyszewski, A.W.; Nowacki, J.P.; Drabik, A.; Szlufik, S.; Habela, P.; Koziorowski, D.M. *Granular Computing (GC) Demonstrates Interactions between Depression and Symptoms Development in Parkinson's Disease Patients*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2019. [CrossRef]
125. Przybyszewski, A.W.; Bojakowska, K.; Nowacki, J.P.; Drabik, A. *Rough Set Rules (RSR) Predominantly Based on Cognitive Tests Can Predict Alzheimer's Related Dementia*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Arti-ficial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2022. [CrossRef]
126. Przybyszewski, A.W. *AI Classifications Applied to Neuropsychological Trials in Normal Individuals that Predict Progression to Cognitive Decline*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2022. [CrossRef]
127. Przybyszewski, A.W. Multi-granular Computing Can Predict Prodromal Alzheimer's Disease Indications in Normal Subjects. In Proceedings of the International Conference on Computational Science, Prague, Czech Republic, 3–5 July 2023. [CrossRef]

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)

applied sciences

MDPI

Article

Classification of Parkinson's Disease Using Machine Learning with MoCA Response Dynamics

Artur Chudzik [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; artur.chudzik@pjwstk.edu.pl
[2] Department of Neurology, UMass Chan Medical School, 65 Lake Avenue, Worcester, MA 01655, USA
* Correspondence: andrzej.przybyszewski@umassmed.edu

Abstract: Neurodegenerative diseases (NDs), including Parkinson's and Alzheimer's disease, pose a significant challenge to global health, and early detection tools are crucial for effective intervention. The adaptation of online screening forms and machine learning methods can lead to better and wider diagnosis, potentially altering the progression of NDs. Therefore, this study examines the diagnostic efficiency of machine learning models using Montreal Cognitive Assessment test results (MoCA) to classify scores of people with Parkinson's disease (PD) and healthy subjects. For data analysis, we implemented both rule-based modeling using rough set theory (RST) and classic machine learning (ML) techniques such as logistic regression, support vector machines, and random forests. Importantly, the diagnostic accuracy of the best performing model (RST) increased from 80.0% to 93.4% and diagnostic specificity increased from 57.2% to 93.4% when the MoCA score was combined with temporal metrics such as IRT—instrumental reaction time and TTS—submission time. This highlights that online platforms are able to detect subtle signs of bradykinesia (a hallmark symptom of Parkinson's disease) and use this as a biomarker to provide more precise and specific diagnosis. Despite the constrained number of participants (15 Parkinson's disease patients and 16 healthy controls), the results suggest that incorporating time-based metrics into cognitive screening algorithms may significantly improve their diagnostic capabilities. Therefore, these findings recommend the inclusion of temporal dynamics in MoCA assessments, which may potentially improve the early detection of NDs.

Keywords: machine learning (ML); Montreal Cognitive Assessment (MoCA); diagnostic accuracy; Parkinson's disease (PD); rough set theory (RST); time-based measurements; web-based cognitive testing; IRT; TTS

Citation: Chudzik, A.; Przybyszewski, A.W. Classification of Parkinson's Disease Using Machine Learning with MoCA Response Dynamics. *Appl. Sci.* **2024**, *14*, 2979. https://doi.org/10.3390/app14072979

Academic Editor: Alexander N. Pisarchik

Received: 24 February 2024
Revised: 25 March 2024
Accepted: 28 March 2024
Published: 1 April 2024

1. Introduction

Neurodegenerative disorders (NDs) are conditions characterized by the progressive loss of specific neuron populations. The most common of these disorders include Alzheimer's disease, Parkinson's disease, Amyotrophic Lateral Sclerosis, and Huntington's disease. These diseases are a significant challenge for clinicians due to their diverse clinical presentations and shared molecular pathology [1].

Despite significant progress in understanding the NDs, particularly in the areas of pathology and pharmacology, the translation of preclinical innovations into effective clinical therapies has been challenging [2].

This study focuses on Parkinson's disease. PD primarily affects older individuals and is associated with a range of motor and nonmotor symptoms [3]. This is a neurodegenerative disorder characterized by the loss of dopaminergic neurons, particularly in the substantia nigra (a small region in the middle of the brain) [4]. In neurological terms, in Parkinson's disease, there is a specific pattern of neuronal loss; the dopaminergic neurons of the midbrain that project to the striatum are selectively and primarily destroyed [5,6].

The result of this neuronal damage is the manifestation of neurological symptoms. Neurological symptoms of Parkinson's disease primarily include tremors, muscle stiffness, slowed movement (bradykinesia), and impaired balance and coordination. Cognitive changes, such as difficulties with memory and problem solving, can also occur as the disease progresses.

Despite the fact that this damage is significant, it often goes unnoticed at first. This is partly because the brain has mechanisms to compensate for the damage. Keeping in mind that neuronal degeneration is a normal part of aging, in NDs, this process occurs much more rapidly and is initially masked by the brain's compensatory responses. That makes it difficult to identify specific biomarkers for early detection, diagnosis, and disease progression [7].

Therefore, the lack of a cure for Parkinson's disease can be attributed to several factors and challenges in hidden manifestations across various domains (including genetic, clinical, and pathophysiological).

1.1. The Intersection of PD and AD

The most common neurodegenerative condition is Alzheimer's disease (AD). In this disease, another small and specific area of the brain is damaged. Research suggests that the early stages of Alzheimer's disease are characterized by changes in the entorhinal–hippocampal system, particularly in the entorhinal cortex (EC) and the hippocampus [8].

The hippocampus is essential for the formation of new memories. It helps in consolidating short-term memory to long-term memory, a process that involves making these memories stable and stored efficiently for later retrieval. The entorhinal cortex, particularly its medial aspect, contains grid cells, which are critical for spatial navigation and mapping environments. This region suffers the most severe neuronal loss among the affected areas in this disease.

The outcome of the neurodegeneration in different brain regions is that patients with Alzheimer's primarily exhibit memory loss, while those with Parkinson's are more likely to suffer from movement disorders. Unfortunately, there can be an overlap in symptoms between these two neurodegenerative diseases. However, the mechanisms underlying the disease, including the role of misfolded proteins and the interaction between dopamine and acetylcholine, remain unclear [9,10].

1.2. The Hidden Progression of the Disease

This complexity and hidden progression make NDs one of the major causes of neurological disability and mortality worldwide. Despite technological advancements, we still pessimistically estimate the number of people with dementia is expected to increase from 57 million cases globally in 2019 to 152 million cases globally in 2050, notably with a three-fold increase projected in China [11,12]. Therefore, we urgently need to address the gap in understanding of these diseases, particularly in terms of the complex interplay of factors that contribute to these conditions.

To bridge this gap, there is a need for more effective screening and diagnostic methods. These should not only facilitate widespread diagnosis but also accurately identify individuals at risk of these diseases. With this objective, we selected the Montreal Cognitive Assessment (MoCA, a pen-and-paper test considered as the gold standard for cognitive evaluation), and we transformed it into an online version. Here, we present the results of applied machine learning algorithms to determine the reliability of this approach.

1.3. The Cognitive Assessment

The Montreal Cognitive Assessment (MoCA) is a brief test used by doctors to find early signs of problems with thinking and memory, which can be linked to Parkinson's disease. MoCA is a one-page 30-point test that includes tasks such as visuospatial/executive functions, naming, memory, attention, language, abstraction, delayed recall, and orientation. It is designed to quickly assess cognitive functioning and screen for mild cognitive

impairment. This test is usually carried out face-to-face in a clinical setup and takes around 15 min to complete.

Despite this, checkups of cognitive state are rare exceptions, as they generate additional workload to the budget of healthcare systems. This leads to the underdiagnosis of neurodegenerative conditions. Because of this challenge, it makes sense to use the web and machine learning models to support the process diagnosis. This removes the geographical and economical constraints, making it possible to screen wider populations, providing valuable data.

However, transitioning to an online environment poses its own set of challenges, especially for elderly users who may struggle with digital tasks that differ from their paper-based counterparts, such as drawing. Difficulty with the technology might result in lower test scores that inaccurately reflect a person's cognitive abilities rather than their familiarity with computers.

Therefore, while the prospects are promising, there are challenges to consider, including ensuring the privacy and security of sensitive medical data, addressing digital literacy and accessibility issues among older populations, and validating the efficacy of machine learning models across diverse populations. Thus, further research is needed to ensure their reliability, accuracy, and integrate them into clinical practice.

Hence, to evaluate available methods, we created an online platform in which participants can solve a self-administered version of the MoCA test. The platform is accessible via web browsers, ensuring ease of access for users from the comfort of their homes. It features a simple, intuitive interface that guides users through the screening or diagnostic process step by step. Given the sensitivity of medical data, the platform adheres to privacy and security standards, such as data encryption, secure data storage, and anonymization techniques.

Then, we analyzed the results, we used machine learning models to classify our participants according to their health status, and we compared our findings with the existing literature.

2. Methods

Taking these challenges into account, we created the online version of the MoCA test. Then, we compared multiple machine learning methods in the task of classification of healthy subjects and patients with Parkinson's disease. Each experiment measured the effectiveness and reliability of selected machine learning algorithms in detecting signs of the disease.

2.1. The Montreal Cognitive Assessment

MoCA is a cognitive health evaluation tool, primarily used for detecting mild cognitive impairment [13]. This test covers a range of cognitive domains, including short-term memory recall, visuospatial abilities, and executive functions. This ability to evaluate various cognitive aspects in a short (around 15 min) time makes it a practical and comprehensive screening tool.

MoCA has proven to be more effective than some other cognitive tests, such as the Mini-Mental State Examination (MMSE), especially in the context of Parkinson's disease [14,15]. Here, research indicates that MoCA is more sensitive in detecting cognitive impairments in PD patients.

Furthermore, it is helpful in predicting cognitive decline in the early stages of PD, with scores of 26 or lower marking a significant risk for progressive cognitive deterioration [16]. This sensitivity to early-stage cognitive issues in PD highlights MoCA's clinical importance.

Additionally, MoCA's adaptability for remote administration is particularly beneficial for patients with movement disorders [17]. This enhances its accessibility in varied clinical settings. MoCA's effectiveness extends to reflecting cognitive reserve, with emphasis on the influence of education and work activity [18].

It is important to acknowledge reported limitations, as for example, cultural and educational biases present a significant challenge. The MoCA's performance can vary across

diverse cultural and educational backgrounds, potentially leading to misinterpretation of results. Moreover, individuals with lower education levels may score lower on the MoCA, not necessarily due to cognitive impairment but due to the test's design, which might favor those with higher educational and cultural backgrounds [19]. However, Gagnon (2013) cautioned that adjusting the MoCA for education could decrease its sensitivity, leading to an increased number of false negatives [20].

Another noteworthy concern is its susceptibility to practice effects, especially between the first and second administrations. This could potentially skew results and requires consideration in clinical interpretation [21].

In summary, while the MoCA is a valuable tool for detecting cognitive impairment, it is essential to be aware of its limitations. These include cultural and educational biases, the potential for overdiagnosis, difficulty in distinguishing between different cognitive disorders, language barriers, and time constraints in clinical administration. Awareness and consideration of these factors are crucial for ensuring appropriate use of the MoCA in nuanced contexts.

2.2. Development of the Online Platform

The development of an online Parkinson's disease screening platform is guided by earlier research demonstrating the effectiveness of online tools in neurological assessment. For example, Youngmann (2019) presented a machine learning algorithm specifically designed for Parkinson's screening through a web platform, showing the feasibility and effectiveness of online tools in diagnosing PD [22]. Furthermore, Kim (2020) developed a more advanced point-of-care platform, which is geared towards the early diagnosis of Parkinson's disease [23]. These studies highlighted the growing trend and potential of using web platforms for various aspects of Parkinson's research, including data collection, patient monitoring, and early diagnostic procedures.

To conduct remote experiments and confirm this method's feasibility and performance, we created a controlled web environment. This validation is important in understanding and improving cognitive assessment tools. We developed the platform using open-source components of React 16.14.0 and Bootstrap 4.5.2, PHP8 API, and MySQL database. This choice allowed us to develop a user-friendly interface with adjustable fonts and contrasts, enhancing accessibility and ensuring the quality of data. This is an important component for participants with cognitive challenges.

The test begins with a supervisor registering a participant. The participant receives a unique URL, including a unique token, to start the test. During the test, we present screens with instructions that the participant should follow. Each of their responses was stored in the database. The test automatically scores simpler MoCA tasks. However, complex tasks like the clock drawing test were evaluated manually, according to established standards.

2.3. Inclusion Criteria for Patients and Healthy Subjects

To evaluate data aggregated by this platform, we invited two groups of people.

Our approach is targeted to establish a distinction between individuals with varying stages of PD and a generally healthy population. This method allowed us to explore and identify patterns related to PD's impact on health status.

(1) The first group included people with Parkinson's disease. 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 had UPDRS III scores above 30 (indicative of advanced symptoms of PD).
(2) For reference, we selected students from the Polish Japanese Academy of Information Technology. This approach was based on the lower likelihood of young people having PD. These individuals did not undergo neurological examination to confirm the absence of PD.

In this study, we applied disease-specific criteria for the patient group and combined convenience, demographic, and health status criteria for selecting healthy subjects. The sample size (n = 31) allows for preliminary comparisons and insights into the differences between patient and reference groups. We plan a study with a larger group to confirm these findings and explore other variables that may influence the results.

Before the study, the plan was reviewed and approved by two groups: the Institutional Review Board at UMass Chan Medical School (protocol code: IRB H0008962) and the Ethics Committee at the Polish Japanese Academy of Information Technology (protocol code: OKE-02-06-2022) to ensure compliance with the Declaration of Helsinki.

2.4. Implementation of Time-Based Measures

Currently, there is a debate in Parkinson's disease research whether digital methods can yield more detailed and precise data than traditional assessment methods, especially in detecting invisible cognitive and motor changes [24,25].

Interestingly, eye-tracking studies in Parkinson's disease confirmed that it is possible to capture these subtle changes with digital technologies [26,27]. In these studies, we can detect if patients show increased saccadic and antisaccadic delays compared to healthy controls. This suggests that certain subtle motor impairments in PD, such as delays in eye movements, can be measured using digital tools.

For this, we included two types of time measurements in the online MoCA test: Instrumental Reaction Time (IRT) and Time to Submit (TTS). Specifically, we measured the time taken to interact with the interface (IRT) and to complete and submit answers (TTS) for each MoCA test question.

IRT and TTS were tracked for each question, presented individually. IRT measures the time from the presentation of a stimulus to the initiation of a response, capturing the cognitive processing period, and TTS records the total time from stimulus presentation to the completion and submission of a response. By design, both measures reflect cognitive decision-making and motor execution abilities.

IRT and TTS are measured in the front-end layer (closest to the user). Both values are on a millisecond scale, captured without network delays using a JavaScript method *performance.now*. These measures were integrated into the digital platform, tracking the response times of each participant as they interacted with the MoCA test.

2.5. Statistical Analysis

To analyze aggregated data, we used the IBM SPSS 29 software. We examined variables such as age, gender, UPDRS group scores, and MoCA scores. We compared these variables between healthy individuals and patients with Parkinson's disease, with a p-value below 0.05 marked as significant.

2.6. Rough Set Theory

For data modeling, we used a method called rough set theory (RST). This is because earlier research shows that RST can be better than other machine learning methods for classifying diseases [28].

Rough set theory (RST), introduced by Zdzisław Pawlak in the 1980s, is a mathematical approach to data analysis that deals with vagueness and uncertainty [29,30]. RST is particularly effective in identifying patterns within imprecise or incomplete information. It operates on the principle of approximating sets by a pair of lower and upper bounds, which represent the crisp sets of all definitely and possibly belonging elements, respectively.

The fundamental operations of RST are based on the concepts of lower and upper approximations. Given a set X within the universe U, the lower approximation (L_x) is the set of elements that are certainly in X based on the available information, while the upper approximation (U_x) includes elements that could possibly belong to X.

$$(L_x) = \{x \in U : [x]_R \subseteq X\}$$

$$(U_x) = \{x \in U : [x]_R \cap X \neq \emptyset\}$$

where $[x]_R$ denotes the equivalence class of x under the equivalence relation R, exposing indiscernibility between objects in the context of available attributes.

Decision rules generated from RST analysis are in the form:

$$\text{IF(condition) THEN (decision)}$$

Decision rules derived from lower approximation sets (L_x) represent certain conclusions, and decision rules derived from upper approximation sets (U_x) represent uncertain conclusions. These rules can be directly interpreted, providing insights into the data's underlying structure and decision-making logic.

Interestingly, RST is like the way primates, like humans, visually understand complex objects they have not seen before [31]. When we see something new, our brain uses what it knows about other objects to guess what this object might be. Sometimes the new object may have features that do not match our existing knowledge, yet our brain gradually recognizes it using rule-based processes similar to RST. Likewise, RST-like mechanisms enable our brains to sort and interpret confusing or conflicting information, gradually narrowing down the gap between concepts of objects to their crisp images.

Intriguingly, this approach often yields better results than other classical machine learning approaches. An added value of RST is that the generated rules, unlike those from classical ML methods, can be easily interpreted by humans. Here, we used Rough Set Exploration System (RSES) 2.2.2 [32]. In this software, we used a 3-fold cross-validation method to test 10 objects per iteration. This means that the data were split into three parts, where a classifier is built on the basis of the training set (random 21 objects) and evaluated on 10 independent and unseen objects.

2.7. Machine Learning Approach

To challenge the Rough Set Exploration System results, we implemented three classic machine learning models in Python using the scikit-learn (sklearn) library, which facilitates a range of machine learning tools. Additionally, we utilized complementary Python libraries for data manipulation and visualization, including pandas, seaborn, and matplotlib [33–36].

Here, we employed a train_test_split approach, allocating 30% of the data for testing before modelling. This split was performed randomly to ensure that 21 subjects were used for training and 10 independent subjects for testing.

We note that a small dataset generates a higher risk of overfitting, where the model learns the noise and specific details of the training data too well and performs poorly on new, unseen data. Therefore, to mitigate the risk of overfitting, we also implemented cross-validation techniques during the model training phase. This approach allowed us to assess the model's performance more accurately and ensure its generalizability to new, unseen data.

Moreover, due to the risk of overfitting, we avoided more complex solutions like deep neural networks. Thus, we considered simpler models with regularization (e.g., L2). Specifically, we selected Logistic Regression, Support Vector Machine (SVM), and Random Forest models. The application of L2 regularization indirectly influences feature selection by shrinking the less important feature's coefficients closer to zero, which helps in identifying more significant predictors. However, L2 regularization does not perform explicit feature elimination but rather adjusts the scale of contribution of each feature.

Each of these models has its strengths and weaknesses, and their performance can vary based on the specific characteristics of the dataset. Therefore, for a better overview of the performance, we analyzed and compared them side by side.

To address the consideration of hyperparameter tuning, we used GridSearchCV with cross-validation, and we limited the range and number of hyperparameters. We used a Stratified K-Fold cross-validation method that ensured the same proportion of classes in

each fold. The dataset was balanced, meaning there was no need for synthetic oversampling techniques like SMOTE.

3. Results

The study involved a total of 31 participants, who were divided into two groups: the Parkinson's disease (PD) group and the healthy subjects (HS) group.

3.1. Statistical Analysis

The PD group contained fifteen patients (eight females and seven males), with an average age of 70.8 years (standard deviation [SD] = 5.931). The reference group contained sixteen healthy subjects: four females and twelve males, with a mean age of 23.26 years (SD = 0.964).

The PD group had a slightly higher percentage of females, representing 53% of the participants. However, males were the majority of the reference group (75%) (Table 1).

Table 1. Comparison of the characteristics of patients and healthy subjects.

Variable	PD Patients (n = 15)	Healthy Subjects (n = 16)	p-Value
Age	70.80 ± 5.931	23.26 ± 0.964	<0.001
Gender (0 = M, 1 = F)	0.53 ± 0.516	0.25 ± 0.447	0.115
UPDRS Group (0 = HS, 1 = MILD, 2 = ADV)	1.47 ± 0.516	0 ± 0	<0.001
Web MoCA Score (total)	24.13 ± 3.543	26.69 ± 1.302	0.017
Web MoCA IRT (avg, ms)	5896.27 ± 1514.26	2894.25 ± 646.974	<0.001
Web MoCA TTS (avg, ms)	13,667.00 ± 3445.42	6881.37 ± 1589.145	<0.001

3.2. Data Profiling

The dataset includes thirty-one observations, and each observation has six variables. Categorical features were managed during preprocessing and modeling stages:

(1) The 'gender' variable was encoded as a binary categorical feature, with 0 representing male (M) and 1 representing female (F). This encoding was straightforward, given the binary nature of the aggregated data, and was directly utilized in machine learning models without further transformation.

(2) The 'is_healthy' variable was also a binary categorical feature indicating the health status of the subjects (0 for patients with Parkinson's disease and 1 for healthy subjects). This binary encoding was chosen to enable clear distinction and modeling of health status as a response variable in predictive analyses.

(3) Although the 'UPDRS Group' variable was included in our dataset for a general overview, it was not used during the training of our models. The decision to exclude this variable from training was made because the UPDRS Group is related to the diagnosis and severity of Parkinson's disease, which could introduce bias into the model when predicting the health status based on broader, non-diagnostic features. Its primary role was to provide context and depth to the clinical profile of the PD patients for the readers and was not intended as a feature for prediction.

To suggest potential connections between variables, we used a matrix of Pearson correlation coefficients, showing that age is strongly negatively correlated with the healthy status and positively with IRT and TTS (Figure 1).

There was a notable age gap among participants, with a strong correlation between age and health status. Preliminary analyses indicated that using age as a predictor would lead to a model that essentially segregates the data into two clusters based solely on age.

Finally, we conducted three experiments on the dataset. The goal was to calculate how distinguishable the web test results of PD patients and healthy subjects are. Surprisingly, unsupervised PD participants performed significantly better than we expected. Their

average MoCA score was higher than noted in the literature, oscillating around 24 points, creating a challenge for classification methods.

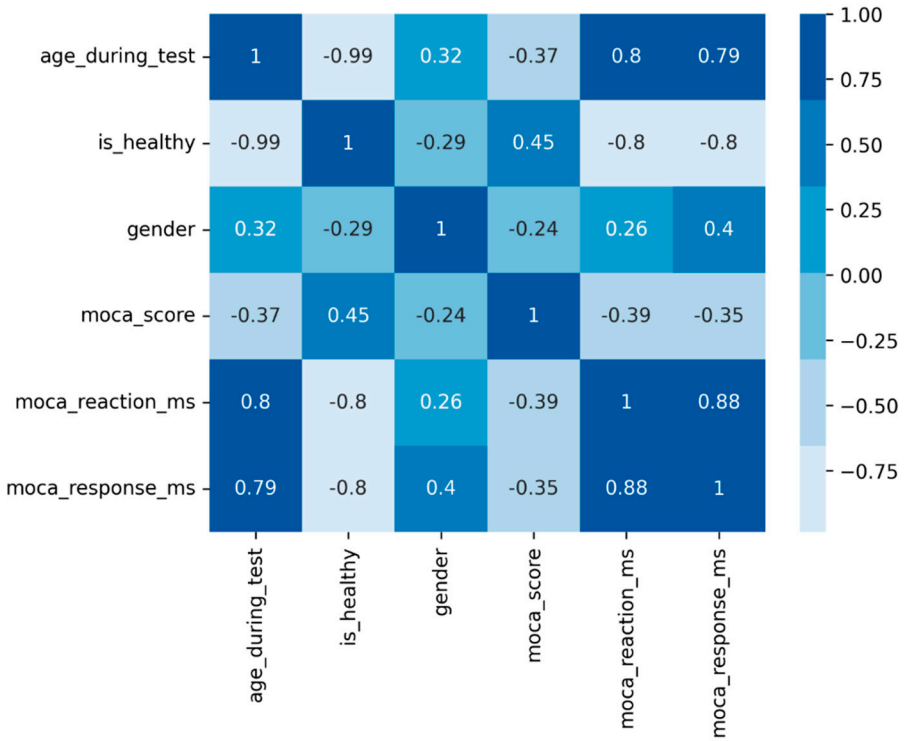

Figure 1. Correlation matrix. The similarity of all paired parameters is correlated with the color of the cell in the matrix.

3.3. Experiment I/III: Predict Health Status Based Solely on MoCA Score

The first experiment targeted the prediction of whether a participant is healthy based only on the total MoCA score. In predicting PD patients (0) using the MoCA score alone, the RSES model demonstrated effectiveness with total accuracy of 80% across all cases it evaluated, presenting its capability to classify individuals accurately (Table 2 and Figure 2). Specifically, the model's performance when identifying healthy subjects (1) was perfect, with a 100% accuracy rate, showing its strength in recognizing individuals without PD. However, the specificity, or the model's ability to correctly identify healthy individuals as healthy, was calculated as 57.13%.

Table 2. Confusion matrix for predicting participant status (is_healthy = 0/1) based on the MoCA score using the RSES model, detailing the total number of tested objects (10), total accuracy (0.8), and coverage (1.0).

		Predicted				
		0	**1**	**No. of Obj.**	**Accuracy**	**Coverage**
Actual	0	2.67	2.00	4.67	0.607	1.000
	1	0	5.33	5.33	1.000	1.000
	True positive rate	1.00	0.72			

To validate, we used hyperparameter-tuned and cross-validated ML models, including Logistic Regression, Support Vector Machine (SVM), and Random Forest. Interestingly, all three models reported a 60% accuracy, which is 20% lower than the RSES model, showing a lesser ability to generalize the prediction of health status based on MoCA scores alone. When predicting health status using only the MoCA score, all ML models achieved

a precision of 50%. This indicates that half of the positive predictions were incorrect, misclassifying healthy individuals as PD patients.

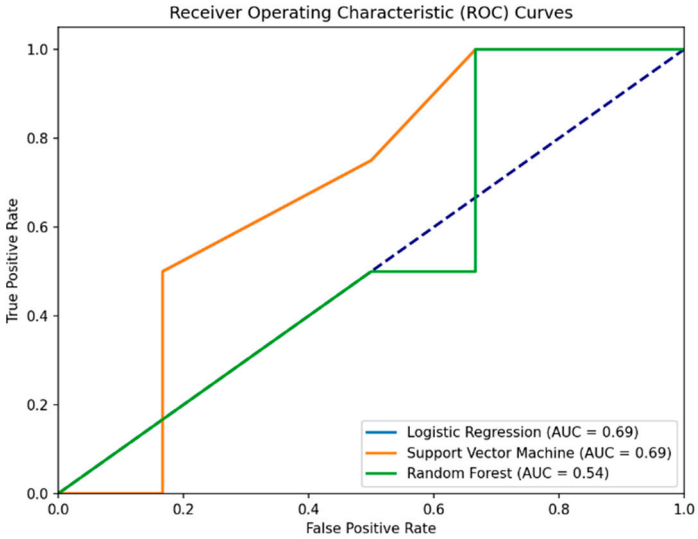

Figure 2. Comparative ROC curves for Logistic Regression, Support Vector Machine, and Random Forest models predicting health status from MoCA scores. The Logistic Regression curve (blue line) is covered by the Support Version Machine curve. The purple dashed line represents the baseline performance of a random classifier (random chance line).

The recall of 100% across models suggests that all PD cases were correctly identified, echoing the RSES's true positive rate but contrasting with its overall accuracy. The consistent confusion matrix results (2 TN, 4 FP, 0 FN, 4 TP) demonstrate the models' tendency to misclassify healthy controls as PD patients, thus presenting specificity issues. The specificity for the Logistic Regression, Support Vector Machine, and Random Forest models in Experiment I/III is all 33.3%. This indicates that each model correctly identified 33.3% of the actual negatives (healthy subjects) as being healthy. This highlights a challenge in distinguishing between PD patients and healthy controls based solely on the MoCA score in this experimental setup.

In summary, the RSES model outperformed traditional ML methods in terms of accuracy in this specific experiment, presenting a potential in handling the predictive task with a limited set of features (MoCA score alone) (Table 3). High recall across all methods highlights a common strength in identifying PD cases but also underscores a shared weakness in specificity, particularly evident in the ML methods where false positives were a significant issue.

Table 3. Comparison of model performance in Experiment I/III.

Model Name and Hyper Tuned Parameters	Accuracy	Precision	Sensitivity	Specificity
RSES	0.8	0.727	0.86	0.572
Logistic Regression {'C': 1}	0.6	0.500	1.00	0.333
SVM {'C': 0.1, 'gamma': 0.01}	0.6	0.500	1.00	0.333
Random Forest {'*n*_estimators': 4}	0.6	0.500	1.00	0.333

3.4. Experiment II/III: Predict Health Status Based on MoCA Score and Gender

The second experiment targeted the prediction of whether a participant is healthy based on both the (a) MoCA score and (b) gender. This experiment examined whether the addition of gender as a predictive variable would enhance the model's performance compared to using the MoCA score alone.

By focusing on gender, we wanted to develop models that could potentially identify health status indicators across more diverse and inclusive demographic profiles. This approach allows for the exploration of health determinants in settings where age may not be the primary factor of interest or where age information is not available.

Interestingly, the RSES analysis with the inclusion of gender showed a decrease in accuracy compared to the first experiment (80% to 70%). The detailed results indicate a mixed accuracy for different classes, with a significant true positive rate (the ability to correctly identify healthy individuals) but a lower accuracy for classifying PD patients (Table 4). The specificity for the RSES model in Experiment II/III is approximately 0.875. This indicates that the RSES model correctly identified 87.5% of the actual healthy individuals as being healthy, demonstrating a better performance in distinguishing between PD patients and healthy controls than the model based solely on MoCA scores.

Table 4. Confusion matrix for predicting participant status (is_healthy = 0/1) based on the MoCA score and gender using the RSES model, detailing the total number of tested objects (10), total accuracy (0.7), and coverage (1.0).

		Predicted				
		0	**1**	**No. of Obj.**	**Accuracy**	**Coverage**
Actual	0	2.33	2.33	4.67	0.389	1.000
	1	0.67	4.67	5.33	0.917	1.000
	True positive rate	0.67	0.66			

Surprisingly, the addition of gender as a predictive variable in Experiment II did not enhance the classical model's performance as well. The Logistic Regression model, despite a high sensitivity (75%), showed a decrease in overall accuracy (50%) and precision (42.86%) when including gender, compared to its performance in the first experiment. SVM maintained its sensitivity at 100% but did not show an improvement in overall accuracy (60%) or precision (50%) with the inclusion of gender. Random Forest showed a significant decrease in performance across all metrics (accuracy = 40%, precision = 33.33%, sensitivity = 50%) compared to the first experiment. The specificity for the Logistic Regression, Support Vector Machine, and Random Forest models in Experiment II is 0.333, or 33.3%. This indicates that each model correctly identified 33.3% of the actual healthy individuals (label 1) as being healthy, highlighting a challenge across all models in this experimental setup (Table 5 and Figure 3).

Table 5. Comparison of model performance in Experiment II/III.

Model Name and Hyper Tuned Parameters	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**
RSES	0.7	0.667	0.500	0.875
Logistic Regression {'C': 1}	0.5	0.429	0.750	0.333
SVM {'C': 0.1, 'gamma': 0.01}	0.6	0.500	1.000	0.333
Random Forest {'*n*_estimators': 5}	0.4	0.333	0.500	0.333

The results highlight the challenge in selecting features for predictive modeling, especially in small datasets. The inclusion of additional variables does not always lead to improved performance and may sometimes detract from the model's accuracy due to overfitting or the introduction of noise.

To conclude, the addition of gender as a feature alongside MoCA scores does not seem to significantly improve the models' performance, especially in the case of Random Forest, where accuracy decreases. However, the RSES model demonstrates a better balance in sensitivity when predicting between PD patients and healthy individuals compared to

the individual ML models, although with a slight drop in overall accuracy compared to Experiment I/III.

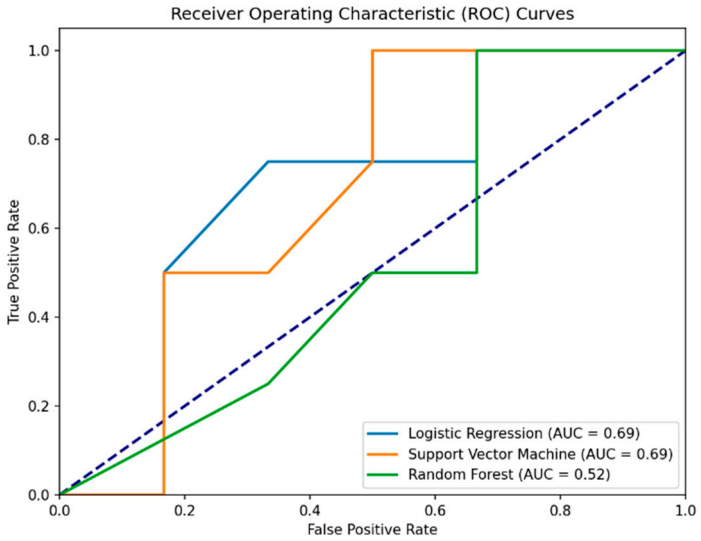

Figure 3. ROC curves for machine learning models predicting health status using MoCA scores and gender as features. The purple dashed line represents the baseline performance of a random classifier (random chance line).

3.5. Experiment III/III: Predict Health Status Based on MoCA Score and IRT + TTS

The third experiment targeted the prediction of whether a participant is healthy based the (a) MoCA score, (b) IRT, and (c) TTS. This experiment examined whether the addition of temporal measures as a predictive variable would enhance the model's performance compared to using the MoCA score alone.

Importantly, the RSES analysis with the inclusion of temporal values showed an increase in accuracy compared to the first experiment (80% to 93%). The detailed results indicate a high accuracy for different classes, with a significant true positive rate and higher accuracy for classifying PD patients (Table 6). The specificity for the RSES model in Experiment III/III is approximately 0.934. This indicates that the RSES model correctly identified 93.4% of the actual healthy individuals as being healthy, demonstrating a better performance in distinguishing between PD patients and healthy controls than based solely on MoCA scores, or MoCA scores together with gender.

Table 6. Confusion matrix for predicting participant status (is_healthy = 0/1) based on the MoCA score, IRT, and TTS using the RSES model, detailing the total number of tested objects (10), total accuracy (0.93), and coverage (1.0).

		Predicted				
		0	**1**	**No. of Obj.**	**Accuracy**	**Coverage**
Actual	0	4.67	0.33	5	0.933	1.000
	1	0.33	4.67	5	0.933	1.000
	True positive rate	0.93	0.93			

The addition of temporal measures as a predictive variable in Experiment III enhanced the classical model's performance as well. The Logistic Regression model achieved perfect sensitivity (100%) and showed an increase in overall accuracy (93%) and precision (80%). SVM maintained its sensitivity at 100% and showed an improvement in overall accuracy (90%) and precision (80%) with the inclusion of IRT and TTS. Random Forest showed a significant increase in performance across all metrics (accuracy = 90%, precision = 100%,

sensitivity = 75%) compared to the first and second experiments. The specificity for all ML models also increased. For the Logistic Regression and Support Vector Machine, this value increased to 83.3%, and Random Forest achieved a perfect (100%) specificity score.

In summary, Experiment III analyzed the impact of adding Instrumental Reaction Time and Time to Submit to MoCA scores for prediction of health status. The RSES model achieved a 93% accuracy and sensitivity rate. Machine learning models like Logistic Regression, Support Vector Machine, and Random Forest also showed marked improvements in their predictive capabilities (Table 7 and Figure 4). Specifically, Random Forest demonstrated perfect specificity, suggesting that temporal measures significantly enhance the model's ability to correctly identify healthy individuals.

Table 7. Comparison of model performance in Experiment III/III.

Model Name and Hyper Tuned Parameters	Accuracy	Precision	Sensitivity	Specificity
RSES	0.933	0.934	0.93	0.934
Logistic Regression {'C': 0.01}	0.900	0.800	1.00	0.833
SVM {'C': 0.01, 'gamma': 0.01}	0.900	0.800	1.00	0.833
Random Forest {'*n*_estimators': 2}	0.900	1.000	0.75	1.000

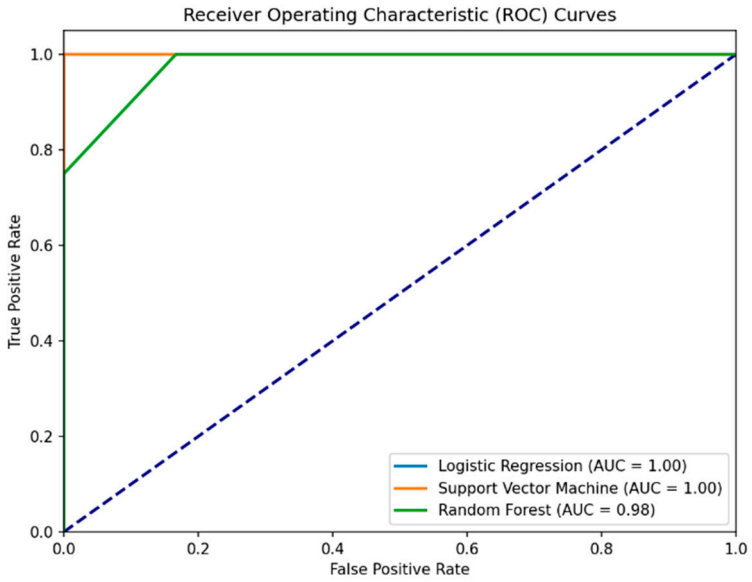

Figure 4. Enhanced ROC curves for Logistic Regression, Support Vector Machine, and Random Forest models predicting health status using MoCA scores, IRT, and TTS. The Logistic Regression curve is covered by the Support Version Machine curve. The Logistic Regression curve (blue line) is covered by the Support Version Machine curve. The purple dashed line represents the baseline performance of a random classifier (random chance line).

The results from Experiment III seem very promising. The AUC values provided for the models in the ROC curves suggest near-perfect or perfect classification ability. Understandably, to confirm the model's predictive accuracy, additional validation using an independent dataset or through a longitudinal prospective study is recommended. We plan future studies with larger samples to confirm these findings and explore other variables that may influence the results.

4. Discussion

We examined the predictive accuracy of health status based on MoCA scores. Experiment I focused solely on MoCA scores, Experiment II added gender as a variable, and

Experiment III included IRT and TTS, alongside MoCA scores. The inclusion of temporal variables, as presented in Experiment III, highlighted how behavioral and temporal data can significantly refine the models' predictive capabilities.

Initially, experiments revealed surprising specificity issues, especially when models were relying solely on MoCA scores. Interestingly, this aligns with the findings about the specificity issues of the Montreal Cognitive Assessment. The literature specifically mentions that while MoCA is good at detecting dementia (high sensitivity), it also produces a high number of false positives (low specificity). This means it often incorrectly identifies people as having dementia when they do not.

This issue was particularly raised by Davis et al. (2021), in their systematic review. They analyzed the accuracy of the MoCA test in detecting dementia across seven studies involving 9422 participants [37]. Their findings highlighted a significant concern; over 40% of non-demented individuals were falsely identified as having dementia at the MoCA cut-off of less than 26 points. These findings are in line with the trends we saw in this study.

Furthermore, Rosenblum et al. (2020) highlight the challenges in accurately identifying mild cognitive impairment in Parkinson's disease (PD-MCI) using MoCA [38]. They emphasize that factors such as age, gender, and education significantly influence the test's accuracy, complicating the early-stage diagnosis of PD-MCI. This reinforces the notion that MoCA, while useful, has limitations in distinguishing between PD-MCI and normal aging or other conditions, especially when demographic variables are significant.

Collectively, insights from the literature validate the patterns we observed in this study and suggest a potential limitation of MoCA's effectiveness, particularly in diverse settings and populations, including the described web-based approach.

Hence, this research advocates for the integration of additional assessment tools or parameters alongside MoCA to enhance diagnostic accuracy. This is significant in advancing the field of cognitive assessment, encouraging a more comprehensive approach that accounts for individual differences.

The findings are particularly crucial for PD-MCI identification, and they show the need for a nuanced approach that considers demographic and behavioral factors. Accordingly, this research suggests that relying solely on MoCA may lead to misdiagnoses or overlooked cases in PD patients.

Importantly, the insights gained from this study can extend to other cognitive assessment tools. They prompt a re-evaluation of current practices and encourage the development of more inclusive and accurate assessment strategies, particularly in digital health platforms.

Furthermore, this study underscores the limitations of MoCA in accurately diagnosing cognitive impairment across different age groups and in web-based settings. This is significant because it challenges the one-size-fits-all approach to cognitive assessment and calls for more personalized or adaptable testing methods.

It is important to acknowledge the limitations of this study. These include the small size of the research groups, potentially limiting the generalizability of these findings. The demographic diversity of the described sample may not fully represent the broader population, particularly in terms of age, gender, ethnicity, and education. The use of a web-based platform for MoCA assessment has its constraints and may not accurately replicate the nuances of an in-person testing environment.

The study's focus on a single tool, MoCA, may not encompass the multifaceted nature of cognitive impairments. The absence of longitudinal data prevents us from observing changes over time. External variables that could affect cognitive performance were not fully controlled. Finally, interpreting interactions in a web-based format poses unique challenges in understanding the complete cognitive assessment scenario.

Therefore, research should aim to address these limitations by including larger, more diverse study groups, employing a combination of cognitive assessment tools, and potentially incorporating longitudinal designs to track cognitive changes over time. Further, more detailed research is crucial for progress in the prevention of neurodegenerative diseases.

5. Conclusions

The advancement of this field of research can be achieved with the integration of time-based measurements and rough set theory. These methods significantly enhance the precision of the Montreal Cognitive Assessment in differentiating between individuals with Parkinson's disease and healthy subjects. This enhancement is shown by an improvement in diagnostic accuracy (80.0% to 93.4%) and precision (57.2% to 93.4%). The use of RST allowed for the development of rule-based models that are more adaptable and sensitive to the nuances in cognitive assessment.

An important insight from this research is the role of movement slowness associated with PD in cognitive assessment. This suggests that cognitive assessments, particularly for conditions like PD, should not be isolated from other symptom dimensions.

In conclusion, the results of this study advocate for a more integrative approach in cognitive assessments, where temporal dynamics and other symptomatic features are considered alongside traditional cognitive measures. Future research should explore similar integrative approaches in other neurodegenerative conditions and different assessment settings, including more extensive web-based platforms.

Author Contributions: Conceptualization, A.C. and A.W.P.; methodology, A.W.P.; software, A.C.; validation, A.W.P.; formal analysis, A.C.; investigation, A.C. and A.W.P.; writing—original draft preparation, A.C.; writing—review and editing, A.C.; visualization, A.C.; supervision, A.W.P.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was reviewed and approved by two groups: the Institutional Review Board at UMass Chan Medical School (protocol code: IRB H0008962) and the Ethics Committee at the Polish Japanese Academy of Information Technology (protocol code: OKE-02-06-2022) 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.

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 and the students from the Polish Japanese Academy of Information Technology.

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

References

1. Taipa, R.; Pinho, J.; Melo-Pires, M. Clinico-Pathological Correlations of the Most Common Neurodegenerative Dementias. *Front. Neurol.* **2012**, *3*, 68. [CrossRef] [PubMed]
2. Mari, Z.; Mestre, T.A. The Disease Modification Conundrum in Parkinson's Disease: Failures and Hopes. *Front. Aging Neurosci.* **2022**, *14*, 810860. [CrossRef] [PubMed]
3. Homayoun, H. Parkinson Disease. *Ann. Intern. Med.* **2018**, *169*, ITC33–ITC48. [CrossRef] [PubMed]
4. Walia, V.; Gakkhar, A.; Garg, M. Parkinson's Disease: A Progressive Disorder of the Nervous System That Affects Movement. In *Handbook of Research on Critical Examinations of Neurodegenerative Disorders*; IGI Global: Hershey, PA, USA, 2019; pp. 252–273. [CrossRef]
5. Donkelaar, H.J.T.; van de Warrenburg, B.; Willemsen, M.; Küsters, B.; Hashizume, Y.; Hori, A. Basal Ganglia. In *Clinical Neuroanatomy*; Springer International Publishing: Cham, Switzerland, 2020; pp. 591–667. [CrossRef]
6. German, D.C.; Manaye, K.; Smith, W.K.; Woodward, D.J.; Saper, C.B. Midbrain dopaminergic cell loss in parkinson's disease: Computer visualization. *Ann. Neurol.* **1989**, *26*, 507–514. [CrossRef] [PubMed]
7. Yilmaz, R.; Hopfner, F.; van Eimeren, T.; Berg, D. Biomarkers of Parkinson's disease: 20 years later. *J. Neural Transm.* **2019**, *126*, 803–813. [CrossRef] [PubMed]
8. Kobro-Flatmoen, A.; Lagartos-Donate, M.J.; Aman, Y.; Edison, P.; Witter, M.P.; Fang, E.F. Re-emphasizing early Alzheimer's disease pathology starting in select entorhinal neurons, with a special focus on mitophagy. *Ageing Res. Rev.* **2021**, *67*, 101307. [CrossRef] [PubMed]
9. Grayson, M. Parkinson's disease. *Nature* **2016**, *538*, S1. [CrossRef] [PubMed]

10. Rizzi, G.; Tan, K.R. Dopamine and Acetylcholine, a Circuit Point of View in Parkinson's Disease. *Front. Neural Circuits* **2017**, *11*, 110. [CrossRef] [PubMed]
11. Li, F.; Qin, W.; Zhu, M.; Jia, J. Model-Based Projection of Dementia Prevalence in China and Worldwide: 2020–2050. *J. Alzheimer's Dis.* **2021**, *82*, 1823–1831. [CrossRef] [PubMed]
12. Nichols, E.; Steinmetz, J.D.; Vollset, S.E.; Fukutaki, K.; Chalek, J.; Abd-Allah, F.; Abdoli, A.; Abualhasan, A.; Abu-Gharbieh, E.; Akram, T.T.; et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* **2022**, *7*, e105–e125. [CrossRef] [PubMed]
13. Hobson, J. The Montreal Cognitive Assessment (MoCA). *Occup. Med.* **2015**, *65*, 764–765. [CrossRef] [PubMed]
14. Zadikoff, C.; Fox, S.H.; Tang-Wai, D.F.; Thomsen, T.; de Bie, R.M.; Wadia, P.; Miyasaki, J.; Duff-Canning, S.; Lang, A.E.; Marras, C. A comparison of the mini mental state exam to the montreal cognitive assessment in identifying cognitive deficits in Parkinson's disease. *Mov. Disord.* **2008**, *23*, 297–299. [CrossRef] [PubMed]
15. Smith, T.; Gildeh, N.; Holmes, C. The Montreal Cognitive Assessment: Validity and Utility in a Memory Clinic Setting. *Can. J. Psychiatry* **2007**, *52*, 329–332. [CrossRef] [PubMed]
16. Kandiah, N.; Zhang, A.; Cenina, A.R.; Au, W.L.; Nadkarni, N.; Tan, L.C. Montreal Cognitive Assessment for the screening and prediction of cognitive decline in early Parkinson's disease. *Park. Relat. Disord.* **2014**, *20*, 1145–1148. [CrossRef] [PubMed]
17. Abdolahi, A.; Bull, M.T.; Darwin, K.C.; Venkataraman, V.; Grana, M.J.; Dorsey, E.R.; Biglan, K.M. A feasibility study of conducting the Montreal Cognitive Assessment remotely in individuals with movement disorders. *Health Inform. J.* **2016**, *22*, 304–311. [CrossRef] [PubMed]
18. Kang, J.M.; Cho, Y.-S.; Park, S.; Lee, B.H.; Sohn, B.K.; Choi, C.H.; Choi, J.-S.; Jeong, H.Y.; Cho, S.-J.; Lee, J.-H.; et al. Montreal cognitive assessment reflects cognitive reserve. *BMC Geriatr.* **2018**, *18*, 261. [CrossRef]
19. Borda, M.G.; Reyes-Ortiz, C.; Pérez-Zepeda, M.U.; Patino-Hernandez, D.; Gómez-Arteaga, C.; Cano-Gutiérrez, C.A. Educational level and its Association with the domains of the Montreal Cognitive Assessment Test. *Aging Ment. Health* **2019**, *23*, 1300–1306. [CrossRef] [PubMed]
20. Gagnon, G.; Hansen, K.T.; Woolmore-Goodwin, S.; Gutmanis, I.; Wells, J.; Borrie, M.; Fogarty, J. Correcting the MoCA for Education: Effect on Sensitivity. *Can. J. Neurol. Sci./J. Can. Sci. Neurol.* **2013**, *40*, 678–683. [CrossRef] [PubMed]
21. Cooley, S.A.; Heaps, J.M.; Bolzenius, J.D.; Salminen, L.E.; Baker, L.M.; Scott, S.E.; Paul, R.H. Longitudinal Change in Performance on the Montreal Cognitive Assessment in Older Adults. *Clin. Neuropsychol.* **2015**, *29*, 824–835. [CrossRef] [PubMed]
22. Youngmann, B.; Allerhand, L.; Paltiel, O.; Yom-Tov, E.; Arkadir, D. A machine learning algorithm successfully screens for Parkinson's in web users. *Ann. Clin. Transl. Neurol.* **2019**, *6*, 2503–2509. [CrossRef]
23. Kim, S.; Cho, M.; Lee, Y. Point-of-Care Platform for Early Diagnosis of Parkinson's Disease. *ACS Appl. Bio Mater.* **2020**, *3*, 8997–9001. [CrossRef] [PubMed]
24. Nair, S.S.; Muddapu, V.R.J.; Sriram, M.; Aditya, R.; Gupta, R.; Chakravarthy, S. Is There a Better Way to Assess Parkinsonian Motor Symptoms?—Experimental and Modelling Approach. In *Techniques for Assessment of Parkinsonism for Diagnosis and Rehabilitation*; Springer: Singapore, 2022. [CrossRef]
25. Ryu, J.; Vero, J.; Dobkin, R.D.; Torres, E.B. Dynamic digital biomarkers of motor and cognitive function in parkinson's disease. *J. Vis. Exp.* **2019**, e59827. [CrossRef]
26. Chudzik, A.; Śledzianowski, A.; Przybyszewski, A.W. Machine Learning and Digital Biomarkers Can Detect Early Stages of Neurodegenerative Diseases. *Sensors* **2024**, *24*, 1572. [CrossRef] [PubMed]
27. Przybyszewski, A.W.; Śledzianowski, A.; Chudzik, A.; Szlufik, S.; Koziorowski, D. Machine Learning and Eye Movements Give Insights into Neurodegenerative Disease Mechanisms. *Sensors* **2023**, *23*, 2145. [CrossRef] [PubMed]
28. 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.* **2020**, *176*, 167–181. [CrossRef]
29. Pawlak, Z. Rough sets. *Int. J. Comput. Inf. Sci.* **1982**, *11*, 341–356. [CrossRef]
30. Pawlak, Z. *Rough Sets: Theoretical Aspects of Reasoning about Data*; Springer Science & Business Media: Dordrecht, The Netherlands, 1991; Volume 9.
31. Przybyszewski, A.W.; Gaska, J.P.; Foote, W.; Pollen, D.A. Striate cortex increases contrast gain of macaque LGN neurons. *Vis. Neurosci.* **2000**, *17*, 485–494. [CrossRef] [PubMed]
32. Bazan, J.G.; Szczuka, M. The rough set exploration system. In *Transactions on Rough Sets III*; Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2005; Volume 3400. [CrossRef]
33. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
34. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 51–56. [CrossRef]
35. Waskom, M.L. seaborn: Statistical data visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [CrossRef]
36. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]

37. Davis, D.H.; Creavin, S.T.; Yip, J.L.; Noel-Storr, A.H.; Brayne, C.; Cullum, S. Montreal Cognitive Assessment for the detection of dementia. *Emergencias* **2021**, *2021*, CD010775. [CrossRef]
38. Rosenblum, S.; Meyer, S.; Gemerman, N.; Mentzer, L.; Richardson, A.; Israeli-Korn, S.; Livneh, V.; Karmon, T.F.; Nevo, T.; Yahalom, G.; et al. The Montreal Cognitive Assessment: Is It Suitable for Identifying Mild Cognitive Impairment in Parkinson's Disease? *Mov. Disord. Clin. Pract.* **2020**, *7*, 648–655. [CrossRef] [PubMed]

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

Artur Chudzik[1], Jerzy P. Nowacki[1], Andrzej W. Przybyszewski[1,2,*]

[1] Polish-Japanese Academy of Information Technology, Faculty of Computer Science,
86 Koszykowa Street, 02-008 Warsaw, Poland
[2] UMass Chan Medical School, Department of Neurology,
65 Lake Avenue, Worcester, MA 01655, USA
przy@pja.edu.pl

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.

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 [Figure 1][Figure 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.

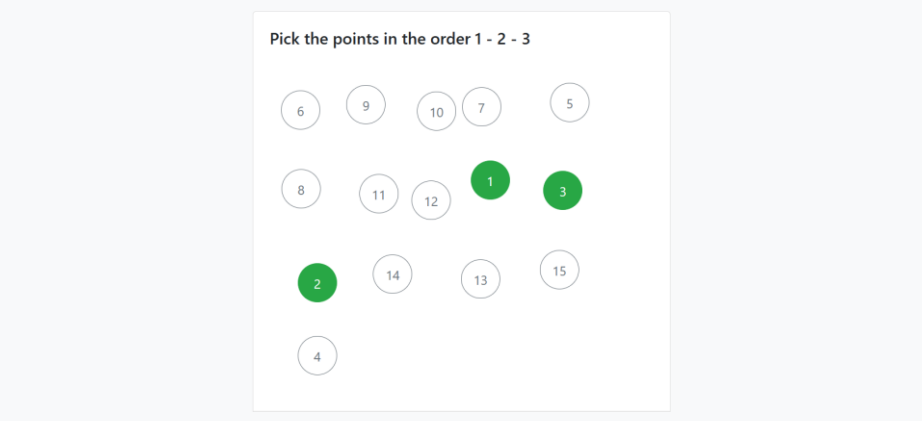

Fig. 1. The image presents an interactive cognitive task screen where the user is asked to pick points in a specific sequence. In TMT A, there are circles labeled with numbers.

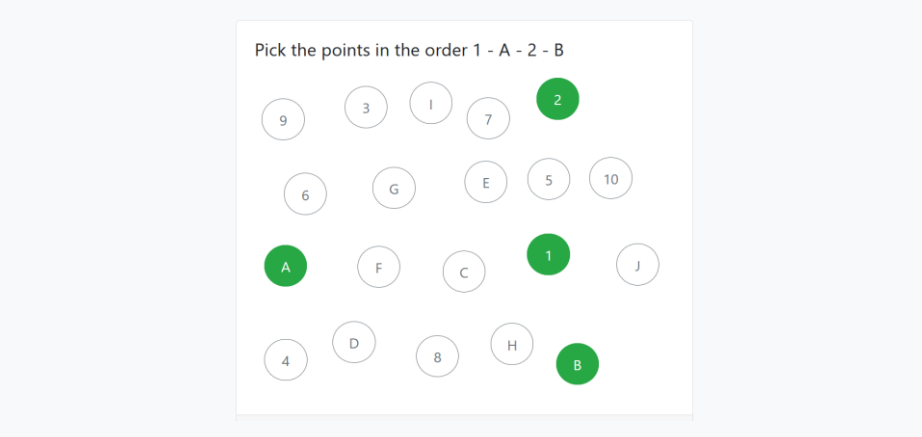

Fig. 2. The image presents an interactive cognitive task screen where the user is asked to pick points in a specific sequence. In TMT B, there are circles labeled with numbers together with letters.

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 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 Figure 3. This report presents outcomes, statistics, and ML models of this preliminary research.

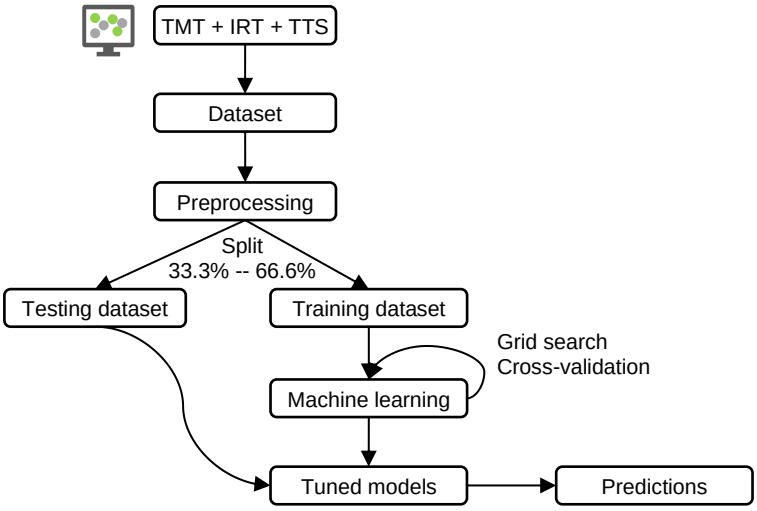

Fig. 3. Flowchart of applied methodology.

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 (Equation 1).

Equation 1. TMT_BA_MEDICAL_RESPONSE_MS = (TMT_B_RESPONSE_MS - TMT_B_REACTION_MS) - (TMT_A_RESPONSE_MS - TMT_A_REACTION_MS)

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 (Equation 2).

Equation 2. TMT_BA_ARTIFICIAL_RESPONSE_MS = (TMT_B_RESPONSE_MS - TMT_B_REACTION_MS) + (TMT_A_RESPONSE_MS - TMT_A_REACTION_MS)

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, Figure 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

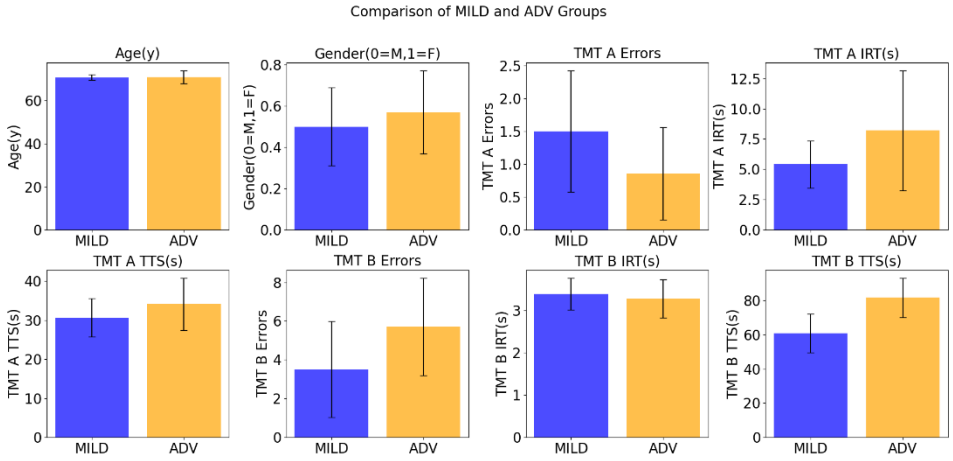

Fig. 4. Comparison of the characteristics of patients in two groups – MILD and ADVANCED (ADV).

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 seconds for MILD and 8.199 seconds for ADVANCED, while response times are 30.637 seconds for MILD and 34.152 seconds 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 signifi-

cantly 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 [Figure 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.

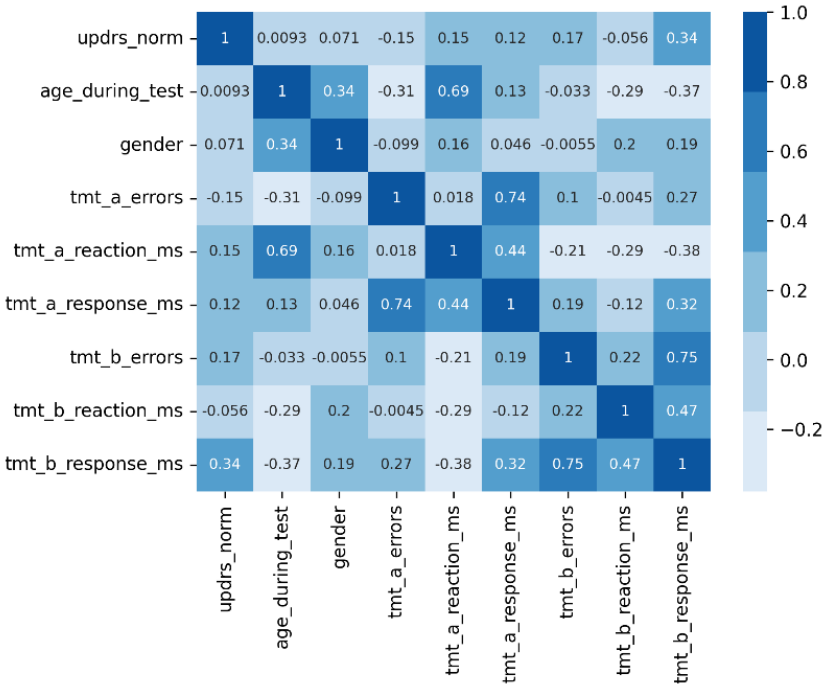

Fig. 5. A heatmap representing Pearson correlation coefficients. The darker the blue, the stronger the positive correlation; the closer to white or the presence of lighter blue indicates a weaker correlation. It's important to note that a correlation does not imply causation. These values simply indicate 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 [Figure 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).

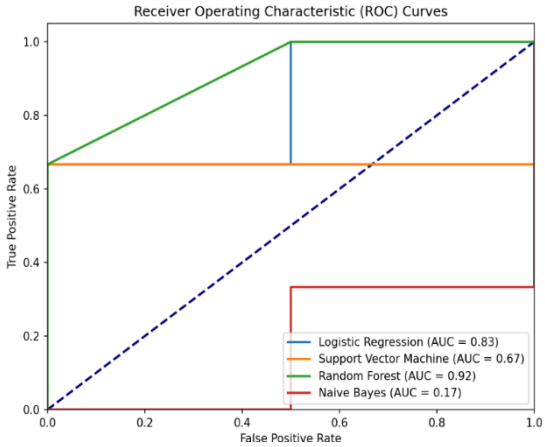

Fig. 6. The Area Under the Curve (AUC) is a measure of the overall performance of a classification model. Using Medical Response Time, Random Forest has the highest AUC of 0.92, indicating it has the best performance among the four classifiers. The Logistic Regression curve (blue line) is covered by 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).

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

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).

The Random Forest has the highest AUC of 0.92, achieving results comparable to the first experiment [Figure 7]. After the analysis of feature importance in both mod-

els 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.

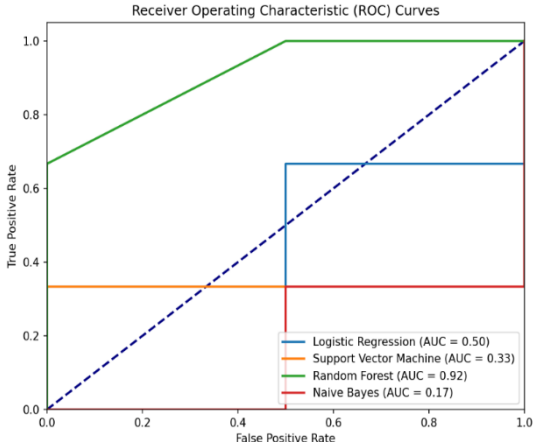

Fig. 7. The Area Under the Curve (AUC) is a measure of the overall performance of a classification model. Using Artificial Response Time, Random Forest has the highest AUC of 0.92, indicating it has the best performance among the four classifiers. The random change line represents the baseline performance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

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.

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.

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.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Talitckii, A., Kovalenko, E., Shcherbak, A., Anikina, A., Bril, E., Zimniakova, O., Semenov, M., Dylov, D. V., Somov, A.: 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. Applied Sciences. 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., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É.: 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., Lofredi, R., Feldmann, L.K., Habets, J., Darcy, N., Neumann, W.-J., Faust, K., Schneider, G.-H., Kühn, A.A.: 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. and 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.

Additional Contributions

showed, these shared hubs IRGs (CXCR4 and FLT1) have brilliant prognostic values (AUC>0.7) and whose expression level was statistically significantly elevated ($p<0.05$) in the PDD brain samples compared with the age-matched controls.
Conclusions: This work highlighted the existence of shared immune-related mechanisms between AD and PD and described the CXCR4 and FLT1 as candidate biomarkers for developing novel strategies for prognostic evaluation, and therapeutic decision making for AD and PD patients.

P 066 (GPT)

CHARCOT-MARIE-TOOTH DISEASE ASSOCIATED WITH PARKINSON DISEASE, ABOUT A CASE

H. Pacheco Mendoza [1], V. Alvarez Rivera [1]. [1] *Universidad Nacional Autonoma de Mexico, Mexico City, Mexico*

Background: Charcot-Marie-Tooth disease (CMT) is a peripheral neuropathy accompanied by weakness, atrophy, pes cavus, tremor, hearing loss and changes in sensitivity. In Parkinson Disease (PD), patients course with bradykinesia, rigidity, tremor, and postural instability. The presentation of both pathologies in a same individual is rare and cases are related to mutations of the *LRSAM1*gene.
Methods: An 81-year-old woman with a history of 2 grandchildren with MTC. Onset at age 17 with deformity in both feet and high plantar arch. He presented generalized weakness and changes in sensitivity, progressed slowly over time with difficulty walking and falls. At 75 years, she presented bradykinesia, generalized and tremor in upper limbs, in addition to cognitive impairment. This a novel case of a Mexican patient who came to Internal Medicine hospitalization, due to liver failure. After it's first evaluation, we found out that patient course with both CMT and PD. We recorded (previously patient consent sign) and performed several scales in her evaluation, such as UPDRS, MDS-UPDRS, NMSS, PDQ-39, MMSE and MoCA test, also a Levodopa test to evaluate PD, SARA and ICARS and clinical examination to evaluate CMT. We also performed neuroimage studies to a better characterization of both entities.
Results:A levodopa test was performed, with improvement in motor symptoms, cognitive tests with severe cognitive impairment, clinimetric scales for moderate ataxia. Simple tomography of the skull with cortico-subcortical atrophy and Magnetic resonance imaging (T1, T2, Flair, Swan).
Conclusions: Few cases of CMT and PD have been reported in the literature, finding 2 altered genes that could be the link between both entities. In addition, there is another important factor, limb ataxia, which has also been found to be related to PD and CMT. There is the possibility of a relationship between both diseases, however, it would be necessary to carry out genetic studies to be able to identify them due to the great heterogenicity of both neurological diseases.

Part III F: Parkinson Disease Subtypes, natural course

P 067

RELATIONSHIP BETWEEN RISK AND PROTECTIVE FACTORS AND CLINICAL FEATURES OF PARKINSON'S DISEASE

M. Costanzo [1], D. Belvisi [2], R. Pellicciari [3], A. Fabbrini [2], G. Ressa [4], S. Pietracupa [1], M. De Lucia [1], N. Modugno [1], F. Magrinelli [5], C. Dallocchio [6], T. Ercoli [7], A. Nicoletti [8], M. Zappia [8], P. Solla [7], M. Bologna [2], G. Fabbrini [2], M. Tinazzi [5], A. Conte [2], A. Berardelli [2], G. Defazio [7]. [1] *IRCCS Neuromed, Pozzilli, Italy;* [2] *Sapienza, University of Rome, Department of Human Neurosciences, Rome, Italy;* [3] *"Aldo Moro" University of Bari, Department of Basic Medical Sciences, Bari, Italy;* [4] *Albert Einstein College of Medicine, Department of Developmental and Molecular Biology, New York, United States;* [5] *University of Verona, Department of Neurosciences, Verona, Italy;* [6] *ASST Pavia-Ospedale Civile di Voghera, Voghera, Italy;* [7] *University of Cagliari, Department of Medical Sciences and Public Health, Cagliari, Italy;* [8] *University of Catania, Department G.F. Ingrassia, Catania, Italy*

Background: Non-genetic risk and protective factors play a relevant role in PD development but the relationship between these factors and PD clinical features is unknown. The aim of the present multicenter study was to investigate possible associations between risk/protective factors and clinical manifestations in a large sample of PD patients.
Methods: Six hundred ninety-four patients with PD participated in the study. Patients underwent a clinical evaluation assessing motor and non-motor symptom severity. Motor symptoms were evaluated by the International Parkinson and Movement Disorder Society-sponsored Unified Parkinson's Disease Rating Scale part III. Non-motor symptoms were evaluated by the Non-Motor Symptoms Scale. Risk and protective factors were previously identified in the present population and included coffee consumption, cigarette smoking, and physical activity as protective factors and a family history of PD, dyspepsia, and exposure to toxic agents and general anesthesia as risk factors. Linear regression models were used to identify possible associations between risk/protective factor profile and clinical variables.
Results: Coffee consumption was associated with an older age at onset and milder motor symptom severity. Non-motor symptom severity was found to be positively associated with dyspepsia and inversely associated with physical activity. We did not find any association between risk/protective factor profile and motor subtype of patients.
Conclusions: Risk and protective factors of PD development are associated with PD clinical features. This finding may represent the first step in the development of new preventive approaches able to slow disease onset and mitigate the extent of clinical manifestations.

P 068 (GPT)

AMANTADINE TREATMENT IN PARKINSON'S DISEASE PATIENTS AS A MODULATORY FACTOR OF SARS-COV-2 INFECTION

S. Szlufik [1], A. Chudzik [2], A. Przybyszewski [2], D. Koziorowski [1]. [1] *Medical University of Warsaw, Department of Neurology, Faculty of Health Sciences, Warszawa, Poland;* [2] *Polish Japanese Academy of Information Technology, Department of Informatics, Warsaw, Poland*

Background: Amantadine has been used for the prevention and treatment of viral influenza A, but more recently it is used mainly in PD patients. Previous studies showed a possible impact of amantadine on COVID-19 severity in patients using this drug due to other (neurological diseases, mainly PD). Therefore the aim of this study was to evaluate the possible impact of amantadine on the SARS-Cov-2 infection in Parkinson's disease (PD) patients.
Methods: It was a nation-wide survey performed in Polish PD population from 01.2021 till 01.2022. All members of Polish PD foundations have been asked to answer a survey – 140 PD patients filled the questionnaire consisting of 35 questions concerning the amantadine treatment, Parkinson's disease and SARS-Cov-2 infection history. The patients were divided into 2 groups: group A+ which was treated with amantadine (57 cases) and group A- (83 non-amantadine takers).
Results: We have observed more slight symptoms and progression of SARS-Cov-2 infection in PD patients taking amantadine (8 patients COVID-19+) than in PD patients not taking amantadine (12 patients COVID-19+). The symptoms of COVID-19 in A+ group were slight weakness, sweating or none of symptoms whereas group A- mainly demonstrated cough, smell loss, high temperature – one of group A- patients was hospitalized.
Conclusions: Amantadine treatment in PD patients can reduce the severity of SARS-Cov-2 infection in PD patients.

Part III G: Parkinson Disease Clinical assessment
P 069

EFFECT OF VENTRICULAR METRICS ON THE TREATMENT RESPONSE IN PARKINSON'S DISEASE

J.-J. Lee [1], J.S. Baik [2]. [1] *Ilsan Paik Hospital, Inje University College of Medicine, Department of Neurology, Goyang, Korea, Republic of;* [2] *Sanggye Paik Hospital, Inje University College of Medicine, Department of Neurology, Seoul, Korea, Republic of*

Background: A myriad of factors affects the mechanism of actions on balance and gait in Parkinson's disease (PD). We sought to explore the impact of cerebral ventricular metrics on treating patients with PD.

Cognitive session

Maciej Gaca
Nencki Institute of Experimental Biology, Polish Academy of Sciences; SWPS University of Social Sciences and Humanities
Natalia Kowalczyk, PhD, SWPS; Monika Myśliwiec, SWPS; Maciek Skorko, IP Polish Academy of Sciences; Paweł Dobrowolski, PhD, IP Polish Academy of Sciences; Artur Chudzik, Polish-Japanese Academy of Information Technology; Aneta Brzezicka, PhD, SWPS/CEDARS-SINAI MEDICAL CENTER, LOS ANGELES, USA

Structural neuroplasticity induced by training in the form of a first-person shooter video game

Over the last few years, computer games have evolved from an unrealistic and straightforward two-dimensional environment to experience becoming more similar to a real-life activity. Due to the increasing impact that video games playing has been having on global society, the cognitive effects of video games have become an exciting matter to be considered on a scientific level. While many studies show that playing action video games has a positive impact on a vast range of cognitive skills, numerous studies show negligible or no cognitive effect. There is nevertheless only a handful of studies that focus on possible structural changes as an effect of playing a video game. Those do show that playing action video games induces grey matter thickness changes in structures like the parahippocampal cortex, somatosensory cortex, superior parietal lobule or insula. The first aim of the presented study was to see whether or not approximately 30 hours of training is sufficient for any cognitive and structural changes to occur. The region of interest (ROI) approach was used to select structures of the brain that could be related to the gaming experience. The second aim of the study was to see whether or not the cortical thickness in specific structures can be predictive of the quality of the training process. A strong effect of insula did occur – while its right side did negatively change its thickness as an effect of training, the left side correlated positively with the game achievements.

Journal of Alzheimer's Disease 99 (2024) 1221–1223
DOI 10.3233/JAD-240231
IOS Press

Commentary

How to Cure Alzheimer's Disease

Andrzej W. Przybyszewski[a,b,*] and Artur Chudzik[a]
[a]*Faculty of Computer Science, Polish-Japanese Academy of Information Technology, Warsaw, Poland*
[b]*Department of Neurology, UMass Chan Medical School, Worcester, MA, USA*

Accepted 18 April 2024
Pre-press 18 May 2024

Abstract. There has been a lot of buzz surrounding new drug discoveries that claim to cure Alzheimer's disease (AD). However, it is crucial to keep in mind that the changes in the brain linked to AD start occurring 20–30 years before the first symptoms arise. By the time symptoms become apparent, many areas of the brain have already been affected. That's why experts are focusing on identifying the onset of the neurodegeneration processes to prevent or cure AD effectively. Scientists use biomarkers and machine learning methods to analyze AD progressions and estimate them "backward" in time to discover the beginning of the disease.

Keywords: Alzheimer's disease, Clinical Dementia Rating SUM of Boxes, cognitive tests, multi-granular computing, rough set theory

The potential impacts of these methods on early detection and treatment strategies are important for AD as the prevalence of this disease is increasing rapidly due to our aging population. By 2050, an estimated 152 million cases could exist worldwide [1]. The earliest biomarker for AD, total tau (t-tau) in the cerebrospinal fluid (CSF), appears 34 years before the onset of the first symptoms, which can be associated with clinical impairment several decades later [2].

Cognitive tests and CSF amyloid-β (Aβ) and phosphorylated tau (p-tau) also show changepoints 10–15 years before symptom onset. Cognitive changes were estimated by Cognitive Test Substitution, Logical Memory Delayed, Pair Associates Delayed, and Boston Naming Test [2]. Magnetic resonance imaging (MRI) visible changes in medial temporal lobes can be observed about 6 years before AD onset [2]. However, MRI changes were not the subject of [3].

*Correspondence to: Andrzej W. Przybyszewski, Faculty of Computer Science, Polish-Japanese Academy of Information Technology, 86 Koszykowa St., 02-008 Warsaw, Poland. Tel.: +48 608264992; E-mail: przy@pjwstk.edu.pl.

Regrettably, no cure for AD currently exists. By the time clinical symptoms and neurological diagnosis appear, many parts of the brain have already been affected inter alia by intercellular t-tau or Aβ, making recovery impossible. Since the neurodegenerative processes begin two to three decades before observed symptoms, the key to fighting AD is to estimate the beginning period of the AD-related brain changes.

In their review, Cabrera-León et al. (2024) focus on state-of-the-art studies on AD's automatic diagnosis, prognosis, and early stages, mainly mild cognitive impairment [3]. Classification methods were based on artificial neural networks, including deep learning, and data from modalities different from brain signals. By looking into different machine learning methods, databases, and measured parameters these examined studies shed some light on the complex mechanisms related to AD in the brain.

There has been a recent increase in cross-sectional studies that use neural computation methods, particularly in the last two years. These studies have utilized both artificial neural networks and deep neural networks-based studies, except in areas not involving neuroimaging where deep learning dominates.

ISSN 1387-2877/$35.00

Characterized by their deep layer structures and extensive computational requirements, deep learning methods present challenges in optimization and cost. Transfer learning has been highlighted as a solution to mitigate these issues; however, its application was limited in the studies reviewed. This could be because convolutional neural networks are not used as much when neuroimaging data are not involved.

The review observed a broad range of dataset sizes and highlighted the challenge of class imbalance in the datasets that can bias model predictions, with various tactics applied to overcome this issue. Importantly, binary classification tasks were more common than multiclass ones, with the most common being the differentiation between cognitively normal individuals and those with AD due to clearer distinctions in early research stages. Prediction of the severity stage of AD was rare. Authors noticed problems, particularly in the context of unbalanced datasets. Therefore, the review advocates the use of performance metrics, such as the area under the curve, sensitivity, specificity, precision, and balanced accuracy, to provide more meaningful comparisons of study outcomes, especially in the context of unbalanced data.

There is indeed a need for more quantitative methods in the diagnosis and prognosis of mild cognitive impairment and AD. While qualitative measures such as cognitively normal, mild cognitive impairment, and dementia are commonly used, they can be subjective and not ideal for machine learning methods. By using more quantitative measures, we can better identify the onset of neurodegenerative processes and potentially develop more effective treatments and prevention strategies.

Furthermore, addressing the issue of homogeneity in binary classification studies within the field, we utilized the intelligent granular computing technique to evaluate AD progressions and potentially identify the beginning of neurodegeneration processes in a backward direction. Intelligent Granular Computing and Rough Set Theory are advanced computational approaches that offer promising avenues for understanding complex data patterns. Our granules are a set of attributes that determine different AD stages. We proposed to implement Intelligent Granular Computing as a Rough Set Theory that introduced rules determining the possibility of dementia in normal subjects observed for almost 30 years in the BIOCARD project [4].

In the BIOCARD study, neurologists evaluated subjects every year with the global score Clinical Dementia Rating (CDR) parameters to determine if a particular individual is normal, has mild cognitive impairment, or has dementia [4]. We used classification based on CDRSUM (sum of boxes) as a more precise and quantitative general index than the global score to provide more information on patients with mild dementia [5]. We were using the cognitive attributes described above with some combinations of additional attributes with the help of multi-granular computing to demonstrate that different sets of attributes were important for the classification of each patient [6]. Additionally, by using rules from patients classified by neurologists as AD and applying them to neurologically normal subjects (CDRSUM = 0), we found that some of them had CDRSUM > 0. Based on the study's rules, we predicted (in 21 normal subjects) that one subject might get mild dementia (CDRSUM > 4.5), one very mild dementia (CDRSUM > 2.25), and five others might get questionable impairment (CDRSUM > 0.75) [5].

Earlier detection and more personalized interventions are possible with artificial intelligence methods. They can uncover patterns invisible to neuropsychologists in the cognitive attributes of normal subjects. That might indicate their pre-dementia stage 5–10 years before symptoms onset [2].

AUTHOR CONTRIBUTIONS

Andrzej W. Przybyszewski (Conceptualization; Investigation; Writing – review & editing); Artur Chudzik (Writing – review & editing).

ACKNOWLEDGMENTS

The authors thank Anja Soldan from BIOCARD for informing us that patients, as we predicted, who might get mild or very mild dementia are still after 3 years, normal (CDR = 0). However, notice that in [2] states "all cognitive tests had changepoints 10–15 years before symptom onset".

FUNDING

The authors have no funding to report.

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

REFERENCES

[1] GBD 2019 Dementia Forecasting Collaborators (2022) Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* **7**, e105-e125.

[2] Younes L, Albert M, Moghekar A, Soldan A, Pettigrew C, Miller MI (2019) Identifying changepoints in biomarkers during the preclinical phase of Alzheimer's disease. *Front Aging Neurosci* **11**, 74.

[3] Cabrera-León Y, Báez PG, Fernández-López P, Suárez-Araujo CP (2024) Neural computation-based methods for the early diagnosis and prognosis of Alzheimer's disease not using neuroimaging biomarkers: A systematic review. *J Alzheimers Dis* **98**, 793-823.

[4] Albert M, Zhu Y, Moghekar A, Mori S, Miller MI, Soldan A, Pettigrew C, Selnes O, Li S, Wang MC (2018) Predicting progression from normal cognition to mild cognitive impairment for individuals at 5 years. *Brain* **141**, 877-887.

[5] O'Bryant SE, Waring SC, Cullum CM, Hall J, Lacritz L, Massman PJ, Lupo PJ, Reisch JS, Doody R; Texas Alzheimer's Research Consortium (2008) Staging dementia using Clinical Dementia Rating Scale Sum of Boxes scores: A Texas Alzheimer's research consortium study. *Arch Neurol* **65**, 1091-1095.

[6] Digital and Analog Biomarkers in AD. https://nd.pja.edu.pl/ad-clinic.html