

Data-Driven Strategies in Neurodegenerative Diseases: Advancing Drug Development and Disease Management

Teil-Kumulative Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)
der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

TAMARA RASCHKA

aus Bad Honnef

Bonn, 2025

Angefertigt mit Genehmigung
der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

- Gutachter/Betreuer: Honorar Prof. Dr. rer. nat. Holger Fröhlich
- Gutachter: Univ.-Prof. Dr. rer. nat. Jan Hasenauer

Tag der Promotion: 06.10.2025
Erscheinungsjahr: 2025

Abstract

Neurodegenerative Disorders, including Alzheimer's, Parkinson's, and Huntington's Disease, pose major challenges to modern medicine due to their progressive nature and impact on patients and the healthcare system. These disorders lead to declines in cognitive, motor, and functional capabilities due to gradual neuron degeneration. Despite extensive research, effective disease-modifying therapies remain elusive, hindered by the complexity of identifying suitable therapeutic targets and the heterogeneity of each condition.

This thesis addresses critical challenges within the pharmaceutical value chain using data-driven methodologies, including machine learning and artificial intelligence, to develop innovative strategies for personalized medical treatments and improved disease management. A key challenge lies in identifying appropriate target structures for active substances. A systems biology approach leverages advanced data analytics for an in-depth analysis of complex interactions within biological pathways, enhancing the understanding of disease mechanisms and guiding the search for effective treatments. Furthermore, determining the appropriate timing and patient selection for treatment is essential. The heterogeneity of symptom progression in Huntington's Disease is examined, identifying two distinct progression subtypes with significant cognitive performance differences. These findings underscore the need for personalized treatment strategies based on individual progression patterns and the importance of recognizing symptom diversity that influences clinical outcomes. Additionally, objective measurements are investigated by evaluating gait sensor data for monitoring symptoms and their progression in Parkinson's Disease. The feasibility of using sensor-based digital gait data as endpoints in clinical trials is assessed by determining necessary sample sizes and measurement effectiveness. This approach aims to establish a reliable framework for integrating health technologies into clinical practice, enhancing patient monitoring and outcomes.

In conclusion, the results of this thesis underscore the potential of data-driven methods in developing disease-modifying treatments for neurodegenerative diseases. Insights gained contribute to understanding potential drug targets, disease progression heterogeneity, and the utility of digital sensors for monitoring, paving the way for more effective drug development and personalized disease management.

Acknowledgments

I would like to express my deepest gratitude to Prof. Dr. Holger Fröhlich, my PhD supervisor, for his invaluable guidance, support, and insightful feedback throughout my journey. I am profoundly grateful for the opportunity to work in his group. His constant scientific guidance, critical inputs, and encouragement have been instrumental in helping me achieve the goals of my thesis. Thank you for mentoring me and for your continuous support.

I would also like to extend my heartfelt appreciation to my second supervisor, Prof. Dr. Jan Hasenauer, for agreeing to be a part of my PhD journey as my second supervisor and reviewer of this thesis. I appreciate the efforts of all my thesis committee members for their invaluable contributions.

To my colleagues and collaborators, thank you for your camaraderie and collaboration. It has been an honor to research alongside you, and I have truly enjoyed and enriched my experience through our stimulating projects. Your support and invaluable scientific advice throughout my studies have made a lasting impact.

I am especially grateful to my friends, whose unwavering belief and unlimited support have been my foundation throughout this journey. I feel blessed and fortunate to be surrounded by so many wonderful people who have always believed in me. Without your constant love and support, I would not have been able to undertake this journey. You have encouraged me to chase my dreams, picked me up every time I fell, and reminded me in overwhelming moments that I could persevere. Your inspiration and emotional support have helped me through every step of this way.

To my partner, Julian Rohrmann, thank you for your endless love, patience, and support. It's hard to express how grateful I am to have you in my life. You have shared the highs and lows of my journey with me, always encouraging me and supporting me emotionally. Your unwavering belief in me has been a source of strength and I am incredibly grateful for everything you have done for me.

Thank you all for being a vital part of this journey.

Publications

Thesis Publications

[†] Equal contribution

[‡] Equal contribution

1. **Raschka, T.**, Sood, M., Schultz, B., Altay, A., Ebeling, C., Fröhlich, H. (2023). AI reveals insights into link between CD33 and cognitive impairment in Alzheimer's Disease. *PLOS Computational Biology*, 19(2), e1009894. doi:10.1371/journal.pcbi.1009894
2. **Raschka, T.**[†], Li, Z.[†], Gassner, H., Kohl, Z., Jukic, J., Marxreiter, F.[‡], Fröhlich, H.[‡] (2024). Unraveling progression subtypes in people with Huntington's Disease. *EPMA Journal*, 15, 275–287. doi:10.1007/s13167-024-00368-2
3. **Raschka, T.**, To, J., Hähnel, T., Sapienza, S., Ibrahim, A., Glaab, E., Gaßner, H., Steidl, R., Winkler, J., Corvol, J. C., Klucken, J., Fröhlich, H. (2024). Objective monitoring of motor symptoms and their progression in Parkinson's Disease using a digital gait device. *Preprint Research Square*. doi:10.21203/rs.3.rs-4521747/v1 This publication is currently under review in Scientific Reports.

Other Publications

4. Domingo-Fernández, D., Baksi, S., Schultz, B., Gadiya, Y., Karki, R., **Raschka, T.**, Ebeling, C., Hofmann-Apitius, M., Kodamullil, A.T. (2021). COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology. *Bioinformatics*, 37, 1332-1334. doi:10.1093/bioinformatics/btaa834

5. Figueiredo, R.Q.[†], **Raschka, T.**[†], Kodamullil, A.T., Hofmann-Apitius, M., Mubeen, S.[‡], Domingo-Fernández, D.[‡] (2021). Towards a global investigation of transcriptomic signatures through co-expression networks and pathway knowledge for the identification of disease mechanisms. *Nucleic Acids Research*, 49, 7939–7953. doi:10.1093/nar/gkab556
6. Schultz, B., Zaliani, A., Ebeling, C., Reinshagen, J., Bojkova, D., Lage-Rupprecht, V., Karki, R., Lukassen, S., Gadiya, Y., Ravindra, N.G., Das, S., Baksi, S., Domingo-Fernández, D., Lentzen, M., Strivens, M., **Raschka, T.**, Cinatl, J., DeLong, L.N., Gribbon, P., Geisslinger, G., Ciesek, S., Van Dijk, D., Gardner, S., Kodamullil, A.T., Fröhlich, H., Peitsch, M., Jacobs, M., Hoeng, J., Eils, R., Claussen, C., Hofmann-Apitius, M. (2021). A method for the rational selection of drug repurposing candidates from multimodal knowledge harmonization. *Scientific Reports*, 11, 11049. doi:10.1038/s41598-021-90296-2
7. Wißfeld, J., Nozaki, I., Mathews, M., **Raschka, T.**, Ebeling, C., Horning, V., Brüstle, O., Neumann, H. (2021). Deletion of Alzheimer’s disease-associated CD33 results in an inflammatory human microglia phenotype. *Glia*, 69, 1393-1412. doi:10.1002/glia.23968
8. Figueiredo, R.Q., Del Ser, S.D., **Raschka, T.**, Hofmann-Apitius, M., Kodamullil, A.T., Mubeen, S.[‡], Domingo-Fernández, D.[‡] (2022). Elucidating gene expression patterns across multiple biological contexts through a large-scale investigation of transcriptomic datasets. *BMC Bioinformatics*, 23, 231. doi:10.1186/s12859-022-04765-0
9. Birkenbihl, C., Ahmad, A., Massat, N.J., **Raschka, T.**, Avbersek, A., Downey, P., Armstrong, M., Fröhlich, H. (2023). Artificial intelligence-based clustering and characterization of Parkinson’s disease trajectories. *Scientific Reports*, 13, 2897. doi:10.1038/s41598-023-30038-8
10. Schultz, B.[†], DeLong, L.N.[†], Masny, A., Lentzen, M., **Raschka, T.**, Van Dijk, D., Zaliani, A., COPERIMOpus, Fröhlich, H. (2023). A machine learning method for the identification and characterization of novel COVID-19 drug targets. *Scientific Reports*, 13, 7159. doi:10.1038/s41598-023-34287-5
11. Bontridder, N., Hähnel, T., Marín Valero, M., Carrasco Marín, L., Corvol, J., A. El-Yacoubi, M., Eskofier, B.M., Filali Razzouki, A., Funck Hansen, A., Glaab, E., Ibrahim, A.A. , Jeancolas, L., Klucken, J., Lage-Rupprecht, V., Le Bars, S., Lehericy, S., Mangone, G., Paccoud, I., Petrovska-Delacretaz, D., **Raschka, T.**, de Terwangne, C., Van

Gyseghem, J., Winkler, J., Fröhlich, H. (2024). Leveraging the Potential of Digital Technology for Personalized Medicine. *Journal of AI Law and Regulation*, 2, 249-253. doi:10.21552/aire/2024/2/12

12. Hähnel, T., **Raschka, T.**, Sapienza, S., Klucken, J., Glaab, E., Corvol, J.-C., Falkenburger, B., Fröhlich, H. (2024). Progression Subtypes in Parkinson's disease identified by a data-driven multi cohort analysis. *npj Parkinson's Disease*, 10(1), 95. doi:10.1038/s41531-024-00712-3

Contents

1	Introduction	1
1.1	Neurodegenerative Diseases	1
1.1.1	Alzheimer’s Disease	1
1.1.2	Parkinson’s Disease	3
1.1.3	Huntington’s Disease	5
1.1.4	The challenges in diagnosis and treatment of NDDs . .	7
1.2	The drug development process	10
1.2.1	Leveraging Data-Driven Approaches for Target Identification	12
1.2.2	Mechanistic modeling in NDDs	14
1.3	Disease management in NDDs	15
1.3.1	Progression subtype identification	17
1.3.2	Objective symptom monitoring	19
1.4	Research contributions of this thesis	21
2	Comprehensive Multi-Scale AD Model for Knock-Down Simulations	25
3	Identification and Predictive Modeling of HD Progression Subtypes	28
4	Gait Sensors for Objective Monitoring PD Symptoms and	

Clinical Trial Endpoints	31
4.1 Introduction	31
4.2 Methodology	33
4.2.1 Overview of datasets	33
4.2.2 LTJMM model	38
4.2.3 Evaluating the relationship between digital gait pat- terns and clinical outcomes	41
4.2.4 ML models for forecasting disease progression	42
4.2.5 Sample size calculation for digital gait as a surrogate endpoint in clinical trials	43
4.3 Results	43
4.3.1 LTJMM model	43
4.3.2 Evaluating the relationship between digital gait pat- terns and clinical outcomes	46
4.3.3 Forecasting disease progression with ML models	49
4.3.4 Sample size calculation for digital gait as surrogate endpoint in clinical trials	50
4.4 Discussion and Conclusion	52
4.4.1 Digital Gait Evaluations for monitoring symptoms and progression	52
4.4.2 Disease progression forecasting with ML methods	53
4.4.3 Potential Value of digital gait as a surrogate endpoint in clinical trials	54
4.4.4 Summary	55

5 Conclusion	56
---------------------	-----------

5.1 Perspective future work	59
Acronyms	62
References	64
A Appendix	90

1 Introduction

1.1 Neurodegenerative Diseases

Neurodegenerative Diseases (NDDs) are characterized by the progressive loss of neurons and subsequent brain deterioration [1]. Among the most prevalent NDDs are Alzheimer’s disease (AD) and Parkinson’s disease (PD) [2, 3]. A less common NDD is Huntington’s disease (HD), which is distinct from AD and PD as it is a monogenic condition caused by a mutation in a single gene [4]. All NDDs exhibit shared pathologies, including abnormal protein conformations and aggregation, such as the amyloid plaques found in AD. In each of these conditions, these protein abnormalities, along with neuronal loss and brain atrophy, typically manifest prior to the appearance of clinical symptoms [1].

1.1.1 Alzheimer’s Disease

AD is a neurodegenerative disorder characterized by cognitive and functional decline, ultimately leading to death [2, 5]. It stands as the most common cause of dementia, affecting over 55 million individuals worldwide, predominantly those aged 65 and older [2]. The incidence of AD rises with age, impacting one-third of people over 85 with women accounting for two-thirds of AD diagnoses [6, 7]. In recent years, a significant proportion of new dementia cases has emerged in low- and middle-income countries, where two-thirds of all AD cases are concentrated [5, 6]. Predictions indicate that the number of individuals living with AD may double by 2050 [5, 7]. Despite considerable research efforts, AD remains one of the top ten leading causes of death in the United States, with reported fatalities increasing by over 140% in the past two decades [2]. Furthermore, the global economic burden of AD was estimated at \$1.3 trillion in 2019, imposing substantial strain on healthcare systems and families alike [8].

AD is a multifactorial disease with its precise underlying causes remain elusive [9]. Two important pathological alterations associated with AD in-

clude the extracellular aggregation of amyloid-beta protein and the formation of neurofibrillary tangles composed of misfolded tau protein [9, 10]. These processes contribute to neurodegeneration by harming neurons, resulting in diminished synaptic strength and synaptic loss [9–11]. Furthermore, various factors, such as cardiovascular health [12], neuroinflammation [13], and cholesterol metabolism [14], are under investigation for their potential roles in shaping the disease phenotype.

AD is a progressive disorder that initially manifests as mild difficulties in memory, language, and cognition, gradually evolving to encompass communication challenges and behavioral changes. Ultimately, severe symptoms emerge, rendering everyday tasks such as walking, speaking, and swallowing increasingly impossible for patients [2, 5]. Clinically, AD is classified into three stages: 1) pre-clinical AD, 2) mild cognitive impairment (MCI), and 3) dementia due to AD [15–17]. During the pre-clinical phase, the aggregation of amyloid-beta and the formation of neurofibrillary tangles occur, yet no recognizable symptoms are present. This phase commences as much as 20 years or more before patients and their families notice the first signs of the disease [15, 18]. Once symptoms manifest, patients transition into the MCI stage, where mild issues, such as memory problems, become apparent but do not significantly interfere with daily living. Individuals with MCI are at an increased risk of progressing to dementia, the third phase, typically within a few years. The duration of this phase varies considerably, influenced by numerous factors, including age, genetics, and lifestyle [16, 18]. As symptoms intensify, patients eventually advance to the dementia phase, which can be further delineated into three stages: mild, moderate, and severe, based on the extent of impairment in daily activities [17]. Historically, AD was diagnosed solely on the basis of symptoms, focusing on stage of dementia and neurobehavioral signs sufficiently pronounced to affect daily functioning [19]. Today, diagnosis integrates clinical assessment with biological markers, including the detection of abnormal levels of amyloid beta and tau in the cerebrospinal fluid (CSF) and blood [20–22]. Recently, criteria for diagnosing AD have been established based exclusively on biological assessments [22].

In addition to age and family history, genetics play an important role in the risk of developing AD [2, 5, 23]. Research has uncovered several genes that influence the biological mechanisms associated with AD, encompassing both risk factors and protective elements [23, 24]. Among these, the APOE gene exerts the most profound effect on the likelihood of late-onset AD [2, 5]. The gene exists in three variants, with the e4 allele markedly elevating the risk of developing AD compared to the e2 and e3 alleles. The risk is further

amplified in individuals carrying two copies of the e4 allele [2, 23, 25, 26]. Other noteworthy genes that impact risk and protection include APP [27], TREM2, [28–30] and CD33 [30–35].

Currently, there is no cure for AD [2, 5, 36]. Most available treatments aim to alleviate symptoms rather than targeting the underlying biological mechanisms, which means they do not slow or halt disease progression [2, 5]. Recently, the U.S. Food and Drug Administration (FDA) approved three drugs: aducanumab, lecanemab, and donanemap, designed as amyloid beta-targeting antibodies to eliminate amyloid- β plaques in the brain [37–40]. Aducanumab received conditional approval, pending further evidence of its cognitive efficacy, while lecanemab and donanemap achieved full approval after demonstrating effectiveness in improving cognitive function [39, 40]. However, the manufacturer of aducanumab has announced the discontinuation of its development. While lecanemab has shown moderate success in slowing cognitive and functional decline in individuals with MCI or mild AD, some researchers remain skeptical about its clinical applicability [2, 41].

1.1.2 Parkinson’s Disease

In addition to AD, PD is the second most prevalent neurodegenerative disorder, affecting over 6 million people as of 2016 [3]. It is the fastest-growing neurodegenerative disease, with cases projected to continue rising [42]. The total economic burden of PD in the US was estimated at \$51.9 billion in 2017, a figure expected to exceed \$79 billion by 2037 [43]. Both the incidence and prevalence of PD increase with age: approximately 25% of patients are diagnosed before the age of 65, and 5-10% receive a diagnosis before the age of 50 [3]. The incidence is notably lower in women, who generally experience a later onset of symptoms [3, 42]. PD is a progressive disease that ultimately leads to death, however many individuals with PD can live for extended periods with the condition, although their life expectancy is reduced compared to that of healthy individuals [3].

PD is a motor syndrome linked to the deposition of alpha-synuclein deposition in the substantia nigra [44]. Although the precise underlying causes remain unclear, similar to AD, PD is characterized by the aggregation of alpha-synuclein, which is associated with mitochondrial dysfunction, synaptic transport issues, immune activation, and alterations in lysosomal and endosomal function, alongside neuroinflammation [44]. This pathological process

leads to the loss of dopaminergic neurons, resulting in decreased dopamine release and slower movements, a condition known as bradykinesia [3]. Recent research indicates that PD pathology may progress in distinct ways: in the body-first type, it originated in the gut and peripheral nervous system before advancing to the brain, while in the brain-first type, it begins in the brain and subsequently spreads to the rest of the body [45].

PD is a progressive condition in which all symptoms initially manifest subtly and worsen over time. The rate of progression and severity of symptoms vary widely among individuals, complicating prognosis [44, 46]. PD is a multifaceted disorder that encompasses both motor and non-motor symptoms [3]. During the prodromal phase, which can last 20 years or more prior to the onset of motor symptoms, typical signs include constipation, hyposmia, depression, and REM sleep behavior disorder (RBD) [47]. Motor symptoms include bradykinesia, muscle stiffness (rigidity), resting tremors, and gait disturbances [48]. Impaired gait patterns are characterized by reduced step length and speed, increased axial rigidity, and impaired rhythmicity [49]. Other prevalent non-motor symptoms include anxiety, cognitive decline, sleep disorders, and pain [50]. The diagnosis of PD primarily relies on the clinical presentation of motor symptoms, confirmed by the presence of bradykinesia in conjunction with either resting tremor, rigidity, or both [51]. Caution is warranted when these indicators are absent, as this may suggest alternative diagnoses such as atypical Parkinsonian syndromes, which typically do not respond to treatment - a key distinguishing factor from PD [51]. A biological diagnostic method is currently under development and is expected to be based on alpha-synuclein seed amplification from blood and CSF [52–54].

Approximately 3-5% of all PD cases are monogenic, meaning that the disease symptoms can be attributed to a mutation in a single gene [3]. Such mutations have been identified in SCNA, LRRK2, PRKN, and PINK1, with PRKN and PINK1 mutations leading to early onset of the disease, while SCNA and LRRK2 mutations are associated with a typical onset [55]. Additionally, mutations in the GBA gene are recognized as a common risk factor for developing PD. In total, 90 different and independent risk variants have been identified, accounting for 16-36% of the heritable risk of non-monogenic PD [55]. Other risk factors include prior head injuries, exposure to environmental toxins such as pesticides, and comorbidities like diabetes or RBD [56]. Patients with RBD are at the highest risk of developing a movement disorder within the next decade [57]. Potential preventive factors include smoking, coffee consumption, physical activity, and the use of anti-inflammatory drugs, however it remains uncertain whether these associations are truly causal [56].

Currently, there are no disease-modifying treatments available that can slow down or stop the progression of PD. Treatment primarily focuses on symptomatic relief, particularly alleviating motor symptoms. The main component of this approach is dopamine replacement therapy, typically utilizing Levodopa, which is favored for its relatively mild side effects at low to moderate doses during initial treatment [58, 59]. However, prolonged use of Levodopa at higher doses may result in dyskinesia [60]. As a result, it is often combined with other dopaminergic medications in the later stages of treatment to address response fluctuations and periods off-treatment [58]. To further minimize off-treatment time, device-aided therapies such as pumps, deep brain stimulation, or lesion surgery may be utilized [58]. Non-motor symptoms are generally managed with standard therapies tailored to the specific symptoms [61]. While several promising drug targets with potential disease-modifying effects have been identified, ongoing studies are mostly in early phases, and others have not met their defined endpoints [62–67].

1.1.3 Huntington’s Disease

HD is a less common, autosomal-dominantly inherited, neurodegenerative disease caused by a mutation in the huntingtin gene (HTT) gene, specifically a CAG repeat expansion [4, 68]. Its prevalence in the Western world is 4.88 per 100,000 individuals, with symptoms typically emerging between the ages of 35 and 45 [69]. In rare cases, the disease can manifest in childhood. The age of onset is inversely correlated with the number of CAG repeats in the HTT gene [70]. Following clinical onset, individuals with HD can expect to live an additional 15 to 20 years [71].

The mutated huntingtin protein induces neurodegeneration in the striatum, specifically causing loss of GABAergic medium spiny neurons [4, 71]. This neuronal loss is driven by the toxic gain-of-function associated with the abnormal conformation of the HTT protein, leading to protein misfolding and aggregation. These abnormalities disrupt cellular metabolic pathways, including the deregulation of ubiquitin and mitochondrial functions, resulting in the production of abnormal metabolites and oxidative stress [72, 73]. This oxidative stress, in turn, triggers inflammation at all stages of the disease [74].

HD clinically presents with a triad of symptoms: motor, cognitive, and psychiatric [4]. The most characteristic clinical symptom is chorea, which

consists of involuntary movements. Other motor symptoms include incoordination, dystonia, bradykinesia, and rigidity. The onset of these motor symptoms marks the manifestation phase of the disease, allowing for a definitive clinical diagnosis [75]. Cognitive and psychiatric symptoms, such as cognitive slowing, decreased attention and emotional recognition, depression, irritability, and apathy, often precede the initial motor symptoms [71, 76]. As the disease progresses, all symptoms worsen, leading to a decline in quality of life and an increased need for assistance [76, 77]. The phase prior to motor symptoms is known as the premanifest phase, which is further subdivided into the presymptomatic phase, characterized by the absence of symptoms and the prodromal phase, which features cognitive and psychiatric changes, along with subtle motor changes that are not clearly attributable to HD [4]. The clinical diagnosis of HD primarily relies on the assessment of symptoms, taking into account family history and genetic testing for the CAG expansion. Notably, neuronal loss can be detected early in the prodromal phase using MRI [53]. To standardize diagnosis and staging in clinical research, a new biological classification system called the HD Integrated Staging System (HD-ISS) has been proposed [53]. The HD-ISS consists of four stages, beginning with stage 0 for CAG expansion carriers. Progression to stage 1 occurs when biomarker evidence of HD pathophysiology is measurable through changes in putamen and/or caudate volume. Stage 2 is reached when cognitive and/or motor symptoms define a clinical phenotype, while stage 3 is characterized by additional functional changes. Other modern diagnostic and symptom recognition approaches, such as CSF or blood biomarkers [78], and digital technologies [79–81], are currently limited to clinical research or have not yet been validated for diagnostic use.

Unlike AD and PD, which are not typically linked to a single gene, HD is monogenic and associated with a mutation in the HTT gene on chromosome 4 [4, 69]. This mutation is characterized by an abnormally long CAG repeat. A normal phenotype consists of 10 to 35 CAG repeats, whereas more than 35 CAG repeats result in an HD phenotype. Individuals with 27 to 35 CAG repeats exhibit a normal phenotype, but these repeats can become unstable during reproduction, leading to de novo cases of HD, which account for up to 7% of all cases [4, 69]. The length of the CAG repeat is inversely correlated with the age of onset, meaning that a higher number of repeats is associated with an earlier onset of HD. CAG repeat length explains 50-70% of the variance in age of onset, while the remaining variation is influenced by other modifying genes, such as GRIN2A, ADORA2A, or PPARGC1A, as well as environmental factors [70]. Additionally, the number of CAG repeats is also related to disease progression, with a greater number of repeats leading to

faster progression [70].

The treatment of HD is purely symptomatic, as no disease-modifying therapies are currently available [73]. In addition to medications for managing symptoms such as chorea and depression, behavioral and social interventions, including physiotherapy and speech therapy, have demonstrated benefits for HD patients [71]. Ongoing research aims to develop treatments that reduce the amount of pathogenic protein by either decreasing its production or enhancing its clearance, thereby preventing HTT aggregation [73]. Although no drug has yet proven effective in clinical trials, a phase III clinical trial investigating the efficacy of pridopidine is currently underway (NCT04556656) [82].

1.1.4 The challenges in diagnosis and treatment of NDDs

In NDDs, the heterogeneity of these conditions and the absence of disease-modifying treatments present significant challenges for current management. These challenges are multifaceted, encompassing the heterogeneity in symptoms and disease progression in AD, PD, and HD, which often complicates diagnosis and leads to delays [3, 71, 83]. Furthermore, the development of disease-modifying treatments is hampered by the repeated failures of clinical trials, resulting in a reliance on symptomatic treatments alone [63, 73, 83].

AD, PD, and HD, exhibit considerable heterogeneity in their symptomatic manifestations and underlying mechanisms. While all patients share key pathologies, such as amyloid aggregation in AD and alpha-synuclein aggregation in PD, the underlying biology is multifactorial. This complexity poses significant obstacles to the development of disease-modifying drugs. Typically, drugs are designed to target specific molecular aspects of a single mechanism of action or biological pathway. However, the varying significance of these mechanisms among individual patients can impair the efficacy of drugs targeting a specific pathway [20]. In addition to the challenges in developing new treatment options, current symptomatic treatments for NDD are complicated by the high variability in symptoms among patients. While individuals share primary symptoms, such as tremors in PD and chorea in HD, the accompanying symptoms can differ markedly from one person to another, both in sequence and intensity [3, 84]. As a result, treatment must be tailored to address the specific type and severity of accompanying symptoms, requiring

clinicians to carefully evaluate potential side effects and drug interactions [58]. Ultimately, the primary goal is to enhance patients' quality of life. In PD, two distinct patterns of pathological progression throughout the body are characterized by different accompanying symptoms and their intensities, as elucidated by the brain-first versus body-first principle [45, 84]. These types are defined by the initial site of pathology: either in the brain or in the gut and peripheral nervous system. Research indicates that these types differ in their symptomatic presentation. The body-first type is often associated with RBD, constipation, and orthostatic hypotension, while the brain-first type less frequently exhibits these symptoms at the onset of parkinsonism but more commonly presents with a tremor-dominant phenotype and asymmetric symptom appearance. This variability in symptom presentation based on the starting point of pathology underscores the interaction between the underlying disease mechanisms pathology and their symptomatic manifestations. Neglecting this distinction may lead to ineffective disease-modifying treatments, as the two subtypes demonstrate significant differences in the spread of alpha-synuclein in the brain.

The underlying pathology of the diseases begins years before the onset of primary clinical symptoms, making early diagnosis essential for timely intervention [3, 71, 83]. While some symptoms may manifest prior to the main disease features, a definitive diagnosis is typically established only after these primary symptoms appear, as current diagnostic procedures do not permit diagnosis in their absence [53, 85]. This often results in delayed diagnosis, particularly for patients who do not initially exhibit key disease symptoms. This situation highlights the urgent need for biomarkers that enable early disease diagnosis independent of clinical criteria. Such biomarkers must demonstrate significant changes as soon as the underlying pathology begins. Despite considerable research in this area over the past few years, reliable biomarkers have only been established for AD, specifically the measurement of amyloid-beta in CSF [20]. For PD, potential biomarkers are still under investigation and have yet to be integrated into clinical diagnostic procedures [85]. In the case of HD, such biomarkers are still lacking [73]. Moreover, while amyloid-beta levels in CSF are currently utilized for AD diagnosis, they are not employed for preventive screening. Consequently, amyloid-beta levels are not routinely assessed in healthy individuals during regular check-ups, but only when there is a suspicion of AD. Additionally, the lumbar puncture required for CSF collection is an invasive procedure that can be uncomfortable for patients and is not typically performed by general practitioners. These limitations could be addressed with more accessible biomarkers, such as blood-based or digital biomarkers. Randomized clinical trials face significant

challenges in selecting appropriate endpoints. Standardized clinical rating scales, such as the Unified Parkinson’s Disease Rating Scale (UPDRS) for PD [86] and the Unified Huntington’s Disease Rating Scale (UHDRS) for HD [75], are commonly employed to assess the efficacy of new drugs. However, this approach has drawn criticism from researchers and clinicians for several reasons [3, 4, 71]:

1. **Disease Stage:** Patients with identical score on these rating scales may be at different stages of the disease based on underlying pathology and affected body structures, despite being classified at the same diagnostic stage.
2. **Symptom Bias:** Clinical rating scales either focus on specific disease symptoms or provide a broad overview that lacks the detail necessary to define subgroups. For example, the UPDRS consists of four components addressing non-motor features, daily living motor aspects, clinical motor examinations, and complications. Relying on total or individual components scores can lead to an overemphasis on motor symptoms, neglecting accompanying non-motor symptoms. Additionally, non-motor symptoms are typically rated via patient or caregiver self-reports, while motor symptoms are assessed by clinicians, introducing potential biases.
3. **Clinician Subjectivity** The subjective nature of these ratings can result in variability. Different clinicians may assess the same patient differently, leading to inconsistencies across study sites and even within the same patient over follow-up visits [3, 71]. Consequently, minor score improvements within a common one-year study timeframe may reflect either the drug’s efficacy or variability in ratings. Furthermore, clinical scales may fail to capture subtle improvements or deteriorations in slowly changing conditions, resulting in only minor score fluctuations.

Therefore, more objective symptom monitoring is essential for accurate disease diagnosis, monitoring, and trial endpoints. Objective scales could enhance the detection of early symptoms and accurately measure subtle changes, thereby facilitating the development of new, potentially disease-modifying treatments.

In conclusion, the treatment of NDDs presents a multifaceted challenge, stemming from the complexity of disease mechanisms and the necessity for precise, objective diagnosis and symptom monitoring. To address these challenges and enhance the potential for success in clinical trials, a systemic and

comprehensive approach is required. This involves untangling the complexity of disease mechanisms, developing targeted drug therapies, establishing methods for objective disease monitoring, and identifying disease subtypes. Such a holistic approach can be viewed as a value chain in biomedical research, encompassing drug development, disease management, and precision medicine.

1.2 The drug development process

Drug development is the intricate process of designing a new pharmaceutical treatment, encompassing five main phases: pre-discovery, drug discovery, pre-clinical research, clinical research, and subsequent stage of reviewing, approval, and post-market monitoring [87, 88]. The primary objective of this process is to ensure the safety and efficacy of the drug, thereby maximizing patient benefit. Below is a brief overview of each stage of the drug development process:

- **Pre-Discovery Stage:** During this phase, potential drug targets, such as proteins or genes, are identified and validated [88–90]. This requires a deep understanding of the underlying disease processes and disease-causing mechanisms. By investigating dysfunctional signaling pathways or molecular mechanisms associated with the disease, researchers can pinpoint viable drug targets.
- **Drug Discovery Stage:** The goal of this stage is to identify molecules or therapeutic strategies that alleviate disease symptoms, interfere with disease progression, or potentially cure the condition.
- **Pre-Clinical Development Stage:** Once drug candidates are identified, their mechanism of action and potential toxicity must be thoroughly investigated. This involves conducting numerous *in vivo* (e.g. animal models) and *in vitro* experiments to demonstrate the safety of the potential drug compounds.
- **Clinical Stage:** This phase aims to evaluate the drug candidates in humans, establishing their safety and efficacy [91]. The clinical stage is divided into three phases:

- Phase I: A small group of healthy individuals or patients, typically 20-100 people, receives the drug to assess safety and tolerability, with multiple doses tested to determine the most beneficial one [87, 88].
- Phase II: The drug is tested in a larger group of 100-500 patients to evaluate its therapeutic effect. During this phase, the optimal dose is further refined, and safety studies continue.
- Phase III: The drug’s efficacy is assessed in an even larger population, usually comprising 300-3000 patients, depending on the disease being investigated. This phase also tests the drug against standard treatments and examines potential interactions with other medications.

- **Reviewing, Approval, and Post-Marketing Monitoring Stage:**

After a successful Phase III study, all study data is submitted to regulatory bodies for review and market approval. Depending on the target market, submissions are made to regulatory agencies such as the FDA in the United States, the European Medical Agency (EMA) in the European Union, or local regulatory agencies in other regions. Once approved, the drug becomes available on the market and can be prescribed to patients. However, post-marketing surveillance studies, also known as Phase IV studies or real-world evidence trials, continue to monitor rare or long-term adverse effects in a larger population.

The drug development process is inherently lengthy, intricate, and costly, with high attrition rates. Out of numerous compounds tested, only a selected few progress to clinical trials, resulting in the approval of a single drug [88]. This entire process typically incurs costs of approximately \$2.8 billion and spans around 12 to 15 years [88, 90, 92]. The pre-discovery phase, lasting 5 to 6 years, and the clinical trial phase, taking 4 to 7 years, are the longest stages. In contrast, pre-clinical stage requires 2 to 3 years, while regulatory approval takes an additional 1 to 2 years [88].

In summary, the drug development process is vital for managing diseases. Newly developed drugs and therapies enhance patient care and improve the quality of life of many individuals. Thus, it is imperative to focus on refining the drug development process. A robust pre-discovery phase, which includes target identification and validation, is essential for successful outcomes. Consequently, substantial efforts are dedicated to understanding disease pathology and the complexities of underlying mechanisms. Modern

technologies, such as data-driven methods and artificial intelligence (AI), hold the potential to expedite this process and deepen our understanding.

1.2.1 Leveraging Data-Driven Approaches for Target Identification

New drug targets can be identified through various methodologies, with literature reviews and genomic studies being the most prevalent. These approaches operate in the premise that inhibiting or activating a specific protein, identified as a potential target, can modify the disease or significantly influence its phenotype [89, 93]. Thus, it is essential not only to identify a potential target but also to validate it to confirm its impact on the disease.

In addition to thorough literature reviews, a common partially data-driven approach is the use of genome-wide association studies (GWAS). GWAS leverage genomic data to explore the relationship between specific genetic mutations and diseases [89, 93]. Various statistical methods are employed to assess the associations between the identified variants and the disease, including single marker tests, burden tests, variance-component tests, and polygenic risk scores for multiple marker analyses [94]. Another data-driven approach focuses on examining mRNA and protein expression levels through transcriptomic or proteomic analyses [89, 93]. In transcriptomics, the emphasis is on identifying which potential targets are expressed in disease states. A range of methods is utilized to analyze the resulting data, including traditional techniques such as differential gene expression (DE) analysis and co-expression network generation [95]. DE analysis aims to pinpoint genes that exhibit differential expression between healthy individuals and those with the disease, thereby shedding light on potential disease-causing genes. Conversely, co-expression networks aim to identify interacting genes within mechanisms or pathways pertinent to the disease process. These analyses are often complemented by pathway or gene-set enrichment analysis (GSEA), which provides functional annotation by assessing the enrichment of specific pathways within a gene set [93]. In transcriptomics, GSEA aids in identifying over- or underexpressed pathways that may be influenced by the disease, thus highlighting these pathways or mechanisms as potential targets for new drug development.

Once a potential target is identified, it is imperative to validate it to ensure that it influences a relevant aspect of the disease biology, rather than

extraneous factors. Traditional validation methods often involve transgenic knock-out (KO) or knock-down (KD) animals, which lack the gene of interest, allowing researchers to observe the resultant phenotypic changes [89, 93]. Similar gene expression perturbation experiments can be conducted using human cell lines, however, these methods are typically time-consuming and costly [89, 93]. Consequently, data-driven validation approaches can facilitate more informed decisions at a reduced expense.

A critical step in drug target identification is understanding the mechanistic relationship between the identified target and the disease. It is essential to delineate the biological pathways and mechanisms that connect the potential target to the disease. This mechanistic understanding is usually achieved through systems biology or systems medicine approaches [96]. Such models provide deeper insights into the molecular mechanisms underlying disease phenotypes by identifying key components and predicting system behavior. The complexity of biological systems is often represented through networks, where biological entities such as genes or proteins act as nodes, and their interactions are depicted as edges [96–98]. Examples of these networks include protein-protein interaction (PPI) networks, drug-target interaction networks, disease-gene association networks, and biological pathways. Current knowledge regarding these interactions is stored in various databases, which facilitate the application of this information in algorithms and models. Notable databases include STRING [99], BioGrid [100], and IntAct [101] for PPIs; OpenTargets [102] and the Therapeutic Targets Database [103] for drug-target interactions; DisGeNET [104] for disease-gene associations; and KEGG [105], Reactome [106], PathwayCommons [107] and WikiPathways [108] for biological pathways. While these databases streamline the integration and utilization of knowledge for various algorithms and models, they primarily provide generic information lacking specific disease context, affected tissues, or cell types. As a result, they may not fully capture the complexities of the disease under investigation.

To derive quantitative insights and predictions from these models, ordinary differential equations (ODEs) and partial differential equations (PDEs) are frequently utilized in systems biology [109–111]. These mathematical and computational frameworks accurately describe the interactions among biological components, enabling simulations of specific biological processes and predictions regarding the effects of perturbations, such as drug treatments or gene KDs [109, 111]. Quantitative models facilitate the prediction of overall phenotypic effects and the assessment of how changes in one part of the network may influence the entire system, including individual components.

However, the formulation of ODEs and PDEs requires a thorough understanding of the underlying biochemical reactions, which is often available only for specific processes, such as insulin signaling in type 2 diabetes [112] or amyloid-beta aggregation in AD [113]. Additionally, fitting differential equations typically necessitates time-resolved data and the ability to conduct intervention experiments for validation. This poses challenges in NDDs, as cell lines and mouse models often replicate only certain aspects of the human disease. Such limitations hinder the reconstruction of complex disease mechanisms, especially in conditions where the underlying biology remains only partially understood. Alternative methods for constructing systems biology networks include probabilistic models, such as probabilistic Boolean networks [114] and Bayesian networks (BNs) [115, 116]. These approaches can yield valuable insights even in the absence of detailed biochemical knowledge or time-resolved data.

Data-driven methods, including machine learning (ML) and AI, are crucial for drug development. These methodologies can expedite the entire process, reduce costs, and enhance the success rate of new drugs by facilitating informed, data-driven decisions. They have proven to be powerful tools for analyzing the rapidly growing volume of data and integrating heterogeneous sources of information to gain a deeper understanding of the molecular mechanisms underlying disease phenotypes.

1.2.2 Mechanistic modeling in NDDs

Current drug targets in NDDs primarily include amyloid beta, tau, APOE4 and inflammation-related proteins, such as TREM2, and CD38 in AD [117]. In PD, key targets include LRRK2, GBA, alpha-synuclein pathology, and PINK1 [118]. However, many additional drug targets have been identified and are currently under investigation. In HD, the main focus is on the huntingtin gene, which is implicated in the disease’s pathology [73].

Developing disease-modifying treatments for NDDs requires a profound understanding of the complex underlying mechanisms involved in disease pathology. Numerous quantitative mechanistic models exist for AD [119–125] and PD [126, 127]. For example, Proctor et al. developed a model for the aggregation of amyloid-beta and tau protein, examining the interplay between GSK3beta, p53, amyloid beta, and tau [113]. This model incorporates components for each of these factors, with the underlying network derived

from prior research and biological knowledge, while parameter estimates are based on data from cellular systems. Another notable model in AD, developed by Hao et al., simulates multiple cell types and their interactions concerning amyloid beta aggregation and tau proteins through a PDE system [120]. This model includes components such as amyloid beta, IL-10, and TNF-alpha, and the numbers of living and dead neurons, with parameter values sourced from previously published research. Van Maanen et al. formulated a model that describes changes in the APP pathway, specifically modeling the effects of BACE1 inhibition on APP metabolites like Abeta42 or Abeta10 in the CSF [121]. While, this model provides valuable insights into the underlying mechanisms, it primarily focuses on the specific effects of a single intervention and does not generalize to the system level. Most quantitative models concentrate on specific aspects of AD, primarily targeting amyloid beta pathology [120, 121, 123], tau pathology [119, 125], or a combination of both [113, 124]. Few models establish a direct link to the disease phenotype beyond protein aggregation. For instance, Geerts et al. calibrated their model using ADAS-Cog score from clinical trials [123], while Mazer et al. included the Clinical Dementia Rating – Sum of Boxes as a clinical outcome [124]. Although the latter incorporated clinical data to estimate certain model parameters, others were derived from cellular data or fixed values determined by the authors. There is a pressing need for a systematic approach that encompasses multiple pathways and mechanisms rather than isolating specific disease aspects. This approach should leverage extensive clinical data, with the network structure inferred directly from the data rather than being manually defined through equations.

1.3 Disease management in NDDs

Disease management refers to a systematic and coordinated approach to healthcare for patients affected with chronic diseases. Its primary goal is to enhance health outcomes and improve patients' quality of life, thereby elevating overall healthcare standards. This approach encompasses patient education regarding their conditions and self-care options, continuous monitoring of health status, the implementation of personalized treatment plans, and coordinated care among healthcare providers. All these elements emphasize patient-centered disease management. While clinicians play a pivotal role in this process, patients are also active participants in managing their own health. Empowering patients through education programs and providing

them with knowledge about their conditions and treatment options is essential for improving adherence and positively influencing disease outcomes. Furthermore, enhancing patients’ understanding fosters greater trust in the healthcare system.

Personalized treatment plans, an essential aspect of disease management, are tailored to each individual patient. Recognizing that symptoms can vary widely among patients and may differ in severity, a diverse array of medical treatments and non-medical interventions, such as physiotherapy, psychotherapy, or speech therapy, are recommended [128, 129]. Clinicians formulate these plans by assessing symptoms and comorbidities while integrating genetic, environmental, and lifestyle factors. For instance, patients with PD have been shown to benefit from activating therapies like physiotherapy and occupational therapy [128, 130]. Overall, physical activity is associated with improved quality of life and better symptom management, while specific therapies target particular symptoms. For example, physiotherapy, aquatic exercise, and endurance training can enhance motor symptoms, whereas resistance and treadmill training improve gait [130, 131]. In terms of medical interventions, careful administration of Levodopa or other dopamine agonists in younger patients is crucial, particularly concerning the risk of dyskinesia associated with long-term Levodopa use [60, 128]. This approach aligns with the principles of precision medicine, also known as personalized [132] or stratified medicine [133], which focuses on delivering the right drug to the right patient at the right time. Unlike traditional “one-size-fits-all” treatment strategies that apply average effects to all patients with the same diagnosis, personalized medicine tailors treatment based on individual characteristics [134]. The underlying principle is that variability in human biology and environmental exposures leads to different disease progressions and drug responses among patients diagnosed with the same condition [135]. Precision medicine emphasized stratifying patients into subgroups based on their individual characteristics and matching them with the treatments or interventions likely to yield the best outcomes [136]. Identifying these subtypes is thus a fundamental task in precision medicine.

In addition to providing optimal treatment, effective disease monitoring is a crucial aspect of disease management, allowing for timely responses to changes and evaluations of treatment plans. Regular visits are essential, as symptoms of NDDs are not static but evolve throughout the course of the disease. Symptoms and their progression in NDDs are typically assessed using clinical rating scales, such as the UPDRS for PD [86], or UHDRS for HD [75]. Both scales evaluate various aspects of the diseases, including non-motor

and motor symptoms in PD through multiple subscales. The UPDRS ranges from 0 to 260, with higher scores indicating greater disability. The UHDRS also comprises multiple subscales focusing separately on motor and cognitive assessment. While higher scores in the motor and behavioral sections indicate greater impairment in HD, higher scores in cognitive and functional assessments suggests better performance. However, these clinical outcome measures are subjective and highly variable due to intra-rater and inter-rater differences [3, 4, 72, 137]. Such variability can affect clinical trial results and influence patient treatment evaluations. For example, if a patient begins treatment with a UPDRS score of 124 and later receives a different rating from another clinician, who assessed the patient more critically, resulting in a score of 136, it becomes unclear whether this change reflects disease progression or merely variability in scoring. Moreover, clinical scales may fail to capture subtle improvements or declines, particularly when changes in the patient’s condition occur gradually, leading to minor score fluctuations. Therefore, there is a pressing need for more objective assessment of symptoms and their progression. In summary, effective disease management relies on both optimal treatment and thorough symptom monitoring, achieved through the identification of progression subtypes and the implementation of objective symptom monitoring.

1.3.1 Progression subtype identification

The experience of each patient with NDDs is uniquely individual, characterized by significant variability in symptoms. The severity and progression of these diseases can vary markedly among individuals, reflecting diverse disease trajectories. Although all NDDs are progressive, leading to a gradual decline in symptoms over time, some patients may face rapid deterioration, while others may experience a more gradual decline. Identifying progression subtypes is essential for effective disease management, as it lays the groundwork for determining optimal treatment strategies and addressing the inherent heterogeneity of these conditions. NDDs manifest a broad spectrum of symptoms and progression patterns, making it vital to comprehend this heterogeneity for implementing precision medicine approaches in research, patient counseling, and the development of successful disease-modifying therapies. Recognizing distinct disease subtypes offers several advantages. First, it enables predictions regarding disease progression, providing valuable prognostic information for patient counseling. Second, it aids in anticipating treatment responses. Lastly, delineating different subtypes can pave the way

for new therapeutic targets, as responses to existing or novel treatments may vary among subtypes. Clinical trials that evaluate the efficacy of new drugs based on identified subtypes may reveal effectiveness in specific groups that broader studies, those that do not account for subtype differences, might overlook. Furthermore, understanding their specific disease subtype empowers patients to manage their condition more effectively, allowing clinicians to deliver more accurate prognoses and customize treatments accordingly.

Subtype identification can be approached through various methods, tailored to the data modalities and techniques employed. A straightforward classification method relies on individual clinical features, such as age of onset, motor phenotype, or the onset of dementia [138]. For example, in HD, patients are categorized by their age of onset: those with juvenile-onset HD present symptoms before age 21, while adult-onset HD is diagnosed thereafter [139]. Other subtypes may be identified using univariate modalities based on the distribution of pathological markers, such as amyloid beta or tau [140, 141]. Additionally, while stratifications can go beyond clinical features to encompass genetic [142] or imaging data [140], the predominant categorization for subtyping in NDDs relies on clinical data that details patient symptoms and clinical phenotypes [143–151]. This clinical data is often supplemented by an additional modality [152, 153], making it the most widely used approach in the field. Traditional clinical approaches often classify patients based on predominant motor symptoms. In PD, two main subtypes have been identified: one characterized by postural instability and gait difficulties, and another defined by tremor [154]. In HD choreatic and hypokinetic-rigid subtypes have been recognized [155]. More recently, imaging studies in PD have revealed the hypothesis of brain-first and body-first subtypes, which correspond to different origins and routes of alpha-synuclein spread in the nervous system [45, 156].

Similar to the various data modalities used for subtype identification, a variety of methods exists for conducting this analysis. Typically, these analyses involve addressing a clustering problem using techniques such as k-means clustering [143, 145, 147–149], hierarchical clustering [152, 153] or a combination of both [151]. These simpler methods are often enhanced by more sophisticated approaches such as gaussian mixture models (GMMs) [144].

The approaches previously mentioned often neglect the variations in disease progression dynamics. Most methods identify subtypes based solely on a single timepoint, while few studies comparing subtype characteristics at additional

timepoints not utilized for clustering [141, 146]. Furthermore, very few studies incorporate progression information directly into the clustering process [157–159], despite the advantages that progression subtype identification could offer in enabling effective early interventions for patients in the initial stages of the disease. Predictive models can classify patients as fast or slow progressors early in their disease trajectory. Disease progression clustering for subtype identification can be accomplished using deep learning techniques, such as Long Short-Term Memory (LSTM) networks [159] or the Variational Deep Embedding with Recurrence (VaDER) method developed by our group, which integrates recurrent variational autoencoders with GMM clustering in latent space [160].

A notable challenge in subtype identification lies in the reproducibility of findings. Although numerous studies have investigated subtypes in NDDs, only a limited number validate or replicate their results using independent datasets. However, a recent study from our group successfully identified reproducible progression subtypes in PD across three datasets, which were further correlated to the brain-first versus body-first principle [158].

In contrast, subtype identification in HD is relatively limited, with only a few recognized subtypes, namely the clinical choreatic versus hypokinetic-rigid subtypes and the distinction between juvenile and adult-onset forms. A recent study employed principal component analysis and k-means clustering to identify three distinct subtypes based on clinical features, marking the first instance where non-motor symptoms, including cognitive and psychiatric manifestations, were taken into account [143]. However, this clustering was conducted at a single timepoint and did not consider the variability of symptoms throughout the disease course. As a result, the identification of progression subtypes remains an important area for further research in HD science.

1.3.2 Objective symptom monitoring

As discussed in chapters 1.1.4 and 1.3, variability among raters in assessing patients using clinical scores impacts symptom monitoring. This inherent subjectivity must be minimized to enhance the success of clinical studies and improve individual disease evaluations, ultimately leading to better disease management for both patients and clinicians. Transforming this subjectivity into objective assessments is essential. Digital technologies can

play a pivotal role by quantitatively recording measures, thereby providing digital biomarkers that facilitate disease monitoring, early diagnosis, and the prediction of clinical outcomes [161].

The landscape of digital technologies for improved disease monitoring is vast. This overview will focus on the most commonly utilized technologies, including smartphone devices and applications, and wearable sensors. I will subsequently discuss the specific features that can be extracted from these technologies and their current applications in research and patient treatment.

Data collection using digital technologies can generally be classified into passive or active methods. Active measurements involve administering questionnaires to patients via computer or smartphone applications. A notable example is the Mobile Parkinson Disease Scores, which employs a ML model to analyze features derived from smartphone activities [162]. In contrast, passive measurement occurs without direct data collection; instead, patients perform specific tasks recorded by devices such as sensors. The data from these tasks is processed to extract digital biomarkers for further analysis. Examples in the NDD area include voice recordings [163–167], videos of facial movements [168–171], and wearable sensors [172–175]. In voice recordings, researchers extract features such as phonation, prosody, and speech pauses for analysis [161, 176]. Video recordings are similarly utilized to examine specific facial action units [171]. However, the most prevalent technology is wearable sensors, which detect gait disturbances in PD patients [172, 173, 177–181]. These sensors, often affixed to patient’s shoes, extract gait parameters such as stride length and gait speed during defined walking tests, such as the Timed Up and Go (TUG) test, where patients stand up, walk 3.5 meters, turn around and return.

While digital gait features have been explored in numerous studies within PD research, most investigations primarily focus on distinguishing PD patients from healthy controls [180, 182–187]. Although some studies classify patients into various disease stages [188–191], few directly investigate the relationship between PD progression and gait [175]. A wide array of devices has been developed; however the reproducibility of findings from the same gait device remains insufficient [137]. As a result, despite significant research on digital technologies for objective symptom monitoring, further work is needed to integrate these technologies into standard disease management and as endpoints in clinical trials. Notable examples include the use of smartphone-based digital biomarkers in a phase I clinical study of PD [192] and a phase II study demonstrating the potential of sensor-based digital gait analysis

to address limitations of traditional rating scales [193]. Before integrating digital biomarkers into routine clinical assessments and trials, several essential steps must be undertaken as outlined by the Movement Disorders Society and others [137, 194, 195]. The primary requirement is to establish the validity and clinical relevance of digital measurements, ensuring a quantitative relationship between these measures and traditional outcome metrics. This will help mitigate skepticism among clinicians and patients regarding the use of devices for disease diagnosis and monitoring. From a regulatory perspective, it is crucial to analyze which disease-relevant concepts the device captures, such as gait, and how these measures can be practically applied. Additionally, devices like gait sensors must receive approval from regulatory agencies to enhance trust and ensure adherence to standards. Finally, the advantages of digital measurements compared to traditional outcomes for both patients and clinicians need to be evaluated before full integration into clinical practice.

In summary, digital technologies are invaluable for objective symptom monitoring and will play an increasingly vital role in management of NDDs and as endpoint in clinical trials. Clinical trials will benefit from the ease of use of these technologies, even within patients' home environments, potentially reducing overall costs. Likewise, disease management will improve through more rapid and objective measurements of symptoms and disease progression, allowing clinicians to adjust treatment plans accordingly. However, to realize these benefits, the validity and clinical relevance of these technologies for disease monitoring must be robustly demonstrated.

1.4 Research contributions of this thesis

In this chapter, we will delineate the contributions of this thesis to drug development and disease management, with a particular emphasis on NDDs such as AD, PD, and HD. This work addresses three pivotal objectives through data-driven methodologies aimed at deepening our understanding and management of these conditions:

- Development of a comprehensive, multi-scale quantitative system medical model for AD, enabling the simulation of perturbation experiments to prioritize drug targets informed by existing biological knowledge [196] (Chapter 2).

- First-time identification, clinical characterization, and predictive modeling of HD progression subtypes, underscoring the essential role of non-motor symptoms in personalized treatment while enhancing predictive modeling to improve patient care and quality of life [197] (Chapter 3).
- Exploration of the potential of gait sensor measurements to monitor PD symptoms and progression, establishing their validity and clinical relevance, as well as their prospective use as surrogate endpoints in clinical trials [198] (Chapter 4).

In Figure 1, we depict the objectives of this thesis within a value chain focused on the development of personalized disease management and enhanced disease management.

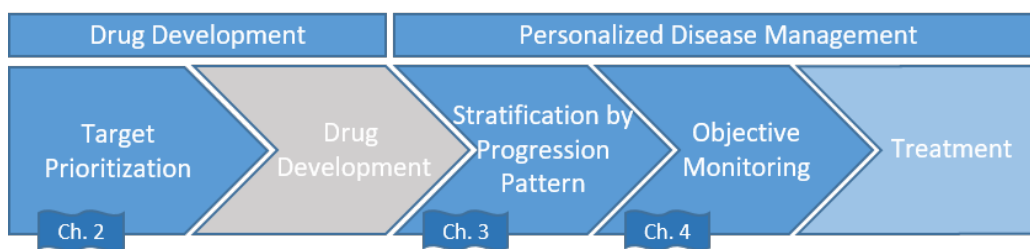


Figure 1: Value chain of drug development and personalized disease management. Drug development includes target prioritization, a critical process where biological modeling, as presented in Chapter 2, serves as a key strategy to investigate potential targets and their effects. Upon identifying a target, it is essential to discover molecules that can intervene in the disease process. While this step is crucial within the value chain, this thesis does not address it. In the field of personalized disease management, progression subtyping, including characterization of identified subtypes, and objective monitoring of symptoms and progression are necessary steps to achieve a more effective personalized treatment strategy. The research conducted in this thesis on these topics is presented in Chapter 3 and Chapter 4.

The objectives of this thesis align with the value chain for developing personalized medical treatments, beginning with the identification of potential drug targets. In Chapter 2, we detail our efforts to discover and validate biological factors critical to the AD process, with a particular focus on the CD33 molecule. The multi-scale quantitative model we developed illustrates the intricate interplay among relevant molecules and biological processes, extending to the phenotype level by integrating patient-level clinical data and existing biological knowledge. We utilized the established Variational

Autoencoder Modular Bayesian Network (VAMBN) model, which employs variational autoencoders to independently encode multiple data modalities, subsequently generating a BN that connects these autoencoder modules [199]. In our study, these modules represent gene components annotated with mechanisms associated with AD. For the first time, we incorporated prior knowledge in the form of a knowledge graph to enhance the generation of the modular BN. Using this model, we predicted how alterations in one part of the network could influence the entire system. Our CD33 knock-down simulation revealed significant changes across multiple disease mechanisms and phenotypic outcomes associated with AD, demonstrating the model’s capacity to prioritize drug targets by predicting the effects of perturbations throughout the network, extending beyond just CD33. We further validated our simulations using data from a CD33 knock-out cell line. Overall, this research underscores the potential of a multi-scale quantitative model that integrates diverse data modalities, including gene expression (transcriptomics), demographic information, brain pathophysiology and cognitive scores, all informed by existing biological knowledge. Through these biological networks, we can forecast the effects of perturbations on clinical scores and disease phenotypes, thereby enhancing the reliability of our predictions.

While a quantitative, mechanistic model of a disease provides valuable insights, effective personalized disease management requires a thorough understanding of which patients are most likely to benefit from new treatments. This necessitates a comprehensive understanding of disease progression, as discussed in Chapter 1.3.1. In Chapter 3, we identify and characterize two distinct progression subtypes of HD for the first time, thereby contributing to the clinical characterization and predictive modeling of these subtypes. To achieve this, we applied the previously developed VaDER model, which integrates GMM with variational autoencoders and LSTM networks. This innovative approach allows us to cluster relatively short time-series data, even in the presence of missing values. Furthermore, we align patient trajectories on a common disease timescale to mitigate confounding effects arising from temporal associations with the study baseline. Our analysis of feature importance within the predictive model for these two subtypes revealed significant differences in cognitive abilities, particularly highlighting a marked distinction in executive function between slow- and fast-progressing patients. This underscores the critical role of non-motor symptoms in predicting disease progression and tailoring personalized treatment. Objective assessments of disease progression not only enhance patient’s quality of life but also enable personalized counseling and treatment strategies. Targeted prevention of cognitive decline is essential, and personalized medical services facilitate the

implementation of appropriate therapies and ongoing adjustments to care.

Once patients start a new treatment, clinicians must carefully monitor its effects and evaluate the overall disease status, including symptomatic expression and disease progression. As discussed in Chapter 1.3.2, objective monitoring of disease symptoms is preferred. In Chapter 4, we assess the potential of digital gait sensor measurements collected from PD patients during various gait tasks. We first explored the relationship between the collected gait features and traditional clinical scores, such as UPDRS, to demonstrate the efficacy of gait assessments in monitoring disease symptoms and their progression. Our analysis indicates that digital features effectively track motor symptoms and their longitudinal progression. Subsequently, we examined the predictive value of these digital gait sensor measurements for assessing PD symptoms at future visits. While we found a clear correlation between gait features and clinical scores related to disease stage, the correlation with disease progression was minimal, and the prognostic value was limited. However, we observed a trend suggesting enhanced prognostic value and monitoring of disease progression when analyzing features from tasks involving longer walking distances. Importantly, our study underscores the potential of digital gait assessments as surrogate endpoints in clinical trials. We found that a digital UPDRS score derived from gait features could effectively serve this purpose. Moreover, using such surrogate endpoints in clinical trials reduces the required sample size, as gait assessments can be repeated at shorter intervals, potentially even in an in-home setting. This underlines the great potential of the gait device as an alternative and easy-to-assess endpoint predictor in clinical trials for PD.

In Chapter 5, we summarize the key findings of this thesis, address their limitations, and outline potential future directions for research in data-driven drug discovery and disease management.

2 Comprehensive Multi-Scale AD Model for Knock-Down Simulations

In this section, we summarize our publication presented fully in **Appendix A.1**).

Raschka, T., Sood, M., Schultz, B., Altay, A., Ebeling, C., Fröhlich, H. (2023). AI reveals insights into link between CD33 and cognitive impairment in Alzheimer’s Disease. *PLOS Computational Biology*, 19(2), e1009894.

Summary

AD is characterized by a progressive decline in the ability to perform daily activities, resulting from the death of brain cells and subsequent brain shrinkage [200]. The field of AD research has faced persistent challenges failures in developing effective medications, largely due to an insufficient understanding of how specific molecules contribute to cognitive decline [200]. Consequently, gaining deeper better mechanistic insights into the disease is essential for creating new, urgently needed treatments that can modify its course. One promising therapeutic target is CD33, a transmembrane receptor protein. Genetic studies have linked CD33 to a reduced risk of AD, while experimental data indicate that its suppression enhances amyloid- β clearance in mice and cell lines [31–35, 201]. Therefore, exploring the biological mechanisms surrounding CD33, without delving into the complex biochemical details required by traditional modeling methods such as ordinary or partial differential equations, could be pivotal for advancing our understanding of AD.

In our publication titled *AI reveals insight into link between CD33 and cognitive impairment in Alzheimer’s Disease*, we integrate prior knowledge about disease mechanisms and their interactions in the form of a disease-focused AD knowledge graph into the development of a quantitative system-level, multi-scale model that connects molecular dynamics to phenotypic outcomes. We utilized joint clinical and patient-level gene expression data from three

independent studies (ROSMAP, Mayo, and cell-line experiments) to train and validate a hybrid AI approach. This approach combines heterogeneous incomplete variational autoencoders with modular Bayesian networks, incorporating a disease-focused knowledge graph. The previously developed Variational Autoencoder Modular Bayesian Networks (VAMBN) algorithm, was trained using patient-level data organized into modules, defined by either clusters within the knowledge graph or by patients’ phenotypes, alongside demographic and clinical variables. This model quantitatively captures the relationships between nodes, generating a probabilistic graphical model that illustrates the connections between modules and original features. It allows for a per-patient score for each module, which can be decoded into feature-level gene expression and phenotype data. Moreover, the model facilitates the simulation of the down-regulation of various putative drug targets, including CD33, to evaluate their potential impact on cognitive impairment and brain pathophysiology. While CD33 is the primary focus of this work, the analysis is not confined to this target, instead all genes incorporated within the model can be examined.

Using ROSMAP data for training, the resulting BN revealed connections between various biological mechanisms involved in AD. While age was the only factor exerting direct influence on CD33, the receptor itself exhibited a predominantly indirect predicted causal influence on every node except the source nodes defined by years of education, age, and expression of the NAV3 gene. The shortest path between CD33 and the disease phenotype was identified through the prostaglandin pathway, with all connections within this path being newly learned and not previously defined in the knowledge graph. Furthermore, the model identified 130 new edges with high bootstrap confidence, including six edges consistently discovered across iterations. These edges illustrate robust pairwise connections, such as between NAV3 and MAVS, a member of the TGF-beta subgraph. External validation of the model using data from the Mayo study demonstrated the models’ superior ability to explain normalized gene expression data from an independent study compared to a randomly permuted network ($p=0.035$) despite clinical differences between the study populations. The model also facilitates intervention experiments, allowing for a simulation of CD33 down-expression to explore the potential systemic consequences of a therapeutic intervention targeting CD33. This simulation yielded significant improvements in cognitive function and perturbations in several biological mechanisms. Noteworthy outcomes included increased activation scores in the prostaglandin module, a significant rise in mini mental state examination (MMSE) scores, and a significant decrease in Braak stages. Additionally, significant effects were observed on 28 mechanisms

and genes, including the calpastatin-calpain and amyloidogenic mechanisms. Experimental validation using a knock-out THP-1 monocyte cell line revealed a high degree of overlap with the predicted molecular mechanisms, further substantiating the model’s predictions.

Overall, this work has advanced the systemic and quantitative understanding of AD. For the first time, we demonstrate the feasibility of integrating gene expression and clinical data with knowledge of cause-and-effect relationships into a quantitative systems medical model for AD. This was achieved through the training and validating of an AI-based model, complemented by a knowledge graph representation. The analysis revealed a connection between CD33 and the phenotype, highlighting correlations with prostaglandins, which previous research has linked to memory, learning, and neuroinflammation - key elements of AD [202–204]. Additionally, newly identified relationships, such as between NAV3 and MAVS, a member of the TGF-beta pathway, are both associated with AD, but lacking a known direct correlation, suggesting promising new avenues for research [205, 206]. Thus, the impact of this study is two-fold. First, we introduced a novel multiscale, quantitative modeling approach that is broadly applicable in systems medicine, particularly in contexts where only a partial mechanistic understanding of biological phenomena exists. Second, the developed model offers a valuable resource for the AD research community, facilitating a deeper understanding of the disease and the identification of novel therapeutic opportunities.

Authors’ contributions

Tamara Raschka contributed significantly to the study methodology, and data curation. She conducted the formal data analysis and interpretation, developed the majority of the necessary software, and created the visualizations. Additionally, she was responsible for drafting the manuscript.

3 Identification and Predictive Modeling of HD Progression Subtypes

In this section, we summarize the publication presented fully in **Appendix A.2**).

Raschka, T.¹, Li, Z.¹, Gassner, H., Kohl, Z., Jukic, J., Marxreiter, F.², Fröhlich, H.² (2024). Unraveling progression subtypes in people with Huntington’s Disease. *EPMA Journal* , 15, 275-287.

Summary

HD is characterized by a diverse range of symptoms, including hyperkinetic movement disorders, psychiatric manifestations, and cognitive deficits, all of which significantly impact patients’ quality of life [68, 69, 76, 77, 207–211]. While the genetic basis of HD is well-established [68, 212, 213], the disease is progressive and multifactorial, marked by a lengthy pre-manifest phase. This complexity results in highly variable and heterogeneous symptoms among patients, complicating counseling efforts [208, 214, 215]. Currently, the diagnosis of HD relies solely on the presence of motor signs, neglecting other critical aspects of the disease, such as cognitive and psychiatric symptoms, despite their significant influence on daily life [216–218]. By integrating both motor and non-motor features of HD into the diagnostic process, we can enhance accuracy and improve patient care [157, 215, 219–222].

In our publication titled *Unraveling progression subtypes in people with Huntington’s Disease*, we identified, characterized, and validated two distinct progression subtypes of HD. This analysis, conducted for the first time, utilized longitudinal, multisymptom disease trajectories from patients enrolled in the Enroll-HD study [223, 224], considering both motor and cognitive symptoms. Recognizing that longitudinal trajectories often correlate temporally with

¹Joint first authors.

²Joint last authors.

the study baseline, we employed a non-linear mixed-effect model to mitigate this confounding effect [225, 226]. This approach allowed us to adjust the longitudinal trajectories of pre-manifest and manifest HD patients along a common disease timescale before incorporating them into the clustering performed by our previously developed VaDER algorithm [160]. Furthermore, we trained and validated ML prediction models [227] to enable early prognosis of the disease progression subtype using data solely from the first visit.

Modeling multiple clinical outcomes, including UHDRS Total Motor Score, Symbol Digit Modality Test (SDMT), and MMSE, from 11093 manifest and pre-manifest patients along a common, potentially unobserved (latent) disease timescale resulted in aligned trajectories. Notably, the trajectories of pre-manifest patients are predominantly shifted leftward on the time axis, indicating an earlier stage of disease progression. This aligns with the natural course of HD, where patients stay in a pre-manifest state prior to the manifest stage. Analysis of the aligned trajectories of manifest patients revealed two distinct subtypes: 1) a large cluster (n=7122) characterized by stable disease progression and minimal impairment across all three outcome measures throughout the common disease timescale, and 2) a smaller cluster (n=411) exhibiting a markedly more progressive disease trajectory, with significant declines in the SDMT and MMSE scores, accompanied by an increase in the Total Motor Score. When training a prognostic model for HD subtypes using data from patients' initial visit, we achieved robust classification performance (ROC-AUC: 95%). This performance improved further with the incorporation of data from subsequent follow-up visits. Our model also outperformed ground-truth models that predict disease progression based on CAG repeat length, age, and sex. Feature importance analysis revealed that cognitive scores are essential for predicting HD progression subtypes, although motor scores, CAG repeat length, and neurobehavioral scores also play significant roles. Further clinical characterization of the subtypes indicated correlations with CAG repeat length and neurobehavioral, psychiatric, and cognitive scores. Notably, cognitive aspects ranked among the top five most influential features, emphasizing that the second, smaller cluster exhibited a higher prevalence of cognitive impairment. These findings were validated by applying both the VaDER and the prognostic model to pre-manifest patients, where we observed that the majority of patients were accurately grouped into the larger cluster, demonstrating the generalization capability of our models. In conclusion, this study elucidates the progression subtypes of HD and facilitates an objective assessment of disease progression by incorporating non-motor symptoms, which are vital for accurate disease prognosis.

Authors' contributions

Tamara Raschka made substantial contributions to the study's methodology and led the data analysis, which was conducted in collaboration with Zexin Li. She interpreted and discussed the results and took the lead in drafting the manuscript, collaborating closely with the other co-authors throughout the process.

4 Gait Sensors for Objective Monitoring PD Symptoms and Clinical Trial Endpoints

This chapter is an adaptation of our work in “Objective monitoring of motor symptoms and their progression in Parkinson’s Disease using a digital gait device, Preprint Research Square, 2024, doi: 10.21203/rs.3.rs-4521747/v1” [228]. All figures presented in this chapter were also used in the manuscript currently under review at Scientific Reports.

4.1 Introduction

Typically, standardized clinical rating scales, such as the UPDRS for PD [86], serve as endpoints in clinical trials to evaluate the efficacy of new drugs. However, this methodology has faced criticism due to the subjective nature of these scales, which can compromise symptom monitoring [3, 4, 71]. A detailed discussion of these challenges is presented in Chapter 1.1.4. It is crucial to transform this subjectivity into objective assessments. In this context, digital technologies can play a pivotal role in PD research by quantitatively capturing data, thus providing digital biomarkers that enhance disease monitoring, facilitate early diagnosis, and predict clinical outcomes [161].

The landscape of digital technologies for disease monitoring is extensive, offering numerous options such as smartphone applications and wearable sensors. The latter are particularly beneficial for monitoring impaired gait patterns in PD [172, 173, 177–181]. These sensors, often attached to patient’s shoes, collect gait parameters such as stride length and gait speed during walking tests. Gait abnormalities significantly impact patient mobility and quality of life, thus, identifying signs of impaired gait patterns, such as reduced gait speed, shorter step length, increased axial rigidity, and impaired rhythmicity, is essential [46, 48, 49, 51]. Differences between PD stages are also notable, beginning with slower speed and shorter steps in the early stage,

progressing to worsened balance and postural control, increasing the risk of falls [49]. Consequently, investigating patients' gait impairments is critical for evaluating disease progression.

Wearable sensors that offer objective monitoring can enhance clinical studies and disease management by optimizing treatment schedules, adjusting dosages, or transitioning between therapies. They are portable, lightweight, cost-effective, and user-friendly for both patients and clinicians, allowing for gait parameter measurements in outpatient settings [177].

Despite the clear advantages of digital gait parameters, there is an ongoing need to rigorously assess their technical robustness and medical diagnostic validity [161, 177, 182, 185, 187, 189]. Furthermore, regulatory agency approval is required for their use in routine medical care. It is essential to provide clear evidence that sensor measurements correlate significantly with established clinical outcomes such as UPDRS. Unfortunately, most current studies primarily focus on classifying disease stages or differentiating PD from healthy individuals using data from digital gait devices, with limited exploration of the direct associations between disease progression and gait [180, 182–191]. Additionally, there has been no investigation into the reproducibility of findings across several PD cohorts [137]. Thus, the capacity of digital gait sensors to reliably and objectively monitor motor symptoms and their progression over time remains uncertain.

In this study, we evaluate the potential benefits of digitally assessed PD symptoms compared to established clinical outcomes across two independent longitudinal cohorts. We utilized the sensor-based digital gait device from Portables HCT¹, which is clipped to the patient's shoe during standardized walking tasks. Initially, PD progression was modeled using several established outcome measures derived from the UPDRS, fitting a latent time joint mixed-effect model (LTJMM) [229] to estimate each patient's true rate of progression and align disease trajectories on a common disease timescale, thus adjusting for the disease stage at the study baseline. Subsequently, we assessed the association between digital gait features and the estimated progression rates and UPDRS-derived outcomes using linear models. Another analysis investigated the potential of an ML model to forecast individual disease progression rates based solely on digital gait features. Finally, we examined the potential benefits of using digitally predicted UPDRS III scores as a surrogate endpoint in clinical trials compared to the original UPDRS III.

¹<https://www.portables-hct.de/en/product/>

4.2 Methodology

4.2.1 Overview of datasets

In this research, we employed two longitudinal studies, each detailed in the following sections.

LuxPARK The Luxembourg Parkinson Study (LuxPARK) (NCT05266872) [230] is an ongoing observational study that periodically collects data from PD patients at all stages on an annual basis. Data has been gathered for up to four years, covering the period from 2015 to 2022. This study encompasses various data types, including clinical, molecular, and device-based information, with the goal of providing a comprehensive phenotyping of PD patients to facilitate biomarker-driven stratification.

Erlangen Data from the University Medical Center Erlangen, Germany, was collected from PD patients across all disease stages during multiple clinical routine visits, which occurred irregularly between 2011 and 2022. The frequency of follow-ups varies significantly among patients.

4.2.1.1 Clinical data

Outcome measures For both studies, we focused on several outcome scores derived from the MDS-Unified Parkinson’s Disease Rating Scale [86]. This clinical rating scale comprises four components: 1) Non-motor features of daily living, 2) Motor aspects of daily living, 3) Motor examination, and 4) Motor complications. Parts I, II, and IV consist of patient questionnaires, while part 3 is evaluated by a trained clinician. In our study, we utilized MDS-UPDRS Parts I, II, and III as outcomes. Additionally, we employed the Tremor Dominance (TD) score [231], the Postural Instability and Gait Difficulty (PIGD) score [231], and an axial score, all derived from specific sub-items of the MDS-UPDRS. The TD score is calculated by averaging the scores from 11 items: MDS-UPDRS II sub-item 10 (tremor), MDS-UPDRS III sub-items 15a (postural tremor right upper extremity), 15b (postural tremor left upper extremity), 16a (kinetic tremor right upper extremity), 16b (kinetic

tremor left upper extremity), 17a (rest tremor right upper extremity), 17b (rest tremor left upper extremity), 17c (rest tremor right lower extremity), 17d (rest tremor left lower extremity), 17e (rest tremor lip/jaw), and 18 (rest constancy). The PIGD score is determined by averaging the MDS-UPDRS II sub-items 12 (walking and balance) and 13 (freezing), alongside MDS-UPDRS III sub-items 10 (gait), 11 (freezing of gait), and 12 (postural stability). The axial scores are calculated by summing the MDS-UPDRS III sub-items 1 (speech), 2 (facial expression), 9 (arising from chair), 10 (gait), 11 (freezing of gait), 12 (postural instability), and 13 (posture). This customized axial score integrated facial changes and freezing of gait more comprehensively, building on the framework described earlier [232]. While all six outcome scores were available in the LuxPARK cohort, the Erlangen cohort included only UPDRS III. Importantly, in Erlangen, UPDRS III was assessed using the original UPDRS [233] rather than the MDS-UPDRS, with differences outlined in the MDS-UPDRS revision [86]. For simplicity, we collectively refer to both the MDS-UPDRS and original UPDRS as UPDRS. The axial score could also be computed for the Erlangen cohort, with the caveat that sub-item 11 (freezing of gait) was absent from the original UPDRS.

Patient selection In this study, we analyzed data from all patients who had at least two visits for each of the outcomes of interest, resulting in a final cohort of 612 patients from LuxPARK and 264 patients from Erlangen. Table 1 presents the characteristics of these participants, including the number of visits of each participant, disease duration, age, sex, clinical scores, and Hoehn & Yahr (H&Y) stages.

4.2.1.2 Digital gait data

In addition to clinical data, digital gait data collected through wearable sensors was available in both cohorts.

Device The digital gait device utilized in this study is a sensor-based system from Portabiles HealthCare Technology GmbH. This certified class I medical device bears a CE mark in Europe and is currently registered for regulatory approval with the FDA. The device comprises multiple sensors, including a gyroscope and an accelerometer. The specific sensors used vary slightly between the studies. In LuxPARK, the Shimmer 3 sensors from

	LuxPARK	Erlangen
Number of patients	612	264
Number of visits	4.46 ± 1.66	10 ± 8.52
Disease duration, years	4.44 ± 4.77	5.76 ± 4.93
Age, years	66.2 ± 10.47	62.85 ± 10.98
Sex		
Male	413 (67%)	161 (61%)
Female	199 (33%)	103 (39%)
Axial score	6.45 ± 4.62	5.66 ± 4.42
PIGD	0.68 ± 0.71	N/A
TD	0.56 ± 0.45	N/A
UPDRS I	10.34 ± 6.74	N/A
UPDRS II	11.14 ± 7.92	N/A
UPDRS III	33.8 ± 15.45	20.66 ± 12.63
H&Y		
0	2	1
1	63	36
1.5	45	20
2	322	52
2.5	83	25
3	58	47
4	25	25
5	13	3
N/A	1	55

Table 1: Patient characteristics from the LuxPARK and Erlangen cohorts.

This table presents the characteristics of patients from both cohorts. Mean values and standard deviations are reported for all features, while absolute and relative values are provided for sex and H&Y stages. Disease duration, age, clinical scores, and H&Y stages are documented at baseline.

Shimmer Sensing (Dublin, Ireland) were employed. Each unit features a tri-axial accelerometer (range $\pm 8g$) and a tri-axial gyroscope (range ± 1000 deg/sec), with a sampling rate of the device was 102.4 Hz. In the Erlangen study, several sensors were utilized over time: the Shimmer 3 sensors, the Shimmer 2R inertial sensor, and the NilsPod. The Shimmer 2R, also produced by Shimmer Sensing, has a tri-axial accelerometer (range $\pm 6g$) and a tri-axial gyroscope (range ± 500 deg/sec), with a sampling rate of 102.4 Hz. The NilsPod, manufactured by Portabiles HealthCare Technology GmbH (Erlangen, Germany), consists of a tri-axial accelerometer (range $\pm 8g$) and a 3-D gyroscope (range ± 2000 deg/sec), also with a sampling rate of 102.4 Hz.

Gait assessment tasks Digital gait data was collected from PD patients with the device clipped to their shoes while performing various tasks. In LuxPARK, patients completed four tasks:

- **Timed Up and Go (TUG):** The patient begins in a seated position and, upon the clinician’s cue, stands up, walks 3.5 meters, turns around, and walks back to sit down again. This TUG test is a standard assessment in PD examinations for diagnosis, where the time taken to complete the task provides the clinician with insights into the patient’s progression.
- **Turn:** Similar to the TUG, but the patient performs a 360° turn at the midpoint of each 3.5-meter segment in both directions.
- **Manual TUG (Tray):** While executing the turn task, the patient balances a tray with a glass of water, engaging in a dual-task scenario that tests two motor skills (turning and tray).
- **Cognitive TUG (Count):** The patient counts backward while performing the Tray task, introducing a cognitive challenge (counting backward) alongside the motor tasks.

In Erlangen, patients performed three tasks:

- **TUG:** Similar to LuxPARK, but with a 3-meter walking distance in one direction.
- **4x10m Preferred Speed without Stop (4x10m):** The patient walks 10 meters at their preferred speed, turns around, walks back, and repeats this sequence another time without pausing.

- **2x10m Preferred Speed with Stop (2x10m):** The patient walks 10 meters, pauses for 2-3 seconds before turning, and then walks back.

Raw signal processing The collected raw signals undergo processing to derive specific gait analysis parameters. Different algorithms are employed in each cohort. In LuxPARK, a proprietary algorithm from Portabiles HCT, based on the work of Hannink [234], utilizes template matching and a feature extraction module to identify gait-related events, such as heel strikes and toe-offs. It isolates segments linked to each side and estimates stride-by-stride biochemical parameters using a statistical model and a pre-trained convolutional neural network. In Erlangen, dynamic time warping by Barth [235] segments individual strides and detects gait events with the Rampp method [236], with temporal and spatial parameters calculated for each stride, implemented through the gaitmap package [237].

Derived digital gait features The derived features include stride time, gait speed, stride length, and the turning angle. The comprehensive list of these features, along with brief descriptions, is given here:

- **Swing Time*:** Duration from Toe Off (TO) until next Heel Strike (HS)
- **Stance Time*:** Duration from HS with the surface until TO
- **Stride Time*:** Duration of one stride, sum of swing and stance time
- **Stride Length*:** Distance between two consecutive HS, the length of one stride
- **Gait Velocity*:** The average walking speed is calculated by dividing the stride length by the stride time.
- **Max. Lateral Excursion*:** The maximum lateral deviation of the foot in the swing phase, measured from an imaginary line between the foot's position at start and end of the swing phase.
- **Max. Sensor Lift:** The maximum elevation of the heel from the ground during the swing phase.
- **Max. Foot Clearance:** The maximum elevation of the foot from the ground during the swing phase.
- **Max. Toe Clearance:** The maximum elevation of the toe from the ground during the swing phase.
- **Turning Angle:** The angle between the direction of the last swing phase (imaginary line between foot position at the beginning and end of the swing phase) and the orientation of the foot in the next stance phase.
- **Toe Off Angle:** The angle between the heel and the surface at the

beginning of the swing phase.

- Heel Strike Angle: The angle between the toes and the surface when the foot lands.
- Landing Impact: The maximum vertical acceleration during landing of the foot.

All described features were extractable from the LuxPARK dataset, while the Erlangen dataset contained only those marked with an asterisk. In both studies, mean values for each parameter were calculated across the aforementioned tasks.

Overview of available gait data for association testing In LuxPARK, a single digital gait assessment was performed during one clinical visit for a total of 343 patients. In the Erlangen cohort, gait assessments were conducted multiple times across visits, yielding data for 802 patients. The measurements were collected irregularly, with a median interval of 364 days between the first and second visits (mean $\pm$ s.d.: 503.11 ± 488.98 days). For the analysis of associations between digital gait assessments and clinical rating scales, we matched clinical and digital gait data, ensuring that analyses were limited to patients with both longitudinal clinical features and digital measurements. This approach resulted in 161 patients from LuxPARK and 178 patients from the Erlangen cohort being included in the analysis. The characteristics of these patients are detailed in Table 2.

4.2.2 LTJMM model

The latent time joint mixed-effect model (LTJMM) aligns the trajectories of patients along a common latent (i.e., unobserved) disease timescale, as first developed by Li et al. [229]. This model integrates fixed and random effects within a multivariate linear mixed-effects framework, capturing the deviation of each patient from a “mean” reference trajectory over time relative to actual outcomes. It presents a piecewise linear progression of multiple clinical outcomes and estimates the extent to which each patient’s timescale deviates from that of the reference. The model is defined as follows:

$$y_{ijk} = x_i\beta_k + \gamma_k(t_{ijk} + \delta_i) + \alpha_{0ik} + \alpha_{1ik}t_{ijk} + \epsilon_{ijk}$$

	LuxPARK	Erlangen
Number of patients	161	178
Number of visits	5.32 ± 1.53	12.08 ± 9.12
Disease duration, years	4.92 ± 5.29	5.31 ± 4.57
Age, years	64.71 ± 10.08	62.39 ± 10.37
Sex		
Male	117 (73%)	110 (62%)
Female	44 (27%)	68 (38%)
Axial score	5.71 ± 3.79	5.35 ± 3.93
PIGD	0.56 ± 0.58	N/A
TD	0.5 ± 0.41	N/A
UPDRS I	9.28 ± 5.68	N/A
UPDRS II	10.32 ± 6.8	N/A
UPDRS III	29.89 ± 14.2	20.2 ± 11.89
H&Y		
0	1	1
1	19	25
1.5	11	14
2	93	35
2.5	20	18
3	12	35
4	4	11
5	1	2
N/A	-	37

Table 2: Patient characteristics from the LuxPARK and Erlangen cohorts for association testing. This table presents the characteristics of patients from both cohorts utilized in the association testing between clinical outcomes and digital gait features. Mean values and standard deviations are reported for all features, while absolute and relative values are provided for sex and H&Y stages. Disease duration, age, clinical scores, and H&Y stages are documented at baseline.

where y_{ijk} is the outcome feature k observed at time point j for individual i . The covariates x_i have corresponding coefficients β_k , representing the fixed effects shared across all individuals. The parameter γ_k indicates the mean slope for outcome feature k , establishing the mean trajectory from which each patient’s deviation is assessed. The time variable t_{ijk} is adjusted by the individual-specific time shift δ_i . Random effects are defined by α_{0ik} and α_{1ik} , representing a random intercept and slope for each outcome and individual, respectively, following a multivariate normal distribution with a mean of zero. Finally, ϵ_{ijk} denotes the measurement error, drawn from a normal distribution with a mean of zero. The inferred patient-specific time shifts δ_i provide estimates of how advanced a particular patient is in terms of symptom severity compared to other patients in the same cohort. Additionally, the model estimates patient-specific slopes of symptom progression, represented by the random slope α_{1ik} , where a higher value suggests a more rapid progression of symptoms compared to peers. The LTJMM is implemented in the `ltjmm` R package [238], which leveraged `rstan` [239] to apply Markov Chain Monte Carlo (MCMC) algorithm for model parameter estimation.

4.2.2.1 Model definition

In this study, the time scale corresponds to the reported time since the patient’s PD diagnosis. The multivariate outcomes include UPDRS I-III, TD, PIGD, and axial scores for LuxPARK cohort, while the Erlangen cohort includes UPDRS III and axial scores. Fixed effects are represented by covariates x_i , which account for age at diagnosis, sex, and medication status (ON/OFF) to control for their influence. The model was estimated using all available longitudinal clinical data, as described in section 4.2.1.1 and Table 1. The MCMC algorithm was executed with 4 chains, 25000 iterations, and 12500 warm-up steps. For subsequent analysis, we utilized the latent time defined as the sum of the original time point t_{ijk} and the estimated patient-specific time shifts δ_i , along with the random slope α_{1ik} .

4.2.2.2 Model validation

To validate the LTJMM, we assessed its accuracy by calculating the correlation between real and predicted outcomes at the last observed visit. This involved fitting the LTJMM to all outcomes while excluding the last measurements for each patient, with the predicted value subsequently derived from the fitted

model. Additionally, we examined the latent time of each patient concerning their H&Y stages. We hypothesized that the time shift of each individual’s trajectory should align with their H&Y stages, whereby patients with lower H&Y stages, indicating earlier disease progression, would be shifted further left on the common disease time scale than those with higher H&Y stages, reflecting more advanced disease. Therefore, we further validated the LTJMM model by analyzing the distribution of latent time across H&Y stages.

4.2.3 Evaluating the relationship between digital gait patterns and clinical outcomes

The relationship between the derived digital gait features and various outcomes measured during the same visit was modeled using linear models. In Erlangen, linear mixed-effects models were employed due to the longitudinal nature of the digital gait data. We utilized the latent disease time and progression rates (random slope) derived from the LTJMM model as model outcomes, alongside the clinical outcome scores. Covariates included age at diagnosis and sex for modeling latent disease time, while the latent disease time at the time of clinical assessment was used otherwise. This approach accounted for temporal variations in observed patient trajectories. In case of linear mixed-effects models for the Erlangen cohort, a patient-specific random intercept was incorporated. We evaluated multiple models with varying feature sets:

- all digital gait features (allGait) - Utilizes all available digital gait features across all tasks.
- individual task features - Employs subsets of digital gait features derived from each specific task.
- single digital gait features (singleGait) - Analyzes each digital gait feature individually, incorporating one feature covariates.

All models were compared against a null model containing only adjustment variables (age and sex or latent disease time) using a likelihood ratio test. The resulting p-values were adjusted for multiple testing using the Benjamini & Yekutieli method [240].

4.2.4 ML models for forecasting disease progression

To predict the random slope from digital gait features, various ML models were employed, including Random Forest [241], XGBoost [227], and Lasso regression [242]. Due to the availability of only one digital gait assessment for LuxPARK, a single prediction was made. Conversely, both the first and first plus second digital gait visits were utilized in Erlangen. The feature sets tested included a null model (using only sex and age as predictors), all digital gait features (allGait), and task-specific subsets. The implementation involved the scikit-learn [243] and xgboost [227] Python packages. Hyperparameter optimization was performed through randomized search within a 5-fold inner cross-validation (CV) framework, as detailed in Table 3. Performance metrics were evaluated using a 5-fold outer CV setting, and R^2 values of the models were compared against the null model using the Wilcoxon signed-rank sum test.

Algorithm	Parameter	Grid values
Random Forest	Max. tree depth	[10, 15, 20, 25, 30, 35, 40, 45, 50]
	Min. samples per split	[0.05, 0.1, 0.15, 0.2, 0.25]
	Min. samples leaf node	[1, 2, 4, 8, 10, 16, 20]
	Cost-complexity pruning	[0.001, 0.01, 0.1]
XGBoost	Max. tree depth	[3, 4, 5]
	Min. child weight	[1, 5, 10]
	Gamma	[0.5, 1, 1.5, 2.5]
	Subsample	[0.6, 0.7, 0.8, 0.9, 1]
	Colsample by tree	[0.6, 0.7, 0.8, 0.9, 1]
Lasso	Alpha	[0.0001, 0.0001, 0.001, 0.01, 0.1, 0, 1, 10, 100]

Table 3: Hyperparameter space used for ML models.

4.2.5 Sample size calculation for digital gait as a surrogate endpoint in clinical trials

A simulated randomized controlled trial was conducted over a one-year observation period, with assumed visits scheduled every 60 (bi-monthly), 30 (monthly), or 7 (weekly) days. The sample size required to assess a 30% efficacy at 80% power and a significance level of 0.01 for a potentially disease-modifying drug was calculated. The primary study endpoints included the original UPDRS III and a gait-predicted UPDRS III, the latter estimated using digital gait features from the 4x10m test in Erlangen through a linear mixed-effects model. The presumed treatment effect was inspired by an ongoing trial [244]. Treatment and control groups were equal in size, with no variations in treatment dosage. Power and sample sizes were calculated using linear mixed-effects models based on a method from Edland [245], implemented in the longpower R package [246].

4.3 Results

4.3.1 LTJMM model

The original longitudinal trajectories were aligned on a common disease time scale utilizing the LTJMM model, as previously described. The results of this alignment for LuxPARK and Erlangen are illustrated in Figures 2 and 3, separately for each of the modeled outcomes.

4.3.1.1 Model validation

To validate the findings from the LTJMM models, we plotted the distribution of the latent (shifted) time for each H&Y stage. The LTJMM model is expected to accurately order the patient trajectories along these stages, with patients in more advanced disease stages (higher H&Y stages) anticipated to shift further to the right, indicating later positioning in the common disease timeline. As shown in Figure 4 and Table 4, the model successfully orders the patient trajectories following the H&Y stages.

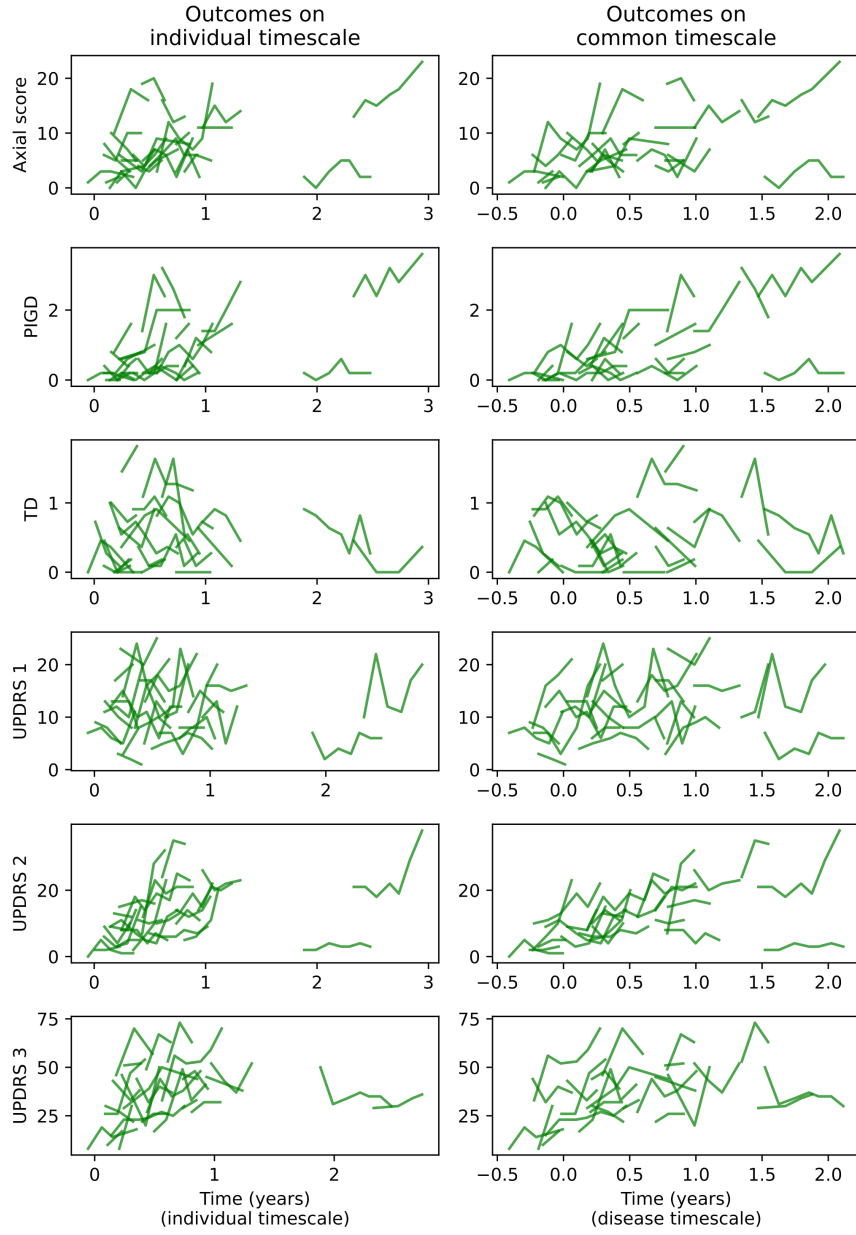


Figure 2: Longitudinal trajectories in LuxPARK after LTJMM alignment. The figure displays the original (left) and shifted (right) trajectories for six outcomes: Axial score, PIGD, TD, UPDRS I, UPDRS II, and UPDRS III. The time dimension is represented in years using individual timescales for the original trajectories and a common disease timescale for the shifted trajectories.

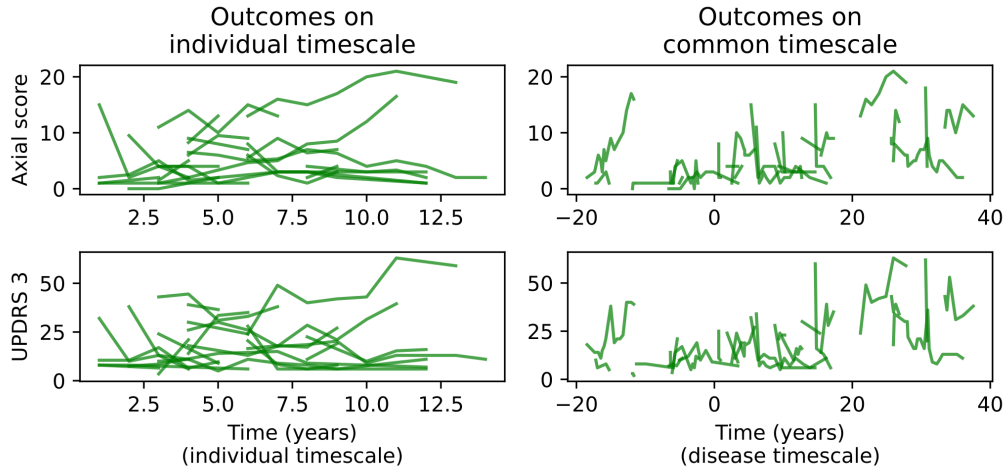


Figure 3: Longitudinal trajectories in Erlangen after LTJMM alignment. The figure displays the original (left) and shifted (right) trajectories for two outcomes: Axial score and UPDRS III. The time dimension is represented in years using individual timescales for the original trajectories and a common disease timescale for the shifted trajectories.

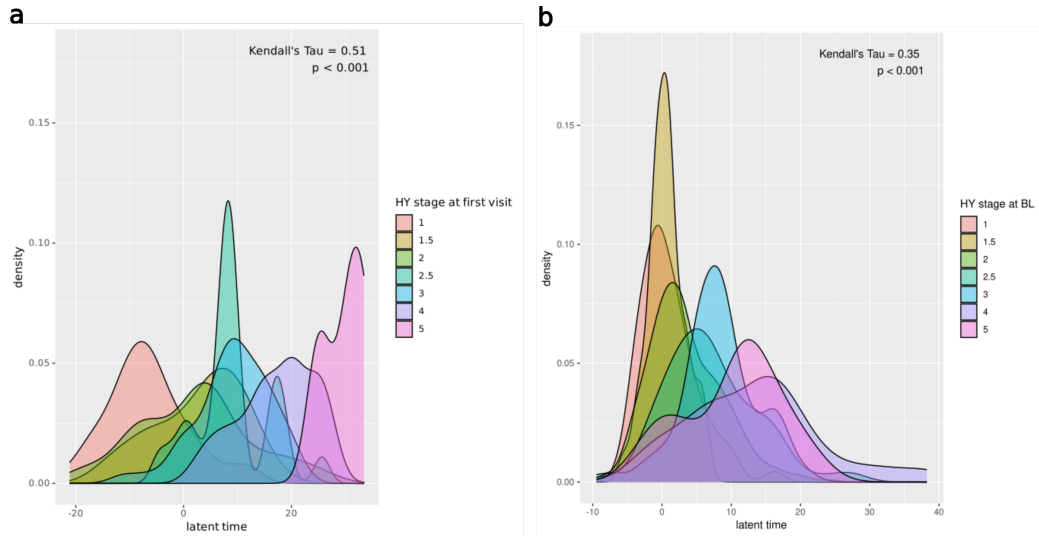


Figure 4: Distribution of latent time across H&Y stages. This figure illustrates the distribution across multiple H&Y stages for the a) Erlangen cohort and b) LuxPARK cohort, demonstrating that the models effectively order patient trajectories according to disease stages. The correlation between latent time and H&Y stages is presented using Kendall's tau and corresponding p-values.

H&Y stage	Erlangen		LuxPARK	
	Mean	Variance	Mean	Variance
1	-6.64	55.43	1.03	17.30
1.5	2.86	66.10	0.64	6.90
2	2.09	109.59	3.38	28.95
2.5	8.96	48.20	7.03	43.65
3	9.25	47.14	8.30	28.32
4	17.25	48.80	13.34	94.52
5	29.74	17.67	10.13	49.52

Table 4: Mean and variance of latent time across multiple H&Y stages. This table presents the mean and variance of latent time for each H&Y stage in both the Erlangen and LuxPARK cohorts.

Furthermore, we calculated the Pearson correlations between actual clinical outcomes and those predicted by the LTJMM, yielding values of $\rho = 0.73$ for Erlangen and $\rho = 0.42$ for LuxPARK. These results indicate a satisfactory fit of the LTJMM models.

4.3.2 Evaluating the relationship between digital gait patterns and clinical outcomes

Statistical associations between clinical outcomes and features derived from the Portabiles digital gait device were analyzed using linear models, following the alignment of observed clinical outcome trajectories on a common disease timescale. Models were fitted for multiple outcomes with varying feature sets. We examined the relationship between digital gait data and symptom severity, indicated by clinical scores, progression, reflected by the random slope of the LTJMM model, and disease stage, defined by latent time. Both task-specific models (Count, Tray, TUG, Turn) and a comprehensive model incorporating all available digital gait features were compared against a null model that included only the integrated covariates. The adjusted p-values from the likelihood ratio tests are presented in Tables 5 and 6 for LuxPARK and Erlangen.

Outcome	All Gait	Count	Tray	TUG	Turn
Clinical score (symptom severity)					
Axial score	0.0055	0.0006	0.0006	0.0001	0.0175
PIGD	0.2152	0.0068	0.0068	<0.0001	<i>0.0629</i>
TD	0.0178	0.4882	0.4882	1	1
UPDRS I	0.2717	<i>0.0783</i>	0.2384	0.2504	<i>0.0783</i>
UPDRS II	0.4154	0.1427	0.1427	0.1427	0.3601
UPDRS III	0.0337	0.2149	<i>0.0538</i>	0.0476	0.1058
Slope (progression)					
Axial score	0.2765	0.5264	<i>0.0534</i>	<i>0.0990</i>	0.4970
PIGD	0.6788	1	1	1	1
TD	0.4895	1	0.1709	0.1709	0.6128
UPDRS I	0.3035	0.8514	<i>0.0621</i>	<i>0.0621</i>	0.5288
UPDRS II	0.3864	0.9143	0.1260	<i>0.0686</i>	0.4988
UPDRS III	0.6779	1	1	1	1
Latent time (disease stage)					
Latent time	0.0001	0.0288	<i>0.0726</i>	0.0005	0.0023

Table 5: Associations of digital gait features with symptom severity, symptom progression, and disease stage in the LuxPARK study. This table presents adjusted p-values from the likelihood ratio test, assessing the relationship between various digital gait feature sets (including all measured digital gait features (All Gait) and task-specific subsets (Count, Tray, TUG, Turn)) and clinical outcomes. Significant results (adjusted $p < 0.05$) are highlighted in bold, while weakly significant results (adjusted $p < 0.1$) are italicized. The measured outcomes include clinical scores (top), their progression (middle), and disease stage (bottom).

The main results are detailed in the following paragraphs, organized by outcome: symptom severity (clinical scores), disease progression (slope), and disease stage (latent time).

Outcome	All Gait	TUG	2x10m	4x10m
Clinical score (symptom severity)				
UPDRS III	0.1574	0.0361	0.0203	0.0110
Axial score	0.0010	0.0068	0.0068	0.0068
Slope (progression)				
UPDRS III	1	1	1	1
Axial score	1	1	1	<0.0001
Latent time (disease stage)				
Latent time	0.0047	<0.0001	0.0001	<0.0001

Table 6: Associations of digital gait features with symptom severity, symptom progression, and disease stage in the Erlangen study. This table presents adjusted p-values from the likelihood ratio test, assessing the relationship between various digital gait feature sets (including all measured digital gait features (allGait) and task-specific subsets (TUG, 2x10m, 4x10m)) and clinical outcomes. Significant results (adjusted $p < 0.05$) are highlighted in bold, while weakly significant results (adjusted $p < 0.1$) are italicized. The measured outcomes include clinical scores (top), their progression (middle), and disease stage (bottom).

4.3.2.1 Monitor motor symptom severity

Our investigation into the relationship between digital gait features and motor symptom severity yielded significant findings, particularly regarding the axial score in both datasets (cf. Tables 5 and 6). In the LuxPARK dataset, significant correlations were also observed for UPDRS III, PIGD, and TD scores. Task-specific analyses indicated distinct task specificity within the LuxPARK cohort, with TUG test emerging as the most informative. The axial score exhibited the strongest correlation with digital gait features, revealing significant effect sizes in the models for stance time ($p=0.04$, Count), swing time ($p=0.009$, Tray), gait speed ($p=0.029$, Tray), stride length ($p=0.015$, Tray), and landing impact ($p=0.047$, TUG). Notably, the landing impact from the TUG test also demonstrated a significant effect size when assessing the PIGD score ($p=0.048$). For this score, gait speed ($p=0.014$, Count) and stride length ($p=0.010$, Count) also exhibited significant effect sizes. In the Erlangen cohort, all three tasks (2x10m, 4x10m, and TUG) revealed significant associations between digital gait features and disease severity,

however, swing time in the 4x10m task for axial score was the only measure to show a significant effect size ($p=0.045$) within the model.

4.3.2.2 Monitor symptom progression

Subsequently, we explored the correlation between digital gait features and disease progression, represented by the random slopes of the LTJMM. Overall, significant associations between digital gait features and the progression of the axial score were identified solely in the Erlangen cohort for the 4x10m test. Additionally, weak significance was observed in the LuxPARK cohort for the axial score in both the Tray and TUG tests. Moreover, weak significance was noted for UPDRS I in the Tray and TUG tests, and UPDRS II in the TUG test.

4.3.2.3 Monitor disease stage

Lastly, we investigated the association between the disease stage (represented through the latent time from LTJMM) and the digital gait features. We could observe a significant association both in LuxPARK and Erlangen in general and task-specific associations for Count test in LuxPARK and all tasks in Erlangen. Significant effect sizes could be observed for maximal sensor lift in LuxPARK for TUG ($p=0.031$) and Turn ($p=0.048$) test, as well as, for stride time in Erlangen and the 2x10m ($p=0.001$) and TUG ($p=0.045$) test.

Altogether, significant associations of digital gait features were thus identifiable with traditional clinical motor scores, their slopes, and disease stage.

4.3.3 Forecasting disease progression with ML models

We developed various ML models to predict the patient-specific slopes, which represent disease progression in motor symptom-related clinical outcomes. The most compelling results were achieved using a 10-fold CV setup with the Random Forest algorithm. Figure 5 displays these results from the Erlangen cohort, where predictions were made using both the first gait assessment alone (left) and the combined data from the first and second assessments (right). In the first scenario, a significant enhancement in prediction performance

was observed when utilizing data from the 4x10m test, compared to a model that relied solely on age and sex as predictors. Predictions were further refined with the inclusion of additional data from the second gait assessment, particularly when all gait tasks were considered. Notably, the prediction of UPDRS III scores demonstrated a substantial improvement when digital gait features were incorporated, rather than depending solely on age and sex. In contrast, due to the availability of only a single gait assessment in the LuxPARK cohort, we were unable to achieve comparable improvements in prediction.

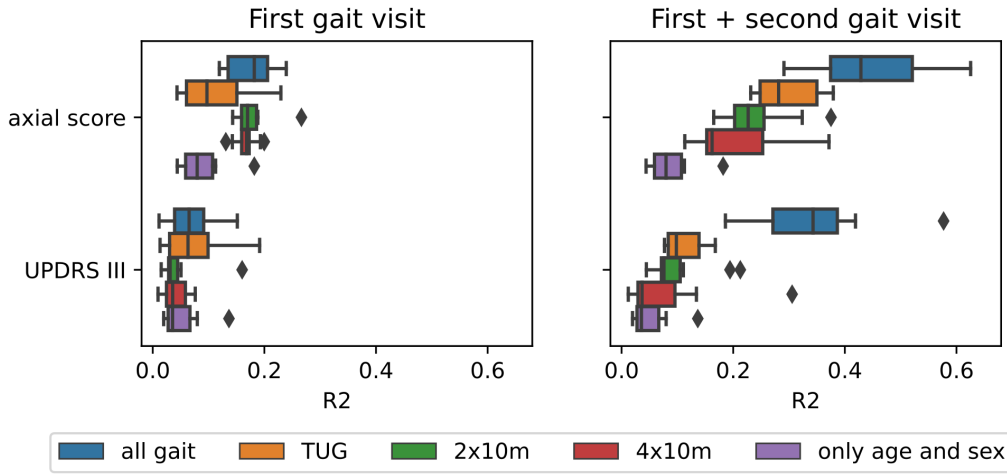


Figure 5: Prediction of patient-specific slopes for traditional clinical outcomes based on digital gait features. Boxplots display the squared coefficient of determination (R^2) from multiple CV repeats of a Random Forest machine learning model. The model predicts the (random) slopes of the axial and UPDRS III scores for each patient, trained on data from either the first gait assessment only (left) or both the first and second gait assessments (right). Different colors represent the feature sets utilized: all digital gait features plus age and sex (blue), TUG task plus age and sex (orange), 2x10m task plus age and sex (green), 4x10m task plus age and sex (red), and only age and sex (purple).

4.3.4 Sample size calculation for digital gait as surrogate endpoint in clinical trials

A longitudinal clinical trial with varying visit frequencies was simulated to assess the potential advantages of digital gait features compared to traditional clinical outcome scores. It is important to note that digital gait features can be

collected more easily and at lower cost with higher frequencies than traditional clinical scores. Furthermore, data collection could potentially take place in a patient's home environment. For the trial simulation, UPDRS III scores were predicted from gait data, utilizing digital gait features obtained from the 4x10m test via a linear mixed-effects model. The estimates demonstrated a strong correlation ($\rho = 0.77$) with the original data (see Figure 6).

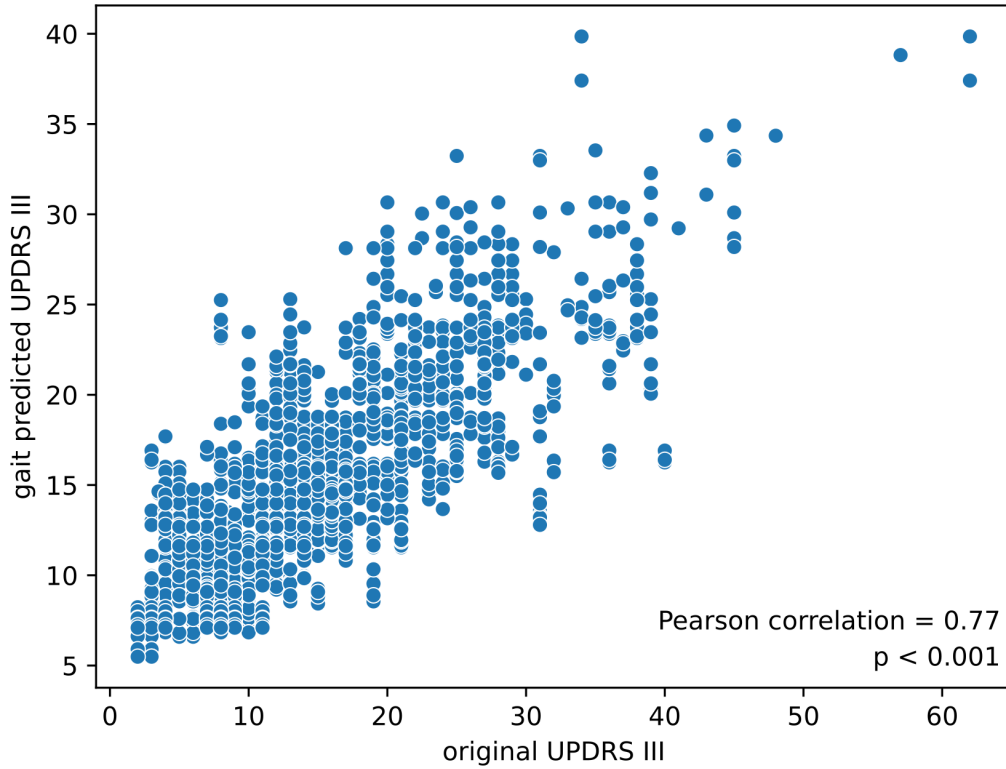


Figure 6: Scatterplot of original versus gait-predicted UPDRS III scores. This scatterplot illustrates the relationship between the original UPDRS III scores and those predicted by gait features, yielding a Pearson correlation of 0.77.

Our statistical sample size calculation, which assumed a treatment effect of 30%, aimed for a statistical power of 80% at a significance level of 1%. Simulating bi-monthly assessments of UPDRS III indicated a required sample size of 690 patients for both the original and gait-predicted UPDRS III (see Figure 7). A monthly assessment could reduce the required sample size to 550 patients using the gait-predicted UPDRS III, while weekly assessments could further decrease this number to 380 patients, representing a potential reduction of 44%. It is crucial to emphasize that conducting traditional UPDRS III at such frequent intervals in a clinical setting would be exceedingly challenging.

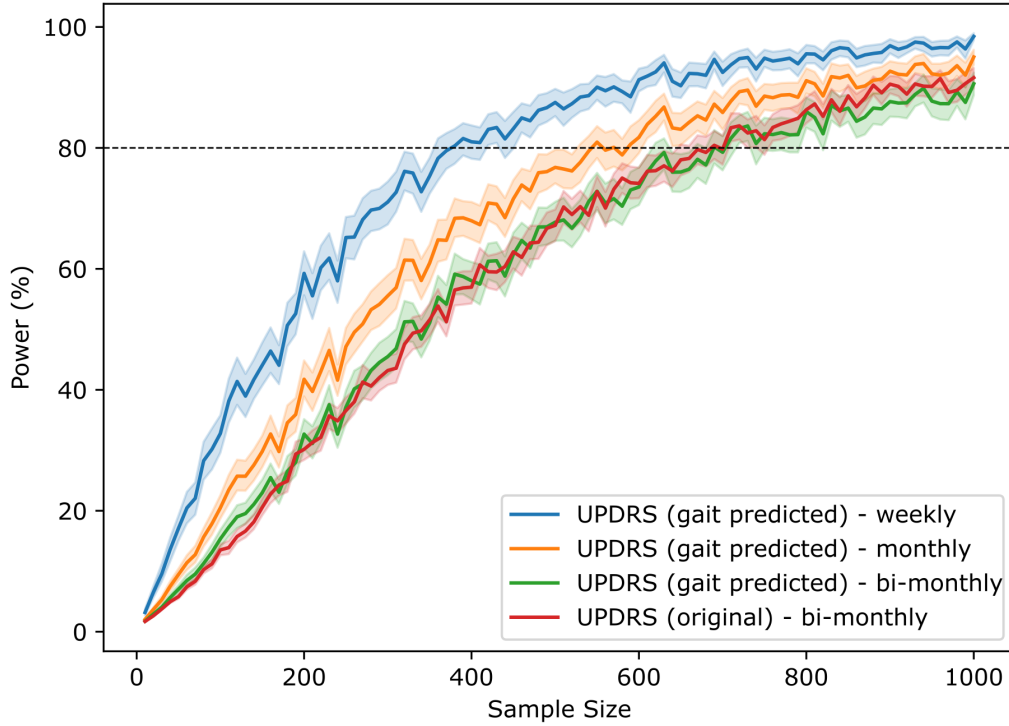


Figure 7: Simulation of a randomized clinical trial with varying visit frequencies. The curves illustrate the statistical power achievable in a clinical study with a specific sample size, anticipating a treatment effect of 30% over a one-year observation period and a significance level of $\alpha = 0.01$. The dashed line represents 80% power. Simulations were conducted for randomized clinical trials using either original or gait-predicted UPDRS III as outcomes, with weekly (blue), monthly (orange), and bi-monthly (green) visit frequencies for gait-predicted UPDRS III, while only bi-monthly visits were simulated for original UPDRS III (red).

4.4 Discussion and Conclusion

4.4.1 Digital Gait Evaluations for monitoring symptoms and progression

Digital gait features from the Portables digital gait device have demonstrated a significant correlation with disease stage, severity as assessed by UPDRS-derived scores, and disease progression. Notably, the axial score

exhibits considerable potential for monitoring disease severity and progression. Previous studies have established associations between axial abnormalities and patient gait, however, a quantitative analysis linking these symptoms to gait characteristics was previously lacking. Most studies analyzing digital gait data tend to focus on micro-analysis, examining single digital features in isolation rather than conducting a comprehensive analysis of multiple digital gait features across various tasks and studies. Additionally, many studies utilize only one or no standardized tasks. This work provides a clear advantage by performing a thorough analysis of both individual and combined features across multiple levels and standardized tasks. The examination of task-specificity offers a systematic overview of the most effective tasks for digital gait analysis. Significant task-specific differences were observed, for instance, the traditional TUG test consistently proved effective for monitoring motor symptom severity across both studies, being the only task to correlate with UPDRS III in the LuxPARK cohort. Furthermore, digital gait features derived from the Tray and Count tasks were linked to the PIGD score, suggesting that dual tasks, which combine multiple motor and cognitive exercises, are effective in monitoring disease severity related to postural instability. Together with previous observations that dual tasks can trigger freezing of gait, this indicates the need for further research into the capacity to monitor cognitive impairment in PD patients. Moreover, a strong association between digital gait features and the progression of UPDRS-derived motor symptoms was established in the Erlangen cohort, where gait measurements were collected over multiple visits and an extended period. However, this association was only observed for the 4x10m task, which involved longer-distance gait measurements compared to other tasks. This underscores the importance of longer-distance walks in identifying changes in motor symptoms over time in individual patients.

4.4.2 Disease progression forecasting with ML methods

A subsequent investigation explored the feasibility of predicting UPDRS III scores from digital gait features. The results indicated a strong correlation with the original data ($\rho=0.88$). This is noteworthy since the prediction task did not incorporate other symptomatic information evaluated in UPDRS III, such as hypomimia, dysarthria, or tremor. This may suggest a potential correlation between the severity of these additional symptoms and digital gait data, or indicate their relatively stable levels that do not influence disease progression as measured by UPDRS III. Further research is needed

to elucidate the relationship between digital gait and other PD symptoms. While previous studies have shown a correlation between digital gait features and various disease stages, they often did not quantify disease progression, typically classifying patients into disease stages or distinguishing PD patients from healthy controls. In contrast, this analysis focused on the quantitative association of digitally assessed gait features with motor symptom progression. However, we observed limited accuracy in the ML-based prediction. Despite analyzing two subsequent clinical visits, the cross-validated R^2 values for predicting axial score progression was only 0.45, and even lower for UPDRS III progression ($R^2 = 0.35$), indicating limited prognostic value for the digital data while affirming its utility in objectively monitoring motor symptom and their progression.

4.4.3 Potential Value of digital gait as a surrogate endpoint in clinical trials

Finally, the potential of digital gait features for clinical trials was evaluated through the simulation of a randomized control trial. The findings revealed that predicting UPDRS III using digital gait data requires a similar number of patients as the standard UPDRS III, when provided data is collected at the same frequency. A weekly frequency of data collection resulted in a substantial reduction in the required sample size. This highlights the significant potential of digital gait sensors in clinical studies, as measurements can be taken remotely and more frequently than traditional in-hospital assessments. This approach may enhance patient quality of life by allowing clinical evaluations to occur at home, potentially increasing willingness to participate in clinical studies by reducing the necessity for clinical visits. Furthermore, data collection via tasks like the 4x10m test is notably quicker and easier than conducting a full UPDRS examination. Despite this, few studies have explored the benefits of utilizing digital gait data, with most citing subjective potential endpoints without validating these hypotheses. Only one study has demonstrated the potential of sensor-based digital gait analysis to address the limitations of UPDRS-based gait assessments in a phase II trial [193]. This work provides an objective comparison of original and gait-predicted UPDRS III outcomes for the first time, quantifying the potential of sensor-based digital gait data as an endpoint in clinical trials concerning the required sample size.

4.4.4 Summary

In summary, this work illustrates the capability of the Portables HCT digital gait device to objectively monitor the severity and progression of motor symptoms. It highlights the potential advantages of employing digital gait metrics as outcome measures in clinical trial contexts, thereby contributing to the growing body of literature that underscores the benefits of digital solutions in the field of PD. A distinguishing feature of this study is its comprehensive quantitative analysis of disease symptoms and stages, in addition to an analysis of symptom progression using a digital gait device and the exploration of its quantifiable benefits for clinical studies. However, it is essential to note that the use of any medical device outside pure research settings in routine medical care requires regulatory agency approval.

5 Conclusion

Despite extensive research aimed at identifying disease-modifying treatments for neurodegenerative diseases such as AD, PD, and HD, only a limited number of effective disease-modifying therapies have been discovered thus far [2–4]. The heterogeneity of these diseases and the challenge of identifying suitable therapeutic targets pose significant obstacles. This thesis contributes to the development of personalized medical treatments and enhances disease management through data-driven approaches designed to address these challenges. Firstly, we developed a patient data-driven multi-scale quantitative model of AD, enabling us to simulate knock-down experiments (Chapter 2). This model facilitated the investigation of the phenotypic effects of targeting CD33, which we validated using cell line data. While CD33 was the primary focus, the model also predicts perturbation effects of other genes and mechanisms, allowing for the prioritization of drug targets based on their potential impact on disease phenotype. Next, we explored the heterogeneity in HD symptom progression by identifying, characterizing, and creating predictive models for two distinct progression subtypes, revealing significant differences in cognitive performance (Chapter 3). This work underscores the importance of non-motor symptoms in predicting disease progression and personalizing treatment. Objective assessments of disease progression enhance patients’ quality of life, while our predictive approach enables tailored counseling and treatment. Targeted prevention of cognitive decline is crucial, and personalized medical services facilitate appropriate therapies and ongoing adjustments to care. Finally, we evaluated the potential of gait sensor measurements for monitoring PD symptoms and their progression (Chapter 4). We also assessed the feasibility of using sensor-based digital gait data as an endpoint in clinical trials by analyzing the required sample sizes. This work highlights the effectiveness of digital gait data for enhanced objective monitoring of PD, enabling more personalized treatment adjustments through continuous assessment. Additionally, home-based evaluations improve patient convenience, ultimately enhancing quality of life. Moreover, objective monitoring empowers clinicians to make data-driven decisions regarding treatment efficacy and adjustments, facilitating informed decision-making. Utilizing surrogate endpoints derived from digital gait data presents a viable option that reduces sample sizes and lowers clinical trial costs while increasing efficiency by expediting trial processes.

The publication presented in Chapter 2 primarily focuses on target identification for drug development. We developed a comprehensive, data-driven multi-scale quantitative model of AD, integrating gene expression, clinical data, and qualitative knowledge of cause-and-effect relationships into a quantitative system medical model of AD. This innovative approach enhances our understanding of intervention effects within a multi-scale biological framework of AD. By simulating the downregulation of the CD33 gene, we predicted a significant impact on cognition and brain pathophysiology via the prostaglandin pathway, with results closely aligning with experimental gene expression knockout data. While CD33 serves as a specific example, our methodology is applicable to other potential drug targets, providing a robust computational framework for predicting the downstream effects of interventions on AD biology across multiple scales: gene expression, disease pathology, and phenotype. Moreover, this patient-driven quantitative modeling approach is broadly applicable in systems medicine, particularly in contexts where the mechanistic understanding of the underlying biology is incomplete. While a quantitative, mechanistic model of a disease offers valuable insights, effective personalized disease management requires an understanding which patients will benefit most from new treatments.

In Chapter 3, we explored the heterogeneity in HD disease progression, identifying two distinct subtypes: a larger group with relatively stable disease trajectories and a smaller cluster exhibiting more progressive patterns. Characterization of these subtypes revealed significant differences in cognitive impairments across multiple cognitive scales, particularly in frontal-executive tests that effectively distinguish the two groups. Our findings underscore the critical role of cognition in HD. Although cognitive deficits are commonly observed in HD patients and their correlation with disease progression has been previously established [247], cognitive assessments have not historically influenced the diagnosis, prevention, or prediction of HD symptoms and progression. Our work highlights the necessity of incorporating cognitive test results into patient counseling and treatment strategies as these are often more indicative of disease progression than motor tests. This research is the first to investigate HD progression from this perspective. Additionally, we utilized both motor and non-motor symptoms, including behavioral or psychological factors, for classification purposes. With the developed model, each patient can be individually assessed based on their unique profile, allowing for a clear and objectively measured prognosis. This personalized projection of symptom development through an AI model provides clinicians with insights to optimize counseling and treatment, guiding interventions toward personalized therapies. Furthermore, this empowers patients to organize and plan their

lives according to their prognosis, ultimately enhancing their quality of life and improving overall disease management.

Finally, in Chapter 4, we explored the use of digital gait assessment for monitoring symptoms and progression in PD. This approach facilitates objective tracking of disease status and progression, enabling clinicians to evaluate the effects of new treatments. Our comprehensive analysis of individual and combined features derived from digital gait sensors revealed significant associations with disease stage and severity. Notably, the correlation between digital gait features and PD symptoms was systematically examined across multiple standardized tasks and two independent cohorts for the first time. We identified notable task-specific differences; for instance, the traditional TUG test consistently proved effective for monitoring motor symptom severity in both studies. Additionally, dual tasks that combined motor and cognitive exercises, such as the Tray and Count task, effectively assessed disease severity related to postural instability. However, significant associations between digital gait features and symptom progression were observed only in Erlangen, where gait assessments were conducted over an extended period and repeated in subsequent visits. This underscores the importance of longer-distance walks in detecting changes in patient’s motor symptoms. While digital gait features demonstrate potential for objectively monitoring motor symptoms and their progression, their prognostic value in ML-based predictions of clinical score progression remains limited. Nonetheless, in another experiment evaluating the utility of digital gait assessment for clinical trials, we found that estimates of the UPDRS III correlated well with the original data when predicted from gait assessments conducted during the same visit. This indicates that such predictions are feasible, even though gait data alone does not capture other symptoms. It remains unclear whether this is due to correlations between other symptoms and digital gait or if those symptoms remain stable in terms of UPDRS III progression. Through sample size estimations for randomized clinical trials (RCTs), we demonstrated that the number of participants required could be significantly reduced by digitally assessing the UPDRS III through weekly gait assessments. The potential advantages of using digital gait sensors for clinical trial evaluation are threefold:

1. Data collection for a single gait task is simpler than conducting a complete UPDRS III assessment.
2. Measurements with digital devices can potentially be taken remotely in patients’ homes, increasing the frequency of assessments without necessitating clinical visits.
3. Utilizing digital gait sensors may alleviate the burden on patients partic-

ipating in RCTs, as they can complete tests at home, thereby enhancing their quality of life and increasing their willingness to participate in studies.

5.1 Perspective future work

Despite our efforts across the various projects outlined in this dissertation, several aspects remain to be improved and explored more thoroughly in the future. As discussed, understanding the heterogeneity of neurodegenerative diseases, particularly regarding symptoms and progression throughout the disease management pipeline, is crucial for developing effective disease-modifying treatments. However, given the significant background of heterogeneity and the limitations of our current studies, numerous follow-up steps can be pursued.

One initial step to consider within the systemic approach presented in Chapter 2 is the integration of additional data modalities into the model’s graph structure. Incorporating known side effects of approved drugs, imaging data, or proteomic data can provide valuable insights into potential outcomes. Specifically, by including side effects, we can not only predict possible phenotypic impacts on clinical scales and cognition but also estimate expected adverse effects. This enhancement contributes to a more comprehensive risk-to-benefit analysis for new drugs in development. Another promising area for future research involves employing more advanced graph ML algorithms, such as Graph Neural Networks [248], which have shown success in predicting adverse drug events and drug-protein interactions in prior studies [249, 250].

Identifying and characterizing progression subtypes of HD is a crucial step toward developing personalized treatment strategies. Further characterization of the two identified subtypes from Chapter 3 using additional data modalities can enhance this initiative. Exploring correlations between these subtypes and biological data such as magnetic resonance imaging, CSF biomarkers, or genetic information, could yield relevant clinical insights and identify potential targets for new treatments. Additionally, the clustering approach itself could be refined by incorporating neurobehavioral or psychiatric scores, which are currently underrepresented. However, expanding the clinical scores used in the initial clustering necessitates the development of more stable and easily estimable non-linear mixed models capable of aligning multivariate

trajectories. The current method is time-consuming and relies on manual adjustments to estimate the final model parameters. One viable option, is the latent-time joint mixed model, previously utilized in Chapter 4 for aligning PD trajectories and in prior work [158] before the progression subgroup clustering. Another alternative is the use of a generalized neural network mixed model, employing the output of a feed-forward neural network instead of the conventional fixed effects [251]. Another area for future work involves the handling of irregular time series. Currently, VaDER can only cluster regular time series, meaning that it is limited to structured clinical study data with defined clinical visits. To extend this approach to irregular time series typically seen in outpatient care, VaDER must be adapted to accommodate such variations. Finally, since subtype identification is the first step in determining optimal treatments, these subtypes could be integrated into clinical studies. Enriching trial populations with individuals exhibiting rapid progression may expedite the demonstration of disease-modifying effects of novel compounds, as this group is more likely to show decline within the typical trial periods of 1 to 2 years. Simulations of this strategy with PD progression subtypes indicated a faster identification of significant outcome changes with fewer participants [158].

As discussed in Chapter 4, features derived from digital gait sensors in PD patients are generally effective for monitoring symptoms and their progression. Notably, these features also predict the UPDRS III during the same visit, despite focusing solely on gait. The underlying reasons for this correlation remain unclear, warranting further investigation into the relationship between digital gait features and non-gait symptoms, particularly cognitive impairments of PD patients. Our findings additionally indicate that longer and repeated gait assessments yield higher correlations with motor symptoms and are the only measurements that can partially predict patient progression. Collecting more data through additional repetitions and longer distances may enhance the prognostic value and provide clearer insights. One shortcoming of our current approach is the reliance on summary statistics from gait features averaged across all steps in a single exercise. Exploring step-wise data could be beneficial, as it may capture information lost when averaging features over all steps performed within one task. We could utilize step-wise gait features, such as gait speed and step length, or even raw sensor data from gyroscopes and accelerometers. Methodologically, convolutional neural networks [252], long short-term memory networks [253], or time-series transformers [254] could be employed, as these methods have already shown success in gait recognition [255–261] and patient classification based on gait [262–265]. However, the use of any medical device outside of pure research

settings in routine clinical care requires regulatory approval and thorough preparation. Moreover, it will be essential to observe the implementation of the new European AI Act and its impact on the ongoing development of algorithms in digital applications for neurodegenerative diseases [266].

Acronyms

2x10m 2x10m Preferred Speed with Stop. 37

4x10m 4x10m Preferred Speed without Stop. 36

AD Alzheimer’s disease. 1

AI artificial intelligence. 12

allGait all digital gait features. 41

BNs Bayesian networks. 14

Count Cognitive TUG. 36

CSF cerebrospinal fluid. 2

CV cross-validation. 42

DE differential gene expression. 12

EMA European Medical Agency. 11

FDA U.S. Food and Drug Administration. 3

GMMs gaussian mixture models. 18

GSEA gene-set enrichment analysis. 12

GWAS genome-wide association studies. 12

H&Y Hoehn & Yahr. 34

HD Huntington’s disease. 1

HD-ISS HD Integrated Staging System. 6

HS Heel Strike. 37

HTT huntingtin gene. 5

KD knock-down. 13

KO knock-out. 13

LSTM Long Short-Term Memory. 19

LTJMM latent time joint mixed-effect model. 32, 38

LuxPARK Luxembourg Parkinson Study. 33

MCI mild cognitive impairment. 2

ML machine learning. 14

MMSE mini mental state examination. 26

NDDs Neurodegenerative Diseases. 1

ODEs ordinary differential equations. 13

PD Parkinson’s disease. 1

PDEs partial differential equations. 13

PIGD Postural Instability and Gait Difficulty. 33

PPI protein-protein interaction. 13

RBD REM sleep behavior disorder. 4

RCTs randomized clinical trials. 58

SDMT Symbol Digit Modality Test. 29

singleGait single digital gait features. 41

TD Tremor Dominance. 33

TO Toe Off. 37

Tray Manual TUG. 36

TUG Timed Up and Go. 20, 36

UHDRS Unified Huntington’s Disease Rating Scale. 9

UPDRS Unified Parkinson’s Disease Rating Scale. 9

VaDER Variational Deep Embedding with Recurrence. 19

VAMBN Variational Autoencoder Modular Bayesian Networks. 26

References

1. Dugger, B. N. & Dickson, D. W. Pathology of Neurodegenerative Diseases. *Cold Spring Harbor Perspectives in Biology* **9**, a028035. ISSN: 1943-0264 (July 2017).
2. 2024 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia* **20**, 3708–3821. ISSN: 1552-5260, 1552-5279 (May 2024).
3. Bloem, B. R., Okun, M. S. & Klein, C. Parkinson's Disease. *The Lancet* **397**, 2284–2303. ISSN: 01406736 (June 2021).
4. Ross, C. A. *et al.* Huntington Disease: Natural History, Biomarkers and Prospects for Therapeutics. *Nature Reviews Neurology* **10**, 204–216. ISSN: 1759-4758, 1759-4766 (Apr. 2014).
5. Scheltens, P. *et al.* Alzheimer's Disease. *The Lancet* **397**, 1577–1590. ISSN: 01406736 (Apr. 2021).
6. International, A. D. *World Alzheimer Report 2023. Report Dementia Risk: Never Too Early, Never Too Late* tech. rep. (Sept. 2023).
7. Rajan, K. B. *et al.* Population Estimate of People with Clinical Alzheimer's Disease and Mild Cognitive Impairment in the United States (2020–2060). *Alzheimer's & Dementia* **17**, 1966–1975. ISSN: 1552-5260, 1552-5279 (Dec. 2021).
8. Wimo, A. *et al.* The Worldwide Costs of Dementia in 2019. *Alzheimer's & dementia : the journal of the Alzheimer's Association* **19**. ISSN: 1552-5279 (July 2023).
9. Monteiro, A. R., Barbosa, D. J., Remião, F. & Silva, R. Alzheimer's Disease: Insights and New Prospects in Disease Pathophysiology, Biomarkers and Disease-Modifying Drugs. *Biochemical Pharmacology* **211**, 115522. ISSN: 00062952 (May 2023).
10. Khan, S., Barve, K. H. & Kumar, M. S. Recent Advancements in Pathogenesis, Diagnostics and Treatment of Alzheimer's Disease. *Current Neuropharmacology* **18**, 1106–1125. ISSN: 1570159X (Nov. 2020).

11. Selkoe, D. J. & Hardy, J. The Amyloid Hypothesis of Alzheimer's Disease at 25 Years. *EMBO Molecular Medicine* **8**, 595–608. ISSN: 1757-4676, 1757-4684 (June 2016).
12. Saeed, A., Lopez, O., Cohen, A. & Reis, S. E. Cardiovascular Disease and Alzheimer's Disease: The Heart–Brain Axis. *Journal of the American Heart Association* **12**, e030780. ISSN: 2047-9980 (Nov. 2023).
13. Heneka, M. T. *et al.* Neuroinflammation in Alzheimer's Disease. *The Lancet Neurology* **14**, 388–405. ISSN: 14744422 (Apr. 2015).
14. Anstey, K. J., Ashby-Mitchell, K. & Peters, R. Updating the Evidence on the Association between Serum Cholesterol and Risk of Late-Life Dementia: Review and Meta-Analysis. *Journal of Alzheimer's Disease* **56**, 215–228. ISSN: 13872877, 18758908 (Jan. 2017).
15. Sperling, R. A. *et al.* Toward Defining the Preclinical Stages of Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease. *Alzheimer's & Dementia* **7**, 280–292. ISSN: 1552-5260, 1552-5279 (May 2011).
16. Albert, M. S. *et al.* The Diagnosis of Mild Cognitive Impairment Due to Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease. *Alzheimer's & Dementia* **7**, 270–279. ISSN: 1552-5260, 1552-5279 (May 2011).
17. McKhann, G. M. *et al.* The Diagnosis of Dementia Due to Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease. *Alzheimer's & Dementia* **7**, 263–269. ISSN: 1552-5260, 1552-5279 (May 2011).
18. Vermunt, L. *et al.* Duration of Preclinical, Prodromal, and Dementia Stages of Alzheimer's Disease in Relation to Age, Sex, and *APOE* Genotype. *Alzheimer's & Dementia* **15**, 888–898. ISSN: 1552-5260, 1552-5279 (July 2019).

19. McKhann, G. *et al.* Clinical Diagnosis of Alzheimer’s Disease: Report of the NINCDS-ADRDA Work Group* under the Auspices of Department of Health and Human Services Task Force on Alzheimer’s Disease. *Neurology* **34**, 939–939. ISSN: 0028-3878, 1526-632X (July 1984).
20. Ferrari, C. & Sorbi, S. The Complexity of Alzheimer’s Disease: An Evolving Puzzle. *Physiological Reviews* **101**, 1047–1081. ISSN: 0031-9333, 1522-1210 (July 2021).
21. Teunissen, C. E. *et al.* Blood-Based Biomarkers for Alzheimer’s Disease: Towards Clinical Implementation. *The Lancet Neurology* **21**, 66–77. ISSN: 1474-4422, 1474-4465 (Jan. 2022).
22. Jack, C. R. *et al.* Revised Criteria for the Diagnosis and Staging of Alzheimer’s Disease. *Nature medicine* **30**, 2121–2124. ISSN: 1078-8956 (Aug. 2024).
23. Bellenguez, C. *et al.* New Insights into the Genetic Etiology of Alzheimer’s Disease and Related Dementias. *Nature Genetics* **54**, 412–436. ISSN: 1061-4036, 1546-1718 (Apr. 2022).
24. Jansen, I. E. *et al.* Genome-Wide Meta-Analysis Identifies New Loci and Functional Pathways Influencing Alzheimer’s Disease Risk. *Nature Genetics* **51**, 404–413. ISSN: 1061-4036, 1546-1718 (Mar. 2019).
25. Saunders, A. M. *et al.* Association of Apolipoprotein E Allele *E4* with Late-onset Familial and Sporadic Alzheimer’s Disease. *Neurology* **43**, 1467–1467. ISSN: 0028-3878, 1526-632X (Aug. 1993).
26. Van Der Lee, S. J. *et al.* The Effect of APOE and Other Common Genetic Variants on the Onset of Alzheimer’s Disease and Dementia: A Community-Based Cohort Study. *The Lancet Neurology* **17**, 434–444. ISSN: 14744422 (May 2018).
27. Jonsson, T. *et al.* A Mutation in APP Protects against Alzheimer’s Disease and Age-Related Cognitive Decline. *Nature* **488**, 96–99. ISSN: 0028-0836, 1476-4687 (Aug. 2012).
28. Guerreiro, R. *et al.* *TREM2* Variants in Alzheimer’s Disease. *New England Journal of Medicine* **368**, 117–127. ISSN: 0028-4793, 1533-4406 (Jan. 2013).
29. Jonsson, T. *et al.* Variant of *TREM2* Associated with the Risk of Alzheimer’s Disease. *New England Journal of Medicine* **368**, 107–116. ISSN: 0028-4793, 1533-4406 (Jan. 2013).

30. Griciuc, A. *et al.* TREM2 Acts Downstream of CD33 in Modulating Microglial Pathology in Alzheimer's Disease. *Neuron* **103**, 820–835.e7. ISSN: 08966273 (Sept. 2019).
31. Estus, S. *et al.* Evaluation of CD33 as a Genetic Risk Factor for Alzheimer's Disease. *Acta Neuropathologica* **138**, 187–199. ISSN: 1432-0533 (Aug. 2019).
32. Jiang, T. *et al.* CD33 in Alzheimer's Disease. *Molecular Neurobiology* **49**, 529–535. ISSN: 1559-1182 (Feb. 2014).
33. Zhao, L. CD33 in Alzheimer's Disease - Biology, Pathogenesis, and Therapeutics: A Mini-Review. *Gerontology* **65**, 323–331. ISSN: 1423-0003 (July 2019).
34. Siddiqui, S. S. *et al.* The Alzheimer's Disease–Protective CD33 Splice Variant Mediates Adaptive Loss of Function via Diversion to an Intracellular Pool. *Journal of Biological Chemistry* **292**, 15312–15320. ISSN: 00219258 (Sept. 2017).
35. Wißfeld, J. *et al.* Deletion of Alzheimer's Disease-Associated CD33 Results in an Inflammatory Human Microglia Phenotype. *Glia* **69**, 1393–1412. ISSN: 1098-1136 (June 2021).
36. Kim, A. Y., Al Jerdi, S., MacDonald, R. & Triggle, C. R. Alzheimer's Disease and Its Treatment—Yesterday, Today, and Tomorrow. *Frontiers in Pharmacology* **15**, 1399121. ISSN: 1663-9812 (May 2024).
37. Cummings, J. *et al.* Aducanumab Produced a Clinically Meaningful Benefit in Association with Amyloid Lowering. *Alzheimer's Research & Therapy* **13**, 98. ISSN: 1758-9193 (Dec. 2021).
38. Beshir, S. A. *et al.* Aducanumab Therapy to Treat Alzheimer's Disease: A Narrative Review. *International Journal of Alzheimer's Disease* **2022** (ed Abate, G.) 1–10. ISSN: 2090-0252, 2090-8024 (Mar. 2022).
39. Van Dyck, C. H. *et al.* Lecanemab in Early Alzheimer's Disease. *New England Journal of Medicine* **388**, 9–21. ISSN: 0028-4793, 1533-4406 (Jan. 2023).
40. Kang, C. Donanemab: First Approval. *Drugs*. ISSN: 1179-1950 (Sept. 2024).
41. Thambisetty, M. & Howard, R. Lecanemab Trial in AD Brings Hope but Requires Greater Clarity. *Nature Reviews Neurology* **19**, 132–133. ISSN: 1759-4758, 1759-4766 (Mar. 2023).

42. Ben-Shlomo, Y. *et al.* The Epidemiology of Parkinson's Disease. *The Lancet* **403**, 283–292. ISSN: 01406736 (Jan. 2024).
43. Yang, W. *et al.* Current and Projected Future Economic Burden of Parkinson's Disease in the U.S. *npj Parkinson's Disease* **6**, 1–9. ISSN: 2373-8057 (July 2020).
44. Morris, H. R., Spillantini, M. G., Sue, C. M. & Williams-Gray, C. H. The Pathogenesis of Parkinson's Disease. *The Lancet* **403**, 293–304. ISSN: 01406736 (Jan. 2024).
45. Horsager, J. *et al.* Brain-First versus Body-First Parkinson's Disease: A Multimodal Imaging Case-Control Study. *Brain* **143**, 3077–3088. ISSN: 0006-8950, 1460-2156 (Oct. 2020).
46. Xia, R. & Mao, Z.-H. Progression of Motor Symptoms in Parkinson's Disease. *Neuroscience Bulletin* **28**, 39–48. ISSN: 1673-7067, 1995-8218 (Feb. 2012).
47. Armstrong, M. J. & Okun, M. S. Diagnosis and Treatment of Parkinson Disease: A Review. *JAMA* **323**, 548. ISSN: 0098-7484 (Feb. 2020).
48. Moustafa, A. A. *et al.* Motor Symptoms in Parkinson's Disease: A Unified Framework. *Neuroscience & Biobehavioral Reviews* **68**, 727–740. ISSN: 01497634 (Sept. 2016).
49. Mirelman, A. *et al.* Gait Impairments in Parkinson's Disease. *The Lancet Neurology* **18**, 697–708. ISSN: 14744422 (July 2019).
50. Weintraub, D. *et al.* The Neuropsychiatry of Parkinson's Disease: Advances and Challenges. *The Lancet Neurology* **21**, 89–102. ISSN: 14744422 (Jan. 2022).
51. Postuma, R. B. *et al.* MDS Clinical Diagnostic Criteria for Parkinson's Disease: MDS-PD Clinical Diagnostic Criteria. *Movement Disorders* **30**, 1591–1601. ISSN: 08853185 (Oct. 2015).
52. Siderowf, A. *et al.* Assessment of Heterogeneity among Participants in the Parkinson's Progression Markers Initiative Cohort Using α -Synuclein Seed Amplification: A Cross-Sectional Study. *The Lancet Neurology* **22**, 407–417. ISSN: 14744422 (May 2023).
53. Tabrizi, S. J. *et al.* A Biological Classification of Huntington's Disease: The Integrated Staging System. *The Lancet Neurology* **21**, 632–644. ISSN: 14744422 (July 2022).

54. Chahine, L. M. *et al.* Proposal for a Biologic Staging System of Parkinson's Disease. *Journal of Parkinson's Disease* **13**, 297–309. ISSN: 18777171, 1877718X (May 2023).
55. Blauwendraat, C., Nalls, M. A. & Singleton, A. B. The Genetic Architecture of Parkinson's Disease. *The Lancet Neurology* **19**, 170–178. ISSN: 14744422 (Feb. 2020).
56. Noyce, A. J. *et al.* Meta-analysis of Early Nonmotor Features and Risk Factors for Parkinson Disease. *Annals of Neurology* **72**, 893–901. ISSN: 0364-5134, 1531-8249 (Dec. 2012).
57. Bohnen, N. I. & Hu, M. T. Sleep Disturbance as Potential Risk and Progression Factor for Parkinson's Disease. *Journal of Parkinson's Disease* **9**, 603–614. ISSN: 18777171, 1877718X (July 2019).
58. Foltynie, T. *et al.* Medical, Surgical, and Physical Treatments for Parkinson's Disease. *The Lancet* **403**, 305–324. ISSN: 01406736 (Jan. 2024).
59. Fox, S. H. *et al.* International Parkinson and Movement Disorder Society Evidence-based Medicine Review: Update on Treatments for the Motor Symptoms of Parkinson's Disease. *Movement Disorders* **33**, 1248–1266. ISSN: 0885-3185, 1531-8257 (Aug. 2018).
60. Pd Med Collaborative Group. Long-Term Effectiveness of Dopamine Agonists and Monoamine Oxidase B Inhibitors Compared with Levodopa as Initial Treatment for Parkinson's Disease (PD MED): A Large, Open-Label, Pragmatic Randomised Trial. *The Lancet* **384**, 1196–1205. ISSN: 01406736 (Sept. 2014).
61. Seppi, K. *et al.* Update on Treatments for Nonmotor Symptoms of Parkinson's Disease—an Evidence-based Medicine Review. *Movement Disorders* **34**, 180–198. ISSN: 0885-3185, 1531-8257 (Feb. 2019).
62. McFarthing, K. *et al.* Parkinson's Disease Drug Therapies in the Clinical Trial Pipeline: 2023 Update. *Journal of Parkinson's Disease* **13**, 427–439. ISSN: 18777171, 1877718X (June 2023).
63. Vijiaratnam, N., Simuni, T., Bandmann, O., Morris, H. R. & Foltynie, T. Progress towards Therapies for Disease Modification in Parkinson's Disease. *The Lancet Neurology* **20**, 559–572. ISSN: 14744422 (July 2021).

64. Lang, A. E. *et al.* Trial of Cinpanemab in Early Parkinson’s Disease. *New England Journal of Medicine* **387**, 408–420. ISSN: 0028-4793, 1533-4406 (Aug. 2022).
65. Pagano, G. *et al.* Trial of Prasinezumab in Early-Stage Parkinson’s Disease. *New England Journal of Medicine* **387**, 421–432. ISSN: 0028-4793, 1533-4406 (Aug. 2022).
66. The Parkinson Study Group SURE-PD3 Investigators *et al.* Effect of Urate-Elevating Inosine on Early Parkinson Disease Progression: The SURE-PD3 Randomized Clinical Trial. *JAMA* **326**, 926. ISSN: 0098-7484 (Sept. 2021).
67. Simuni, T. *et al.* Efficacy of Nilotinib in Patients With Moderately Advanced Parkinson Disease: A Randomized Clinical Trial. *JAMA Neurology* **78**, 312. ISSN: 2168-6149 (Mar. 2021).
68. MacDonald, M. E. *et al.* A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington’s Disease Chromosomes. *Cell* **72**, 971–983. ISSN: 0092-8674, 1097-4172 (Mar. 1993).
69. Medina, A., Mahjoub, Y., Shaver, L. & Pringsheim, T. Prevalence and Incidence of Huntington’s Disease: An Updated Systematic Review and Meta-Analysis. *Movement Disorders: Official Journal of the Movement Disorder Society* **37**, 2327–2335. ISSN: 1531-8257 (Dec. 2022).
70. Chao, T.-K., Hu, J. & Pringsheim, T. Risk Factors for the Onset and Progression of Huntington Disease. *NeuroToxicology* **61**, 79–99. ISSN: 0161813X (July 2017).
71. Ross, C. A. & Tabrizi, S. J. Huntington’s Disease: From Molecular Pathogenesis to Clinical Treatment. *The Lancet Neurology* **10**, 83–98. ISSN: 14744422 (Jan. 2011).
72. Browne, S. E. *et al.* Oxidative Damage and Metabolic Dysfunction in Huntington’s Disease: Selective Vulnerability of the Basal Ganglia. *Annals of Neurology* **41**, 646–653. ISSN: 0364-5134, 1531-8249 (May 1997).
73. Tabrizi, S. J. *et al.* Potential Disease-Modifying Therapies for Huntington’s Disease: Lessons Learned and Future Opportunities. *The Lancet Neurology* **21**, 645–658. ISSN: 14744422 (July 2022).
74. Politis, M. *et al.* Increased Central Microglial Activation Associated with Peripheral Cytokine Levels in Premanifest Huntington’s Disease Gene Carriers. *Neurobiology of Disease* **83**, 115–121. ISSN: 1095-953X (Nov. 2015).

75. Unified Huntington's Disease Rating Scale: Reliability and Consistency. *Movement Disorders* **11**, 136–142. ISSN: 0885-3185, 1531-8257 (Mar. 1996).
76. Ross, C. A., Pantelyat, A., Kogan, J. & Brandt, J. Determinants of Functional Disability in Huntington's Disease: Role of Cognitive and Motor Dysfunction. *Movement Disorders: Official Journal of the Movement Disorder Society* **29**, 1351–1358. ISSN: 1531-8257 (Sept. 2014).
77. van Walsem, M. R., Howe, E. I., Ruud, G. A., Frich, J. C. & Andelic, N. Health-Related Quality of Life and Unmet Healthcare Needs in Huntington's Disease. *Health and Quality of Life Outcomes* **15**, 6. ISSN: 1477-7525 (Jan. 2017).
78. Tabrizi, S. J., Flower, M. D., Ross, C. A. & Wild, E. J. Huntington Disease: New Insights into Molecular Pathogenesis and Therapeutic Opportunities. *Nature Reviews Neurology* **16**, 529–546. ISSN: 1759-4758, 1759-4766 (Oct. 2020).
79. Adams, J. L. *et al.* Multiple Wearable Sensors in Parkinson and Huntington Disease Individuals: A Pilot Study in Clinic and at Home. *Digital Biomarkers* **1**, 52–63. ISSN: 2504-110X (Aug. 2017).
80. Collett, J. *et al.* Insights into Gait Disorders: Walking Variability Using Phase Plot Analysis, Huntington's Disease. *Gait & Posture* **40**, 694–700. ISSN: 09666362 (Sept. 2014).
81. Gaßner, H. *et al.* Gait Variability as Digital Biomarker of Disease Severity in Huntington's Disease. *Journal of Neurology* **267**, 1594–1601. ISSN: 0340-5354, 1432-1459 (June 2020).
82. Prilenia. *PRidopidine's Outcome On Function in Huntington Disease, PROOF-HD* 2022.
83. Kim, C. K. *et al.* Alzheimer's Disease: Key Insights from Two Decades of Clinical Trial Failures. *Journal of Alzheimer's Disease* **87**, 83–100. ISSN: 13872877, 18758908 (May 2022).
84. Horsager, J. & Borghammer, P. Brain-First vs. Body-First Parkinson's Disease: An Update on Recent Evidence. *Parkinsonism & Related Disorders* **122**, 106101. ISSN: 13538020 (May 2024).
85. Tolosa, E., Garrido, A., Scholz, S. W. & Poewe, W. Challenges in the Diagnosis of Parkinson's Disease. *The Lancet Neurology* **20**, 385–397. ISSN: 14744422 (May 2021).

86. Goetz, C. G. *et al.* 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. ISSN: 1531-8257 (Nov. 2008).
87. FDA. *The Drug Development Process* <https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process>. Feb. 2020.
88. Singh, N. *et al.* Drug Discovery and Development: Introduction to the General Public and Patient Groups. *Frontiers in Drug Discovery* **3**, 1201419. ISSN: 2674-0338 (May 2023).
89. Hughes, J., Rees, S., Kalindjian, S. & Philpott, K. Principles of Early Drug Discovery. *British Journal of Pharmacology* **162**, 1239–1249. ISSN: 0007-1188, 1476-5381 (Mar. 2011).
90. Mohs, R. C. & Greig, N. H. Drug Discovery and Development: Role of Basic Biological Research. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* **3**, 651–657. ISSN: 2352-8737, 2352-8737 (Nov. 2017).
91. Kandi, V., Vadakedath, S., Kandi, V. & Vadakedath, S. Clinical Trials and Clinical Research: A Comprehensive Review. *Cureus* **15**. ISSN: 2168-8184 (Feb. 2023).
92. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. *Journal of Health Economics* **47**, 20–33. ISSN: 0167-6296 (May 2016).
93. Gashaw, I., Ellinghaus, P., Sommer, A. & Asadullah, K. What Makes a Good Drug Target? *Drug Discovery Today. Strategic Approach to Target Identification and Validation: A Supplement to Drug Discovery Today* **17**, S24–S30. ISSN: 1359-6446 (Feb. 2012).
94. Aborageh, M., Krawitz, P. & Fröhlich, H. Genetics in Parkinson's Disease: From Better Disease Understanding to Machine Learning Based Precision Medicine. *Frontiers in Molecular Medicine* **2**, 933383. ISSN: 2674-0095 (Oct. 2022).
95. Manzoni, C. *et al.* Genome, Transcriptome and Proteome: The Rise of Omics Data and Their Integration in Biomedical Sciences. *Briefings in Bioinformatics* **19**, 286–302. ISSN: 1467-5463, 1477-4054 (Mar. 2018).

96. Butcher, E. C., Berg, E. L. & Kunkel, E. J. Systems Biology in Drug Discovery. *Nature Biotechnology* **22**, 1253–1259. ISSN: 1087-0156, 1546-1696 (Oct. 2004).
97. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network Medicine: A Network-Based Approach to Human Disease. *Nature Reviews Genetics* **12**, 56–68. ISSN: 1471-0056, 1471-0064 (Jan. 2011).
98. Katsila, T., Spyroulias, G. A., Patrinos, G. P. & Matsoukas, M.-T. Computational Approaches in Target Identification and Drug Discovery. *Computational and Structural Biotechnology Journal* **14**, 177–184. ISSN: 20010370 (2016).
99. Szklarczyk, D. *et al.* STRING V11: Protein–Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Research* **47**, D607–D613. ISSN: 0305-1048, 1362-4962 (Jan. 2019).
100. Oughtred, R. *et al.* The BioGRID Interaction Database: 2019 Update. *Nucleic Acids Research* **47**, D529–D541. ISSN: 0305-1048, 1362-4962 (Jan. 2019).
101. Orchard, S. *et al.* The MIntAct Project—IntAct as a Common Curation Platform for 11 Molecular Interaction Databases. *Nucleic Acids Research* **42**, D358–D363. ISSN: 0305-1048, 1362-4962 (Jan. 2014).
102. Carvalho-Silva, D. *et al.* Open Targets Platform: New Developments and Updates Two Years On. *Nucleic Acids Research* **47**, D1056–D1065. ISSN: 0305-1048, 1362-4962 (Jan. 2019).
103. Wang, Y. *et al.* Therapeutic Target Database 2020: Enriched Resource for Facilitating Research and Early Development of Targeted Therapeutics. *Nucleic Acids Research* **48**, D1031–D1041. ISSN: 0305-1048 (Jan. 2020).
104. Piñero, J. *et al.* The DisGeNET Knowledge Platform for Disease Genomics: 2019 Update. *Nucleic Acids Research* **48**, D845–D855. ISSN: 1362-4962 (Jan. 2020).
105. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Research* **45**, D353–D361. ISSN: 0305-1048 (Jan. 2017).
106. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Research* **46**, D649–D655. ISSN: 0305-1048 (Jan. 2018).

107. Cerami, E. G. *et al.* Pathway Commons, a Web Resource for Biological Pathway Data. *Nucleic Acids Research* **39**, D685–D690. ISSN: 0305-1048 (Jan. 2011).
108. Slenter, D. N. *et al.* WikiPathways: A Multifaceted Pathway Database Bridging Metabolomics to Other Omics Research. *Nucleic Acids Research* **46**, D661–D667. ISSN: 0305-1048 (Jan. 2018).
109. Sobie, E. A., Lee, Y.-S., Jenkins, S. L. & Iyengar, R. Systems Biology—Biomedical Modeling. *Science Signaling* **4**. ISSN: 1945-0877, 1937-9145 (Sept. 2011).
110. Aldridge, B. B., Burke, J. M., Lauffenburger, D. A. & Sorger, P. K. Physicochemical Modelling of Cell Signalling Pathways. *Nature Cell Biology* **8**, 1195–1203. ISSN: 1476-4679 (Nov. 2006).
111. Kitano, H. Computational Systems Biology. *Nature* **420**, 206–210. ISSN: 0028-0836, 1476-4687 (Nov. 2002).
112. Brännmark, C. *et al.* Insulin Signaling in Type 2 Diabetes. *The Journal of Biological Chemistry* **288**, 9867–9880. ISSN: 0021-9258 (Apr. 2013).
113. Proctor, C. J., Boche, D., Gray, D. A. & Nicoll, J. A. R. Investigating Interventions in Alzheimer’s Disease with Computer Simulation Models. *PloS One* **8**, e73631. ISSN: 1932-6203 (2013).
114. Shmulevich, I., Dougherty, E. R., Kim, S. & Zhang, W. Probabilistic Boolean Networks: A Rule-Based Uncertainty Model for Gene Regulatory Networks. *Bioinformatics* **18**, 261–274. ISSN: 1367-4811, 1367-4803 (Feb. 2002).
115. Wilkinson, D. J. Bayesian Methods in Bioinformatics and Computational Systems Biology. *Briefings in Bioinformatics* **8**, 109–116. ISSN: 1467-5463 (Mar. 2007).
116. Needham, C. J., Bradford, J. R., Bulpitt, A. J. & Westhead, D. R. A Primer on Learning in Bayesian Networks for Computational Biology. *PLOS Computational Biology* **3**, e129. ISSN: 1553-7358 (Aug. 2007).
117. Cummings, J. *et al.* Alzheimer’s Disease Drug Development Pipeline: 2023. *Alzheimer’s & Dementia: Translational Research & Clinical Interventions* **9**, e12385. ISSN: 2352-8737, 2352-8737 (Apr. 2023).
118. Wolff, A. *et al.* Parkinson’s Disease Therapy: What Lies Ahead? *Journal of Neural Transmission* **130**, 793–820. ISSN: 0300-9564 (2023).

119. Stepanov, A., Karelina, T., Markevich, N., Demin, O. & Nicholas, T. A Mathematical Model of Multisite Phosphorylation of Tau Protein. *PLoS ONE* **13**, e0192519. ISSN: 1932-6203 (Feb. 2018).
120. Hao, W. & Friedman, A. Mathematical Model on Alzheimer's Disease. *BMC Systems Biology* **10**, 108. ISSN: 1752-0509 (Nov. 2016).
121. Van Maanen, E. M. T. *et al.* Systems Pharmacology Analysis of the Amyloid Cascade after β -Secretase Inhibition Enables the Identification of an A β 42 Oligomer Pool. *Journal of Pharmacology and Experimental Therapeutics* **357**, 205–216. ISSN: 0022-3565, 1521-0103 (Apr. 2016).
122. Geerts, H., Roberts, P., Spiros, A. & Carr, R. A Strategy for Developing New Treatment Paradigms for Neuropsychiatric and Neurocognitive Symptoms in Alzheimer's Disease. *Frontiers in Pharmacology* **4**, 47. ISSN: 1663-9812 (Apr. 2013).
123. Geerts, H., Spiros, A. & Roberts, P. Impact of Amyloid-Beta Changes on Cognitive Outcomes in Alzheimer's Disease: Analysis of Clinical Trials Using a Quantitative Systems Pharmacology Model. *Alzheimer's Research & Therapy* **10**, 14. ISSN: 1758-9193 (Feb. 2018).
124. Mazer, N. A. *et al.* Development of a Quantitative Semi-Mechanistic Model of Alzheimer's Disease Based on the Amyloid/Tau/Neurodegeneration Framework (the Q-ATN Model). *Alzheimer's & Dementia* **19**, 2287–2297. ISSN: 1552-5279 (2023).
125. Kuznetsov, I. A. & Kuznetsov, A. V. Simulating the Effect of Formation of Amyloid Plaques on Aggregation of Tau Protein. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **474**, 20180511 (Dec. 2018).
126. Ouzounoglou, E. *et al.* In Silico Modeling of the Effects of Alpha-Synuclein Oligomerization on Dopaminergic Neuronal Homeostasis. *BMC Systems Biology* **8**, 54. ISSN: 1752-0509 (May 2014).
127. Roberts, P., Spiros, A. & Geerts, H. A Humanized Clinically Calibrated Quantitative Systems Pharmacology Model for Hypokinetic Motor Symptoms in Parkinson's Disease. *Frontiers in Pharmacology* **7**, 6. ISSN: 1663-9812 (Feb. 2016).

128. Höglinger, G., Trenkwalder, D. C., vla & vlo. Parkinson-Krankheit, S2k-Leitlinie, 2023. *Leitlinien für Diagnostik und Therapie in der Neurologie* (2023).
129. Saft, C. S2k-Leitlinie Chorea/Morbus Huntington (2022).
130. Radder, D. L. M. *et al.* Physiotherapy in Parkinson's Disease: A Meta-Analysis of Present Treatment Modalities. *Neurorehabilitation and Neural Repair* **34**, 871–880. ISSN: 1545-9683, 1552-6844 (Oct. 2020).
131. Ernst, M. *et al.* Physical Exercise for People with Parkinson's Disease: A Systematic Review and Network Meta-Analysis. *Cochrane Database of Systematic Reviews* **2024** (ed Cochrane Movement Disorders Group) ISSN: 14651858 (Apr. 2024).
132. Fröhlich, H. *et al.* From Hype to Reality: Data Science Enabling Personalized Medicine. *BMC Medicine* **16**, 150. ISSN: 1741-7015 (Dec. 2018).
133. Lonergan, M. *et al.* Defining Drug Response for Stratified Medicine. *Drug Discovery Today* **22**, 173–179. ISSN: 13596446 (Jan. 2017).
134. Ginsburg, G. S. & Phillips, K. A. Precision Medicine: From Science to Value. *Health affairs (Project Hope)* **37**, 694–701. ISSN: 0278-2715 (May 2018).
135. Hartl, D. *et al.* Translational Precision Medicine: An Industry Perspective. *Journal of Translational Medicine* **19**, 245. ISSN: 1479-5876 (June 2021).
136. Kosorok, M. R. & Laber, E. B. Precision Medicine. *Annual Review of Statistics and Its Application* **6**, 263–286. ISSN: 2326-8298, 2326-831X (Mar. 2019).
137. Dorsey, E. R., Papapetropoulos, S., Xiong, M. & Kieburtz, K. The First Frontier: Digital Biomarkers for Neurodegenerative Disorders. *Digital Biomarkers* **1**, 6–13. ISSN: 2504-110X (July 2017).
138. Berg, D. *et al.* Prodromal Parkinson Disease Subtypes — Key to Understanding Heterogeneity. *Nature Reviews Neurology* **17**, 349–361. ISSN: 1759-4766 (June 2021).
139. Bakels, H. S., Roos, R. A., van Roon-Mom, W. M. & de Bot, S. T. Juvenile-Onset Huntington Disease Pathophysiology and Neurodevelopment: A Review. *Movement Disorders* **37**, 16–24. ISSN: 0885-3185 (Jan. 2022).
140. Ferreira, D., Nordberg, A. & Westman, E. Biological Subtypes of Alzheimer Disease. *Neurology* **94**, 436–448 (Mar. 2020).

141. Vogel, J. W. *et al.* Four Distinct Trajectories of Tau Deposition Identified in Alzheimer's Disease. *Nature Medicine* **27**, 871–881. ISSN: 1078-8956, 1546-170X (May 2021).
142. Emon, M. A. *et al.* Clustering of Alzheimer's and Parkinson's Disease Based on Genetic Burden of Shared Molecular Mechanisms. *Scientific Reports* **10**, 19097. ISSN: 2045-2322 (Nov. 2020).
143. Cao, L.-X., Yin, J.-H., Du, G., Yang, Q. & Huang, Y. Identifying and Verifying Huntington's Disease Subtypes: Clinical Features, Neuroimaging, and Cytokine Changes. *Brain and Behavior* **14**, e3469. ISSN: 2162-3279 (Mar. 2024).
144. Dadu, A. *et al.* Identification and Prediction of Parkinson's Disease Subtypes and Progression Using Machine Learning in Two Cohorts. *npj Parkinson's Disease* **8**, 172. ISSN: 2373-8057 (Dec. 2022).
145. Erro, R. *et al.* Clinical Clusters and Dopaminergic Dysfunction in De-Novo Parkinson Disease. *Parkinsonism & Related Disorders* **28**, 137–140. ISSN: 1353-8020 (July 2016).
146. Fereshtehnejad, S.-M. *et al.* New Clinical Subtypes of Parkinson Disease and Their Longitudinal Progression: A Prospective Cohort Comparison With Other Phenotypes. *JAMA Neurology* **72**, 863–873. ISSN: 2168-6149 (Aug. 2015).
147. Graham, J. M. & Sagar, H. J. A Data-Driven Approach to the Study of Heterogeneity in Idiopathic Parkinson's Disease: Identification of Three Distinct Subtypes. *Movement Disorders* **14**, 10–20. ISSN: 1531-8257 (1999).
148. Lawton, M. *et al.* Developing and Validating Parkinson's Disease Subtypes and Their Motor and Cognitive Progression. *Journal of Neurology, Neurosurgery & Psychiatry* **89**, 1279–1287. ISSN: 0022-3050, 1468-330X (Dec. 2018).
149. Shakya, S., Prevett, J., Hu, X. & Xiao, R. Characterization of Parkinson's Disease Subtypes and Related Attributes. *Frontiers in Neurology* **13**, 810038. ISSN: 1664-2295 (May 2022).
150. van Rooden, S. M. *et al.* Clinical Subtypes of Parkinson's Disease. *Movement Disorders: Official Journal of the Movement Disorder Society* **26**, 51–58. ISSN: 1531-8257 (Jan. 2011).

151. Zhou, Z. *et al.* Subtyping of Early-Onset Parkinson's Disease Using Cluster Analysis: A Large Cohort Study. *Frontiers in Aging Neuroscience* **14**, 1040293. ISSN: 1663-4365 (Nov. 2022).
152. Fereshtehnejad, S.-M., Zeighami, Y., Dagher, A. & Postuma, R. B. Clinical Criteria for Subtyping Parkinson's Disease: Biomarkers and Longitudinal Progression. *Brain: A Journal of Neurology* **140**, 1959–1976. ISSN: 1460-2156 (July 2017).
153. Blanken, A. E., Dutt, S., Li, Y., Nation, D. A. & Alzheimer's Disease Neuroimaging Initiative. Disentangling Heterogeneity in Alzheimer's Disease: Two Empirically-Derived Subtypes. *Journal of Alzheimer's disease: JAD* **70**, 227–239. ISSN: 1875-8908 (2019).
154. Zetusk, W. J., Jankovic, J. & Pirozzolo, F. J. The Heterogeneity of Parkinson's Disease: Clinical and Prognostic Implications. *Neurology* **35**, 522–526. ISSN: 0028-3878 (Apr. 1985).
155. Hart, E. P. *et al.* Better Global and Cognitive Functioning in Choreatic versus Hypokinetic-Rigid Huntington's Disease. *Movement Disorders* **28**, 1142–1145. ISSN: 1531-8257 (2013).
156. Borghammer, P. The α -Synuclein Origin and Connectome Model (SOC Model) of Parkinson's Disease: Explaining Motor Asymmetry, Non-Motor Phenotypes, and Cognitive Decline. *Journal of Parkinson's Disease* **11**, 455–474. ISSN: 18777171, 1877718X (Apr. 2021).
157. Birkenbihl, C. *et al.* Artificial Intelligence-Based Clustering and Characterization of Parkinson's Disease Trajectories. *Scientific Reports* **13**, 2897. ISSN: 2045-2322 (Feb. 2023).
158. Hähnel, T. *et al.* *Predictive Modeling to Uncover Parkinson's Disease Characteristics That Delay Diagnosis* Mar. 2024.
159. Zhang, X. *et al.* Data-Driven Subtyping of Parkinson's Disease Using Longitudinal Clinical Records: A Cohort Study. *Scientific Reports* **9**, 797. ISSN: 2045-2322 (Jan. 2019).
160. de Jong, J. *et al.* Deep Learning for Clustering of Multivariate Clinical Patient Trajectories with Missing Values. *GigaScience* **8**, giz134. ISSN: 2047-217X (Nov. 2019).

161. Fröhlich, H. *et al.* Leveraging the Potential of Digital Technology for Better Individualized Treatment of Parkinson’s Disease. *Frontiers in Neurology* **13**, 788427. ISSN: 1664-2295 (Feb. 2022).
162. Zhan, A. *et al.* Using Smartphones and Machine Learning to Quantify Parkinson Disease Severity: The Mobile Parkinson Disease Score. *JAMA Neurology* **75**, 876. ISSN: 2168-6149 (July 2018).
163. Arora, S. *et al.* Smartphone Motor Testing to Distinguish Idiopathic REM Sleep Behavior Disorder, Controls, and PD. *Neurology* **91**. ISSN: 0028-3878, 1526-632X (Oct. 2018).
164. Robin, J., Xu, M., Kaufman, L. D. & Simpson, W. Using Digital Speech Assessments to Detect Early Signs of Cognitive Impairment. *Frontiers in Digital Health* **3**. ISSN: 2673-253X (Oct. 2021).
165. Jeancolas, L. *et al.* X-Vectors: New Quantitative Biomarkers for Early Parkinson’s Disease Detection From Speech. *Frontiers in Neuroinformatics* **15**. ISSN: 1662-5196 (Feb. 2021).
166. Gómez-Vilda, P. *et al.* Parkinson Disease Detection from Speech Articulation Neuromechanics. *Frontiers in Neuroinformatics* **11**. ISSN: 1662-5196 (Aug. 2017).
167. Shahbakhi, M., Far, D. T. & Tahami, E. Speech Analysis for Diagnosis of Parkinson’s Disease Using Genetic Algorithm and Support Vector Machine. *Journal of Biomedical Science and Engineering* **7**, 147–156 (Mar. 2014).
168. Wang, M. & Deng, W. Deep Face Recognition: A Survey. *Neurocomputing* **429**, 215–244. ISSN: 0925-2312 (Mar. 2021).
169. Rajnoha, M. *et al.* *Towards Identification of Hypomimia in Parkinson’s Disease Based on Face Recognition Methods in 2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)* (Nov. 2018), 1–4.
170. Bandini, A. *et al.* Analysis of Facial Expressions in Parkinson’s Disease through Video-Based Automatic Methods. *Journal of Neuroscience Methods* **281**, 7–20. ISSN: 0165-0270 (Apr. 2017).
171. Filali Razzouki, A. *et al.* *Early-Stage Parkinson’s Disease Detection Based on Action Unit Derivatives in Colloque En TéléSANTé et Dispositifs Biomédicaux (JETSAN)* (Université Paris 8, CNRS, Paris Saint Denis, France, June 2023).

172. Andrzejewski, K. L. *et al.* Wearable Sensors in Huntington Disease: A Pilot Study. *Journal of Huntington's Disease* **5**, 199–206. ISSN: 18796397, 18796400 (July 2016).
173. Borzì, L. *et al.* Prediction of Freezing of Gait in Parkinson's Disease Using Wearables and Machine Learning. *Sensors* **21**, 614. ISSN: 1424-8220 (Jan. 2021).
174. Sjaelland, N. S., Gramkow, M. H., Hasselbalch, S. G. & Frederiksen, K. S. Digital Biomarkers for the Assessment of Non-Cognitive Symptoms in Patients with Dementia with Lewy Bodies: A Systematic Review. *Journal of Alzheimer's Disease* (ed Duraes, J.) 1–21. ISSN: 13872877, 18758908 (June 2024).
175. Sotirakis, C. *et al.* Identification of Motor Progression in Parkinson's Disease Using Wearable Sensors and Machine Learning. *npj Parkinson's Disease* **9**, 142. ISSN: 2373-8057 (Oct. 2023).
176. Sun, Y.-m., Wang, Z.-y., Liang, Y.-y., Hao, C.-w. & Shi, C.-h. Digital Biomarkers for Precision Diagnosis and Monitoring in Parkinson's Disease. *npj Digital Medicine* **7**, 1–10. ISSN: 2398-6352 (Aug. 2024).
177. Chandrabhatla, A. S., Pomeranec, I. J. & Ksendzovsky, A. Co-Evolution of Machine Learning and Digital Technologies to Improve Monitoring of Parkinson's Disease Motor Symptoms. *npj Digital Medicine* **5**, 32. ISSN: 2398-6352 (Mar. 2022).
178. Elbatanouny, H. *et al.* Insights into Parkinson's Disease-Related Freezing of Gait Detection and Prediction Approaches: A Meta Analysis. *Sensors (Basel, Switzerland)* **24**, 3959. ISSN: 1424-8220 (June 2024).
179. Hssayeni, M. D., Burack, M. A., Jimenez-Shahed, J. & Ghoraani, B. Assessment of Response to Medication in Individuals with Parkinson's Disease. *Medical Engineering & Physics* **67**, 33–43. ISSN: 1873-4030 (May 2019).
180. Castelli Gattinara Di Zubiena, F. *et al.* Machine Learning and Wearable Sensors for the Early Detection of Balance Disorders in Parkinson's Disease. *Sensors* **22**, 9903. ISSN: 1424-8220 (Dec. 2022).
181. Maetzler, W., Domingos, J., Srulijes, K., Ferreira, J. J. & Bloem, B. R. Quantitative Wearable Sensors for Objective Assessment of Parkinson's Disease. *Movement Disorders* **28**, 1628–1637. ISSN: 0885-3185, 1531-8257 (Oct. 2013).

182. Kirk, C. *et al.* Can Digital Mobility Assessment Enhance the Clinical Assessment of Disease Severity in Parkinson’s Disease? *Journal of Parkinson’s Disease* **13**, 999–1009. ISSN: 18777171, 1877718X (Sept. 2023).
183. Nair, P., Shojaei Baghini, M., Pendharkar, G. & Chung, H. Detecting Early-Stage Parkinson’s Disease from Gait Data. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 09544119231197090. ISSN: 0954-4119, 2041-3033 (Nov. 2023).
184. Deng, K. *et al.* Heterogeneous Digital Biomarker Integration Out-Performs Patient Self-Reports in Predicting Parkinson’s Disease. *Communications Biology* **5**, 58. ISSN: 2399-3642 (Jan. 2022).
185. Djurić-Jovičić, M., Belić, M., Stanković, I., Radovanović, S. & Kostić, V. S. Selection of Gait Parameters for Differential Diagnostics of Patients with *de Novo* Parkinson’s Disease. *Neurological Research* **39**, 853–861. ISSN: 0161-6412, 1743-1328 (Oct. 2017).
186. Adams, J. L. *et al.* Using a Smartwatch and Smartphone to Assess Early Parkinson’s Disease in the WATCH-PD Study. *npj Parkinson’s Disease* **9**, 64. ISSN: 2373-8057 (Apr. 2023).
187. Pistacchi, M. Gait Analysis and Clinical Correlations in Early Parkinson’s Disease. *Functional Neurology* **32**, 28. ISSN: 0393-5264 (2017).
188. Cai, G. *et al.* Specific Distribution of Digital Gait Biomarkers in Parkinson’s Disease Using Body-Worn Sensors and Machine Learning. *The Journals of Gerontology: Series A* **78** (ed Duque, G.) 1348–1354. ISSN: 1079-5006, 1758-535X (Aug. 2023).
189. Beswick, E. *et al.* A Systematic Review of Digital Technology to Evaluate Motor Function and Disease Progression in Motor Neuron Disease. *Journal of Neurology* **269**, 6254–6268. ISSN: 0340-5354, 1432-1459 (Dec. 2022).
190. Zhu, S. *et al.* Gait Analysis with Wearables Is a Potential Progression Marker in Parkinson’s Disease. *Brain Sciences* **12**, 1213. ISSN: 2076-3425 (Sept. 2022).
191. Diao, J. A., Raza, M. M., Venkatesh, K. P. & Kvedar, J. C. Watching Parkinson’s Disease with Wrist-Based Sensors. *npj Digital Medicine* **5**, 73, s41746-022-00619–4. ISSN: 2398-6352 (June 2022).

192. Lipsmeier, F. *et al.* Evaluation of Smartphone-based Testing to Generate Exploratory Outcome Measures in a Phase 1 Parkinson’s Disease Clinical Trial. *Movement Disorders* **33**, 1287–1297. ISSN: 0885-3185, 1531-8257 (Aug. 2018).
193. Payne, T. *et al.* A Double-Blind, Randomized, Placebo-Controlled Trial of Ursodeoxycholic Acid (UDCA) in Parkinson’s Disease. *Movement Disorders* **38**, 1493–1502. ISSN: 0885-3185, 1531-8257 (Aug. 2023).
194. Espay, A. J. *et al.* A Roadmap for Implementation of Patient-centered Digital Outcome Measures in Parkinson’s Disease Obtained Using Mobile Health Technologies. *Movement Disorders* **34**, 657–663. ISSN: 0885-3185, 1531-8257 (May 2019).
195. Brem, A.-K. *et al.* Digital Endpoints in Clinical Trials of Alzheimer’s Disease and Other Neurodegenerative Diseases: Challenges and Opportunities. *Frontiers in Neurology* **14**, 1210974. ISSN: 1664-2295 (June 2023).
196. Raschka, T. *et al.* AI Reveals Insights into Link between CD33 and Cognitive Impairment in Alzheimer’s Disease. *PLOS Computational Biology* **19** (ed Jensen, L. J.) e1009894. ISSN: 1553-7358 (Feb. 2023).
197. Raschka, T. *et al.* Unraveling Progression Subtypes in People with Huntington’s Disease. *EPMA Journal* **15**, 275–287. ISSN: 1878-5085 (May 2024).
198. Raschka, T. *et al.* *Objective Monitoring of Motor Symptom Severity and Their Progression in Parkinson’s Disease Using a Digital Gait Device* June 2024.
199. Gootjes-Dreesbach, L., Sood, M., Sahay, A., Hofmann-Apitius, M. & Fröhlich, H. Variational Autoencoder Modular Bayesian Networks for Simulation of Heterogeneous Clinical Study Data. *Frontiers in Big Data* **3**. ISSN: 2624-909X (May 2020).
200. Hodson, R. Alzheimer’s Disease. *Nature* **559**, S1. ISSN: 1476-4687 (July 2018).
201. Griciu, A. *et al.* Alzheimer’s Disease Risk Gene CD33 Inhibits Microglial Uptake of Amyloid Beta. *Neuron* **78**, 631–643. ISSN: 1097-4199 (May 2013).
202. Tassoni, D., Kaur, G., Weisinger, R. S. & Sinclair, A. J. The Role of Eicosanoids in the Brain. *Asia Pacific Journal of Clinical Nutrition* **17 Suppl 1**, 220–228. ISSN: 0964-7058 (2008).

203. Ardura-Fabregat, A. *et al.* Targeting Neuroinflammation to Treat Alzheimer's Disease. *CNS drugs* **31**, 1057–1082. ISSN: 1179-1934 (Dec. 2017).
204. Biringer, R. G. The Role of Eicosanoids in Alzheimer's Disease. *International Journal of Environmental Research and Public Health* **16**, 2560. ISSN: 1660-4601 (July 2019).
205. Shioya, M. *et al.* Aberrant microRNA Expression in the Brains of Neurodegenerative Diseases: miR-29a Decreased in Alzheimer Disease Brains Targets Neurone Navigator 3. *Neuropathology and Applied Neurobiology* **36**, 320–330. ISSN: 1365-2990 (June 2010).
206. Seth, R. B., Sun, L., Ea, C.-K. & Chen, Z. J. Identification and Characterization of MAVS, a Mitochondrial Antiviral Signaling Protein That Activates NF-kappaB and IRF 3. *Cell* **122**, 669–682. ISSN: 0092-8674 (Sept. 2005).
207. Horta-Barba, A. *et al.* Measuring Cognitive Impairment and Monitoring Cognitive Decline in Huntington's Disease: A Comparison of Assessment Instruments. *Journal of Neurology* **270**, 5408–5417. ISSN: 0340-5354, 1432-1459 (Nov. 2023).
208. Snowden, J. S. The Neuropsychology of Huntington's Disease. *Archives of Clinical Neuropsychology* **32**, 876–887. ISSN: 0887-6177, 1873-5843 (Nov. 2017).
209. Ho, A. K., Gilbert, A. S., Mason, S. L., Goodman, A. O. & Barker, R. A. Health-related Quality of Life in Huntington's Disease: Which Factors Matter Most? *Movement Disorders* **24**, 574–578. ISSN: 0885-3185, 1531-8257 (Mar. 2009).
210. Banaszkiewicz, K. *et al.* Huntington's Disease from the Patient, Caregiver and Physician's Perspectives: Three Sides of the Same Coin? *Journal of Neural Transmission* **119**, 1361–1365. ISSN: 0300-9564, 1435-1463 (Nov. 2012).
211. Read, J. *et al.* Quality of Life in Huntington's Disease: A Comparative Study Investigating the Impact for Those with Pre-Manifest and Early Manifest Disease, and Their Partners. *Journal of Huntington's Disease* **2**, 159–175. ISSN: 18796397 (2013).
212. Lee, J.-M. *et al.* Identification of Genetic Factors That Modify Clinical Onset of Huntington's Disease. *Cell* **162**, 516–526. ISSN: 00928674 (July 2015).

213. Langbehn, D. R. *et al.* A New Model for Prediction of the Age of Onset and Penetrance for Huntington’s Disease Based on CAG Length. *Clinical Genetics* **65**, 267–277. ISSN: 0009-9163 (Apr. 2004).
214. Tabrizi, S. J. *et al.* Biological and Clinical Manifestations of Huntington’s Disease in the Longitudinal TRACK-HD Study: Cross-Sectional Analysis of Baseline Data. *The Lancet. Neurology* **8**, 791–801. ISSN: 1474-4422 (Sept. 2009).
215. Abeyasinghe, P. M. *et al.* Tracking Huntington’s Disease Progression Using Motor, Functional, Cognitive, and Imaging Markers. *Movement Disorders* **36**, 2282–2292. ISSN: 0885-3185, 1531-8257 (Oct. 2021).
216. Hogarth, P. *et al.* Interrater Agreement in the Assessment of Motor Manifestations of Huntington’s Disease. *Movement Disorders* **20**, 293–297. ISSN: 0885-3185, 1531-8257 (Mar. 2005).
217. Reilmann, R., Leavitt, B. R. & Ross, C. A. Diagnostic Criteria for Huntington’s Disease Based on Natural History. *Movement Disorders* **29**, 1335–1341. ISSN: 0885-3185, 1531-8257 (Sept. 2014).
218. Mandel, S. & Korczyn, A. D. in *Neurodegenerative Diseases: Integrative PPPM Approach as the Medicine of the Future* (ed Mandel, S.) 95–140 (Springer Netherlands, Dordrecht, 2013).
219. EPMA, Golubnitschaja, O. & Costigliola, V. General Report & Recommendations in Predictive, Preventive and Personalised Medicine 2012: White Paper of the European Association for Predictive, Preventive and Personalised Medicine. *EPMA Journal* **3**, 14. ISSN: 1878-5077, 1878-5085 (Dec. 2012).
220. Wang, M. *et al.* Statistical Methods for Studying Disease Subtype Heterogeneity. *Statistics in Medicine* **35**, 782–800. ISSN: 1097-0258 (Feb. 2016).
221. Giannoula, A., Gutierrez-Sacristán, A., Bravo, Á., Sanz, F. & Furlong, L. I. Identifying Temporal Patterns in Patient Disease Trajectories Using Dynamic Time Warping: A Population-Based Study. *Scientific Reports* **8**, 4216. ISSN: 2045-2322 (Mar. 2018).
222. Golubnitschaja, O. *et al.* Medicine in the Early Twenty-First Century: Paradigm and Anticipation - EPMA Position Paper 2016. *EPMA Journal* **7**, 23. ISSN: 1878-5077, 1878-5085 (Dec. 2016).

223. Landwehrmeyer, G. B. *et al.* Data Analytics from Enroll-HD, a Global Clinical Research Platform for Huntington’s Disease. *Movement Disorders Clinical Practice* **4**, 212–224. ISSN: 2330-1619, 2330-1619 (Mar. 2017).
224. Sathe, S. *et al.* Enroll-HD: An Integrated Clinical Research Platform and Worldwide Observational Study for Huntington’s Disease. *Frontiers in Neurology* **12**, 667420. ISSN: 1664-2295 (Aug. 2021).
225. Kühnel, L., Berger, A.-K., Markussen, B. & Raket, L. L. Simultaneous Modeling of Alzheimer’s Disease Progression via Multiple Cognitive Scales. *Statistics in Medicine* **40**, 3251–3266. ISSN: 1097-0258 (June 2021).
226. Raket, L. L. Statistical Disease Progression Modeling in Alzheimer Disease. *Frontiers in Big Data* **3**, 24. ISSN: 2624-909X (Aug. 2020).
227. Chen, T. & Guestrin, C. *XGBoost: A Scalable Tree Boosting System* in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, San Francisco California USA, Aug. 2016), 785–794. ISBN: 978-1-4503-4232-2.
228. Raschka, T. *et al.* *Objective Monitoring of Motor Symptom Severity and Their Progression in Parkinson’s Disease Using a Digital Gait Device* June 2024.
229. Li, D., Iddi, S., Thompson, W. K., Donohue, M. C. & Alzheimer’s Disease Neuroimaging Initiative. Bayesian Latent Time Joint Mixed Effect Models for Multicohort Longitudinal Data. *Statistical Methods in Medical Research* **28**, 835–845. ISSN: 1477-0334 (Mar. 2019).
230. Hipp, G. *et al.* The Luxembourg Parkinson’s Study: A Comprehensive Approach for Stratification and Early Diagnosis. *Frontiers in Aging Neuroscience* **10**, 326. ISSN: 1663-4365 (Oct. 2018).
231. Aleksovski, D., Miljkovic, D., Bravi, D. & Antonini, A. Disease Progression in Parkinson Subtypes: The PPMI Dataset. *Neurological Sciences* **39**, 1971–1976. ISSN: 1590-1874, 1590-3478 (Nov. 2018).
232. Hayashi, Y. *et al.* Unilateral GPI-DBS Improves Ipsilateral and Axial Motor Symptoms in Parkinson’s Disease as Evidenced by a Brain Perfusion Single Photon Emission Computed Tomography Study. *Frontiers in Human Neuroscience* **16**, 888701. ISSN: 1662-5161 (May 2022).
233. Fahn, . . Unified Parkinson’s Disease Rating Scale. *Recent developments in Parkinson’s disease*, 153–163 (1987).

234. Hannink, J. *et al.* Sensor-Based Gait Parameter Extraction With Deep Convolutional Neural Networks. *IEEE Journal of Biomedical and Health Informatics* **21**, 85–93. ISSN: 2168-2194, 2168-2208 (Jan. 2017).
235. Barth, J. *et al.* Stride Segmentation during Free Walk Movements Using Multi-Dimensional Subsequence Dynamic Time Warping on Inertial Sensor Data. *Sensors* **15**, 6419–6440. ISSN: 1424-8220 (Mar. 2015).
236. Rampp, A. *et al.* Inertial Sensor-Based Stride Parameter Calculation From Gait Sequences in Geriatric Patients. *IEEE Transactions on Biomedical Engineering* **62**, 1089–1097. ISSN: 0018-9294, 1558-2531 (Apr. 2015).
237. Küderle, A. *et al.* *Gaitmap - An Open Ecosystem for IMU-based Human Gait Analysis and Algorithm Benchmarking* Preprint (Sept. 2023).
238. Donohue, M. *Mdonohue / Ltjmm — Bitbucket* 2017.
239. Carpenter, B. *et al.* Stan : A Probabilistic Programming Language. *Journal of Statistical Software* **76**. ISSN: 1548-7660 (2017).
240. Benjamini, Y. & Yekutieli, D. The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics* **29**. ISSN: 0090-5364 (Aug. 2001).
241. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32. ISSN: 08856125 (2001).
242. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 267–288. ISSN: 0035-9246, 2517-6161 (Jan. 1996).
243. Pedregosa, F. *et al.* Scikit-Learn: Machine Learning in Python. *the Journal of machine Learning research* **12**, 2825–2830 (2011).
244. SRL., U. B. A Double-Blind, Placebo-Controlled, Randomized, 18-Month Phase 2a Study to Evaluate the Efficacy, Safety, Tolerability, and Pharmacokinetics of Oral UCB0599 in Study Participants With Early Parkinson’s Disease (2023).
245. Ard, M. C. & Edland, S. D. Power Calculations for Clinical Trials in Alzheimer’s Disease. *Journal of Alzheimer’s Disease* **26** (eds Ashford, J. W. *et al.*) 369–377. ISSN: 18758908, 13872877 (Oct. 2011).

246. Iddi, S. & C Donohue, M. Power and Sample Size for Longitudinal Models in R – The Longpower Package and Shiny App. *The R Journal* **14**, 264–282. ISSN: 2073-4859 (July 2022).
247. Tabrizi, S. J. *et al.* Predictors of Phenotypic Progression and Disease Onset in Premanifest and Early-Stage Huntington’s Disease in the TRACK-HD Study: Analysis of 36-Month Observational Data. *The Lancet Neurology* **12**, 637–649. ISSN: 14744422 (July 2013).
248. Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M. & Monfardini, G. The Graph Neural Network Model. *IEEE transactions on neural networks* **20**, 61–80. ISSN: 1941-0093 (Jan. 2009).
249. Krix, S. *et al.* MultiGML: Multimodal Graph Machine Learning for Prediction of Adverse Drug Events. *Heliyon* **9**. ISSN: 2405-8440 (Sept. 2023).
250. Wu, Y., Gao, M., Zeng, M., Zhang, J. & Li, M. BridgeDPI: A Novel Graph Neural Network for Predicting Drug–Protein Interactions. *Bioinformatics* **38**, 2571–2578. ISSN: 1367-4803 (Apr. 2022).
251. Mandel, F., Ghosh, R. P. & Barnett, I. Neural Networks for Clustered and Longitudinal Data Using Mixed Effects Models. *Biometrics* **79**, 711–721. ISSN: 1541-0420 (2023).
252. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE* **86**, 2278–2324. ISSN: 1558-2256 (Nov. 1998).
253. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **9**, 1735–1780. ISSN: 0899-7667 (Nov. 1997).
254. Wu, N., Green, B., Ben, X. & O’Banion, S. *Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case* Jan. 2020. arXiv: 2001.08317 [cs, stat].
255. Bari, A. S. M. H. & Gavrilova, M. L. KinectGaitNet: Kinect-Based Gait Recognition Using Deep Convolutional Neural Network. *Sensors (Basel, Switzerland)* **22**, 2631. ISSN: 1424-8220 (Mar. 2022).
256. Cheriet, M., Dentamaro, V., Hamdan, M., Impedovo, D. & Pirlo, G. Multi-Speed Transformer Network for Neurodegenerative Disease Assessment and Activity Recognition. *Computer Methods and Programs in Biomedicine* **230**, 107344. ISSN: 1872-7565 (Mar. 2023).

257. Dehzangi, O., Taherisadr, M. & ChagalVala, R. IMU-Based Gait Recognition Using Convolutional Neural Networks and Multi-Sensor Fusion. *Sensors (Basel, Switzerland)* **17**, 2735. ISSN: 1424-8220 (Nov. 2017).
258. Kim, J., Seo, H., Naseem, M. T. & Lee, C.-S. Pathological-Gait Recognition Using Spatiotemporal Graph Convolutional Networks and Attention Model. *Sensors (Basel, Switzerland)* **22**, 4863. ISSN: 1424-8220 (June 2022).
259. Li, J., Liang, W., Yin, X., Li, J. & Guan, W. Multimodal Gait Abnormality Recognition Using a Convolutional Neural Network-Bidirectional Long Short-Term Memory (CNN-BiLSTM) Network Based on Multi-Sensor Data Fusion. *Sensors (Basel, Switzerland)* **23**, 9101. ISSN: 1424-8220 (Nov. 2023).
260. Mogan, J. N., Lee, C. P., Lim, K. M. & Muthu, K. S. Gait-ViT: Gait Recognition with Vision Transformer. *Sensors (Basel, Switzerland)* **22**, 7362. ISSN: 1424-8220 (Sept. 2022).
261. Mogan, J. N., Lee, C. P., Lim, K. M., Ali, M. & Alqahtani, A. Gait-CNN-ViT: Multi-Model Gait Recognition with Convolutional Neural Networks and Vision Transformer. *Sensors (Basel, Switzerland)* **23**, 3809. ISSN: 1424-8220 (Apr. 2023).
262. Naimi, S., Bouachir, W. & Bilodeau, G.-A. *HCT: Hybrid Convnet-Transformer for Parkinson's Disease Detection and Severity Prediction from Gait* Oct. 2023. arXiv: 2310.17078 [cs].
263. Hayashi, S., Saho, K., Shioiri, K., Fujimoto, M. & Masugi, M. Utilization of Micro-Doppler Radar to Classify Gait Patterns of Young and Elderly Adults: An Approach Using a Long Short-Term Memory Network. *Sensors (Basel, Switzerland)* **21**, 3643. ISSN: 1424-8220 (May 2021).
264. Jung, D., Nguyen, M. D., Park, M., Kim, J. & Mun, K.-R. Multiple Classification of Gait Using Time-Frequency Representations and Deep Convolutional Neural Networks. *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society* **28**, 997–1005. ISSN: 1558-0210 (Apr. 2020).
265. Turner, A. & Hayes, S. The Classification of Minor Gait Alterations Using Wearable Sensors and Deep Learning. *IEEE transactions on bio-medical engineering* **66**, 3136–3145. ISSN: 1558-2531 (Nov. 2019).

266. Aboy, M., Minssen, T. & Vayena, E. Navigating the EU AI Act: Implications for Regulated Digital Medical Products. *npj Digital Medicine* **7**, 1–6. ISSN: 2398-6352 (Sept. 2024).

A Appendix

A.1 AI reveals insights into link between CD33 and cognitive impairment in Alzheimer’s Disease

Reprinted with permission from “**Raschka, T.**, Sood, M., Schultz, B., Altay, A., Ebeling, C., Fröhlich, H. (2023). AI reveals insights into link between CD33 and cognitive impairment in Alzheimer’s Disease. *PLOS Computational Biology*, 19(2), e1009894. doi:10.1371/journal.pcbi.1009894”.

Copyright © Raschka, T., *et al.*, 2023

RESEARCH ARTICLE

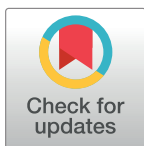
AI reveals insights into link between CD33 and cognitive impairment in Alzheimer's Disease

Tamara Raschka^{1,2,3*}, Meemansa Sood^{1,2}, Bruce Schultz¹, Aybuge Altay^{1,2*}, Christian Ebeling¹, Holger Fröhlich^{1,2*}

1 Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, Germany, **2** Bonn-Aachen International Center for Information Technology (B-IT), University of Bonn, Bonn, Germany, **3** Fraunhofer Center for Machine Learning, Sankt Augustin, Germany

* Current address: Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany

* tamara.raschka@scai.fraunhofer.de (TR); holger.froehlich@scai.fraunhofer.de (HF)



OPEN ACCESS

Citation: Raschka T, Sood M, Schultz B, Altay A, Ebeling C, Fröhlich H (2023) AI reveals insights into link between CD33 and cognitive impairment in Alzheimer's Disease. *PLoS Comput Biol* 19(2): e1009894. <https://doi.org/10.1371/journal.pcbi.1009894>

Editor: Lars Juhl Jensen, University of Copenhagen Faculty of Health and Medical Sciences: Kobenhavns Universitet Sundhedsvidenskabelige Fakultet, DENMARK

Received: February 3, 2023

Accepted: January 18, 2023

Published: February 13, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1009894>

Copyright: © 2023 Raschka et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The source code used to produce the results and analyses

Abstract

Modeling biological mechanisms is a key for disease understanding and drug-target identification. However, formulating quantitative models in the field of Alzheimer's Disease is challenged by a lack of detailed knowledge of relevant biochemical processes. Additionally, fitting differential equation systems usually requires time resolved data and the possibility to perform intervention experiments, which is difficult in neurological disorders. This work addresses these challenges by employing the recently published Variational Autoencoder Modular Bayesian Networks (VAMBN) method, which we here trained on combined clinical and patient level gene expression data while incorporating a disease focused knowledge graph. Our approach, called iVAMBN, resulted in a quantitative model that allowed us to simulate a down-expression of the putative drug target CD33, including potential impact on cognitive impairment and brain pathophysiology. Experimental validation demonstrated a high overlap of molecular mechanism predicted to be altered by CD33 perturbation with cell line data. Altogether, our modeling approach may help to select promising drug targets.

Author summary

For the last 20 years, the field of Alzheimer's Disease (AD) is marked by a series of continuous failures to deliver demonstrably effective medications to patients. One of the reasons for the continuous failure of trials in AD is the lack of understanding of how targeting a certain molecule would affect cognitive impairment in humans. One way to address this issue is the development of quantitative system level models connecting the molecular level with the phenotype. In this paper we propose a novel hybrid Artificial Intelligence (AI) approach, named Integrative Variational Autoencoder Modular Bayesian Networks (iVAMBN), combining clinical and patient level gene expression data while incorporating a disease focused knowledge graph. The model showed connections between various biological mechanisms playing a role in AD and allowed us to simulate a down-expression of the putative drug target CD33. Results showed a significantly increased cognition and predicted perturbation of several biological mechanisms. We experimentally validated these

presented in this manuscript are available on a GitHub repository at <https://github.com/traschka/iVAMBN>. The ROSMAP and Mayo data is available at <https://adknowledgeportal.synapse.org/Explore/Studies/DetailsPage?Study=syn21241740>. The CD33 KO cell line data is available at: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155567>.

Funding: This project has received partial funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115976. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. This work was partially developed in the Fraunhofer Cluster of Excellence "Cognitive Internet Technologies" and partially supported via the Fraunhofer Center for Machine Learning. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

predictions using gene expression data from a knock-out THP-1 monocyte cell line, which confirmed our model predictions up to a very high extent. To our knowledge, we thus developed the first experimentally validated, quantitative, multi-scale model connecting molecular mechanisms with clinical outcomes in the AD field.

Introduction

Alzheimer's Disease (AD) is a neurodegenerative disorder affecting about 50 million people worldwide, resulting in the inability to perform necessary, daily activities before leading to an often early death [1]. Despite decades of research and more than 2000 clinical studies listed on ClinicalTrials.gov, to date there is no cure, and all existing treatments are purely symptomatic [1]. New disease modifying treatments are urgently needed, but require a better mechanistic understanding of the disease.

A common starting point in this context is to map out the existing knowledge landscape about the disease. In the past few decades, a large number of databases have been developed in the bioinformatics community, such as databases for biological pathways (like KEGG [2], PathwayCommons [3], WikiPathways [4], Reactome [5]), drug-target interactions (like OpenTargets [6], Therapeutic Targets Database [7]), disease-gene associations (like DisGeNET [8]) or protein-protein interactions (like STRING [9], IntAct [10]). All these databases simplify the usage of the respective knowledge for algorithms and models, especially in the field of drug target identification. Moreover, none of these databases have been compiled in a disease focused manner. The Biological Expression Language (BEL) provides this opportunity and can be used to represent literature-derived, disease focused knowledge in the form of attributed graphs in a precise manner. For AD a knowledge graph has been published in [11] and represents the manually curated, disease focused mechanistic interplay between genetic variants, proteins, biological processes and pathways described in the literature, enabling the user to computationally query and integrate knowledge graphs into drug target identification algorithms.

One of the interesting molecules in the AD field is CD33, a transmembrane receptor protein expressed primarily in myeloid lineage cells. It has been associated with decreased risk of AD in GWAS studies [12–18] and discussed as a potential therapeutic target, for example via immunotherapy [14]. In an AD mouse model, a knockout of CD33 mitigated amyloid- β clearance and improved cognition [13, 17, 18]. Similarly, a positive effect on amyloid- β phagocytosis could be observed in CD33 knock-out THP-1 macrophages [16]. In humans a correlation between CD33, cognition and amyloid clearance is known, however, the concrete underlying mechanisms are still not well understood. There is an ongoing clinical trial that is testing the effects of a CD33 inhibitor in patients with mild to moderate AD (NCT03822208). Along those lines, the EU-wide PHAGO project (<https://www.phago.eu>) funded via the Innovative Medicines Initiatives aimed to develop tools and methods to study the functioning of CD33 and related pathways in AD in order to facilitate decisions about potential drug development programs.

While graphs are useful for describing the disease focused knowledge landscape about AD, the principal incompleteness of disease focused biological knowledge may result in disagreements to observed data. Moreover, graphs do not allow to produce quantitative insights and predictions. For this purpose ordinary (ODEs) and partial differential equations (PDEs) are frequently used in systems biology and systems medicine, as they are able to describe biological mechanisms in a quantitative way. However, their formulation requires a detailed understanding of biochemical reactions, which in the AD field is only available for specific processes, like

for example amyloid- β aggregation [19, 20]. Moreover, fitting differential equations usually requires time resolved data and the possibility to perform intervention experiments (as knock-downs or stimulation), which is challenged by the fact that cell lines and mouse models in the AD field can most likely only mimic specific aspects of the human disease [21–23].

A principle alternative to differential equation systems are probabilistic graphical models and in particular Bayesian Networks (BNs), which are quantitative as well. However, standard BN implementations require normally or multinomially distributed data, which is not the case in many applications. Furthermore, structure learning of BNs is an NP hard problem, where the number of possible network structures grows super-exponentially with the number of nodes in the network [24]. Hence, modeling higher dimensional data with a BN raises severe concerns regarding structure identifiability.

Altogether, these challenges lead to the fact that the AD field lacks a comprehensive quantitative model of the interplay between relevant molecules and biological processes, including the role of CD33, up to the phenotype level.

In this work, we developed a—to our knowledge—first quantitative, multi-scale model focused on the multitude of mechanisms governing the CD33 molecule. Our model spans a variety of modalities, including gene expression, brain pathophysiology, demographic information and cognition scores. To address the challenges mentioned before, we started with a disease focused knowledge graph reconstruction, which we clustered into modules to significantly reduce dimensionality. In the following we use the term “module” to denote a set of objects grouped together. Subsequently, we relied on our recently published Variational Auto-encoder Modular Bayesian Network (VAMBN) algorithm [25], which is a hybrid Artificial Intelligence (AI) approach combining variational autoencoders [26] with modular Bayesian Networks [27], which is able to model arbitrary statistical distributions. We trained VAMBN on joint clinical and patient level gene expression data while employing a clustered knowledge graph reflecting incomplete prior knowledge about disease mechanisms and their interplay. A simulated knock-down of CD33 and predicted downstream effects could be experimentally validated with gene expression data from a cell line. Overall, we believe that our work helps to move closer towards a systemic and quantitative understanding of the disease, which is the prerequisite for finding urgently needed novel therapeutic options.

Results

In this work, we relied on AD patient data from the Religious Orders Study and Memory and Aging Project (ROSMAP) [28–30] for model training and specificity analysis and the Mayo RNAseq Study (Mayo) [31] for external validation and specificity analysis. The data was retrieved from the RNASeq Harmonization study through the AMP-AD Knowledge Portal. [Table 1](#) shows an overview about the clinical characteristics of the AD patients, which were used for model training and external validation. These patient samples were selected, because for all of them gene expression data from post mortem cerebral cortex tissues was available. We would like to mention at this point that gene regulation and thus gene expression is tissue specific [32]. Available data of other brain regions, also of healthy controls were thus kept separate for specificity analysis. A more detailed description of the used samples in each step of the analysis can be found in [S3 Note](#).

Overview about modeling strategy

[Fig 1](#) shows an overview about our modeling strategy, which we call integrative VAMBN (iVAMBN), combining clinical and patient-level gene expression data with disease focused knowledge graphs. The first step of our workflow compiles an AD focused knowledge graph

Table 1. Patient statistics. Shown are the number of patients, their age in years (with mean and standard deviation), sex, APOE genotype (binary encoding for at least one present E4 allele), MMSE score (with mean and sd) and Braak stage.

	ROSMAP	Mayo
no. patients	221	82
age	87.95 ± 3.38	82.66 ± 7.61
sex		
male	68	33
female	153	49
APOE		
0	138	39
1	83	43
MMSE	13.16 ± 8.38	-
Braak		
1	7	-
2	6	-
3	42	-
4	71	6
5	88	35
6	7	41

<https://doi.org/10.1371/journal.pcbi.1009894.t001>

describing cause and effect relationships between biological processes, genes and pathologies. The generated graph consisted of 383 nodes and 607 edges. The graph was subsequently clustered into modules with the help of the Markov Clustering algorithm [33] to significantly reduce the number of variables for subsequent modeling steps. Markov Clustering was chosen

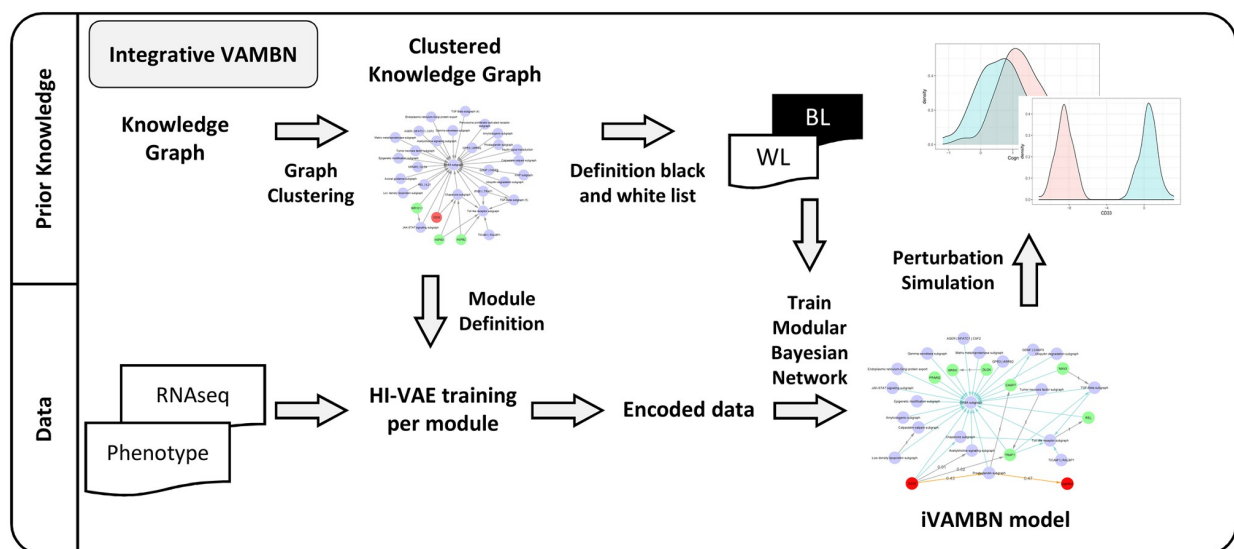


Fig 1. The Integrative VAMBN (iVAMBN) approach. The iVAMBN approach integrates gene expression data, clinical and patho-physiological (phenotype) measures (bottom left) into a joint quantitative, probabilistic graphical model. The method initially uses a knowledge graph (top left) for defining modules and for informing about potential connections between them. In a second step, a representation of each module using a Heterogeneous Incomplete Variational Autoencoder (HI-VAE) is learned. In a third step a modular Bayesian Network between autoencoded modules is learned while taking into account the information derived from the knowledge graph. Finally, the iVAMBN model is used to simulate gene perturbation (top right).

<https://doi.org/10.1371/journal.pcbi.1009894.g001>

above other methods, as an evaluation of different graph clustering algorithms showed best metrics for this approach (see [Methods](#)). Genes within modules were annotated with AD disease mechanisms coming from the NeuroMMSig gene set collection [34].

Using patient-level clinical and gene expression data from post-mortem cerebral cortex tissues, in a second step the VAMBN algorithm was employed to quantitatively model relationships between gene modules as well as phenotype related scores (Mini-Mental State Examination (MMSE), Braak staging) and demographic features based on ROSMAP data. ROSMAP was chosen for training of the algorithm, because of the comparably large number of patients (more than 200) and available MMSE plus Braak scores. VAMBN takes as input patient-level data hierarchically organized into pre-defined modules (here: either gene modules or a phenotype related module including i.e. MMSE plus Braak stages), original features (here: demographic and clinical variables like age, sex, APOE genotype, and brain region) and prior knowledge regarding their possible connections. The output is a probabilistic graphical model describing connections between modules and original features. There is a per-patient score for each module, and each of these scores can be further decoded into feature-level gene expression and phenotype data, respectively.

In the third step of our strategy we evaluated, whether our iVAMBN model could also explain gene expression data from the Mayo study. Notably, at this step we only considered the Braak stage in the phenotype module, because the Mayo study does not report MMSE scores. For that purpose we first re-trained our iVAMBN model on ROSMAP while leaving out MMSE scores and then assessed the marginal log-likelihood of the modified model on the Mayo dataset. We then tested the marginal log-likelihood of the true model against randomly permuted versions of the learned probabilistic graph. This allowed us to assess, in how far the model learned on ROSMAP could explain Mayo data better than expected by pure chance.

For the last step, we used our iVAMBN model trained on ROSMAP to simulate several therapeutic interventions, including a CD33 inhibition. Based on available data, we were able to experimentally validate the predicted effects of a CD33 inhibition using CD33 knockout gene expression data from a THP-1 monocyte cell line. More details about the entire iVAMBN approach can be found in the Methods section of this paper.

In the following we elaborate on the results obtained in each of these different steps, while technical details are provided in the Methods part of this article.

Knowledge graph compilation

As outlined in the previous section, our modeling approach started with the compilation and Markov clustering of a knowledge graph. The Markov clustering resulted in 32 modules, including 4 single gene modules, namely CD33, HSPB2, HSPB3, and MIR101-1. Most of the non-single gene modules comprised only two genes, while others had multiple combinations, like the GABA subgraph module with 289 genes. The exact number of genes clustered together as well as the result of a statistical over-representation analysis (hypergeometric test) using the AD focused gene set collection NeuroMMSig [34] can be found in [S1 Table](#). A complete list of molecules within each module can be found in [S2 Table](#). The modules were considered as nodes of a graph between them, where an edge was set between modules M_1 , M_2 , if in the original knowledge graph there was at least one gene in M_1 and one in M_2 that was connected via a directed path. The resulting (acyclic) module graph is shown in [S1 Fig](#).

Integrative variational autoencoder modular bayesian network model

Integrative VAMBN combines the advantages of Bayesian Networks with the capabilities of variational autoencoders, more specifically Heterogeneous Incomplete Variational Autoencoders

(HI-VAEs) [35]. Briefly, the idea is to initially learn a low dimensional Gaussian representation of features mapping to each of the defined modules. HI-VAEs differ from classical variational autoencoders in the sense that they can be applied to heterogeneous input data of different numerical scales, potentially containing missing values. In a second step a Bayesian Network structure is then learned over the low dimensional representations of modules, resulting in a modular Bayesian Network. More details are presented in the Methods part of this paper and in [25].

We here trained an iVAMBN model using the identified modules (i.e. feature groups in the original data) as—potentially multivariate—nodes of a probabilistic graphical model. Noteworthy exceptions are described in detail in [S1 Note](#). In cases where multiple features map to one and the same module (i.e. the corresponding node / random variable in the probabilistic graphical model is multivariate), our method initially learns a low dimensional representation using a HI-VAE. Second, we learned the Bayesian Network structure connecting these modules. At this stage it is possible to provide information about possible connections between modules given in the knowledge derived module graph ([S1 Fig](#)). We tested three different strategies to incorporate the information provided in the module graph:

- *completely data driven*: the entire Bayesian Network was only learned from data,
- *knowledge informed*: the module graph was either used to only initialize Bayesian Network structure learning, to enforce / white list the existence of specific edges, or used for a combination of both, and
- *completely knowledge driven*: strictly constrain edges between modules to those provided via the module graph, and additionally learned ones are only allowed to connect cognition scores, patho-physiological stages, and demographic features. All other possible edges are black listed, i.e. not allowed.

A systematic comparison of these strategies via a cross-validation yielded a better performance of the second strategy (knowledge informed), in which we used the module graph to white list edges and to initialize a greedy hill climbing based structure learning, see details in [Methods](#) Section and [S2 Note](#). That means, iVAMBN was allowed to add additional edges, if the data provided according evidence.

We repeated the knowledge informed modular Bayesian Network learning 1000 times on random bootstrap sub-samples of the data drawn with replacement, hence allowing to quantify the statistical confidence of each inferred edge. The results of this analysis can be found in [S3 Table](#).

In the following we only focus on the 130 edges that were found in at least 40% of the 1000 modular Bayesian Network reconstructions ([Fig 2](#)). Notably, this threshold was only chosen for better visualization purposes and to limit the subsequent discussion. Edges with lower bootstrap probability might also exist in reality despite lower statistical confidence. Nodes corresponding to sex, APOE status, and brain region were not connected to any other nodes with sufficient statistical confidence, meaning that these features might have no impact on the rest of the network. Nodes with only outgoing edges in the network (i.e. source nodes) were: the years of education, the age, and the single gene NAV3. The GABA subgraph (containing more than 280 genes) and the phenotype module were leaf nodes, meaning they had no outgoing edges. Only patient age had a direct influence on CD33. CD33 had eight directly influenced molecular mechanisms: the GABA subgraph, the Amyloidogenic subgraph (containing genes SRC and APBA2), the Acetylcholine signaling subgraph (containing genes ACHE and PRNP), the Prostaglandin subgraph, and the Chaperone subgraph (containing genes HSPB6, CXCL8,

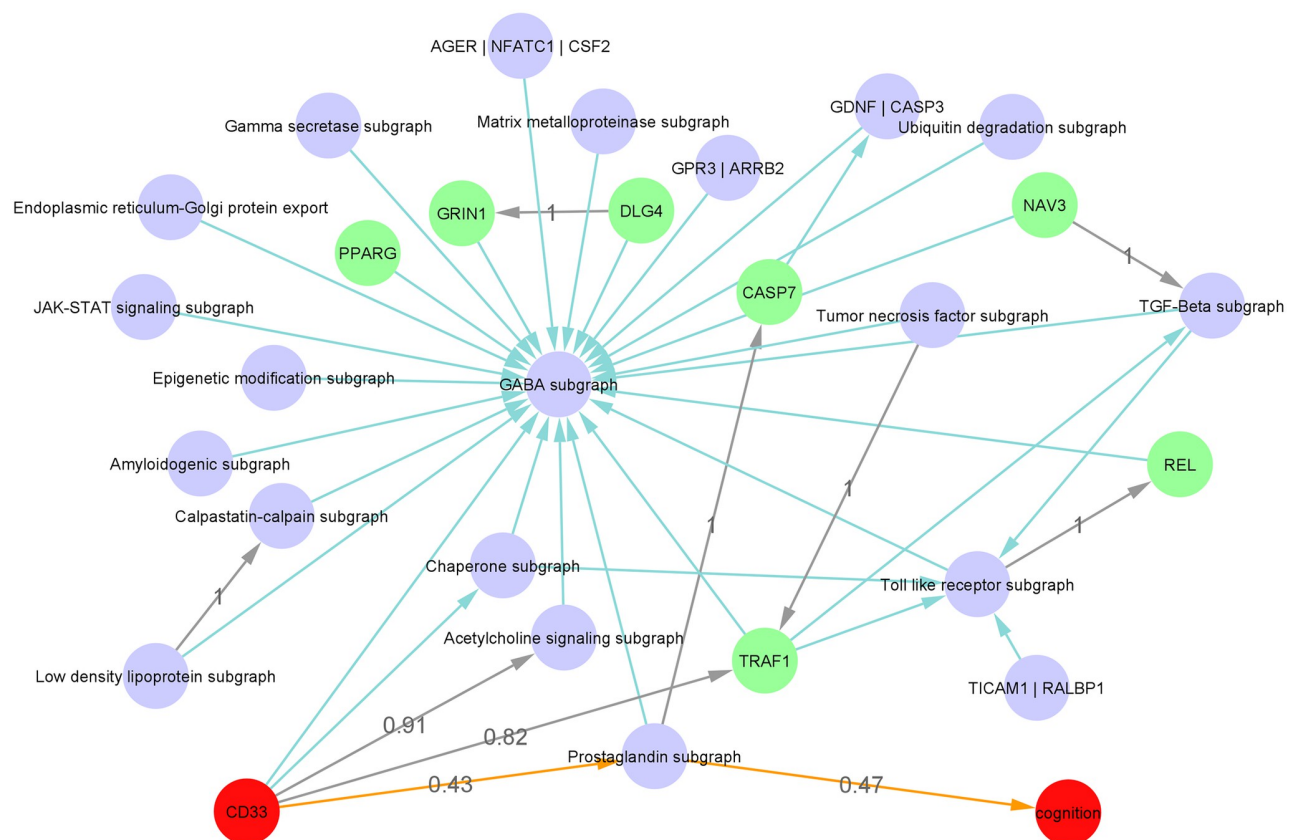


Fig 2. Network representation of iVAMBN model for ROSMAP data. Shown are the learned (grey) and knowledge-derived (green) edges between gene modules (purple nodes), single gene modules (green) and CD33 and phenotype module (red). All these edges appeared with bootstrap frequency > 0.4. The newly inferred shortest path between CD33 and phenotype is displayed in orange. Other edges with bootstrap frequency > 0.4 have been removed for visualization purposes, except for those six edges which were trained with a bootstrap confidence of 1.

<https://doi.org/10.1371/journal.pcbi.1009894.g002>

and CCR2). Also, the single gene module, TRAF1, was a child of CD33. Altogether, CD33 had a predicted causal influence on every node, except for the source nodes.

Model reveals path between CD33 and disease phenotype. As shown in Fig 2 the shortest path between CD33 and the disease phenotype was observed through the Prostaglandin subgraph. All the edges from this connection were newly learned from data, meaning that they had not previously been identified in the knowledge graph. Nevertheless, these correlations have been previously reported in the literature: Prostaglandines are eicosanoides, which were found to play a role in memory learning and neuroinflammation [36, 37]. A major producer is activated microglia, which itself is activated through amyloid- β and produces inflammatory cytokines [38]. Currently, microglia and their effects on AD is a major focus of research [39, 40]. Also, PGD2, a prostaglandin mainly synthesized in neurons, was previously found to be upregulated in AD patients [41]. Pairwise correlation plots between the genes of the prostaglandin pathway and CD33 or phenotype can be found in S2 Fig.

In total, 130 of the 162 edges of the bootstrapped iVAMBN model were newly learned from the data and had not been previously identified within the literature derived knowledge graph. Out of these 130 edges, six edges had a bootstrap confidence of 100%, meaning that they were learned consistently from 1000 random sub-samples of the data. A list of these edges can be found in Table 2.

Table 2. Consistently newly learned edges in iVAMBN model. All edges were found in each of 1000 network reconstructions from randomly subsampled data.

from	to
DLG4	GRIN1
Tumor necrosis factor subgraph	TRAF1
Toll like receptor subgraph	REL
Low density lipoprotein subgraph	Calpastatin-calpain subgraph
Prostaglandin subgraph	CASP7
NAV3	TGF-Beta subgraph

<https://doi.org/10.1371/journal.pcbi.1009894.t002>

These high confidence edges demonstrated strong pairwise correlations between connected modules. NAV3, for example, had a strong negative correlation with MAVS, a member of the TGF-Beta subgraph module (Fig 3 left). In contrast to that SRSF10 and CREB1, members of the Low density lipoprotein subgraph and Calpastatin-calpain subgraph modules, were strongly positive correlated (Fig 3 right).

Although no direct correlation between NAV3 and MAVS is known, their effects are both linked to AD. NAV3, which is predominantly expressed in the nervous system, is increased in AD patients [42], while MAVS encodes a gene that is needed for the expression of beta interferon and thus contributes to antiviral innate immunity and may protect the cells from apoptosis [43]. Together with the strong negative correlation seen in the data, one can hypothesize that the increased level of NAV3 in AD leads to a decreased level of MAVS, which elevates apoptosis of the cells.

The strong positive correlation between SRSF10 and CREB1 linked the Low density lipoprotein (LDL) and Calpastatin-calpain subgraphs. LDL is a major APOE receptor, which is the strongest genetic factor for AD, where different alleles are either risk or protective alleles [44]. APOE is also linked to amyloid- β , whose production is increased with elevated activity of calpain due to the decreased levels of calpastatin. Calpastatin is also linked to synaptic dysfunction and to the tau pathology of AD [45, 46]. Tau is another protein that accumulates in the brains of AD patients. The exact underlying mechanisms here are still unknown, but regulatory mechanisms of calpain are highly influenced by Calcium (Ca^{2+}) influx and increased

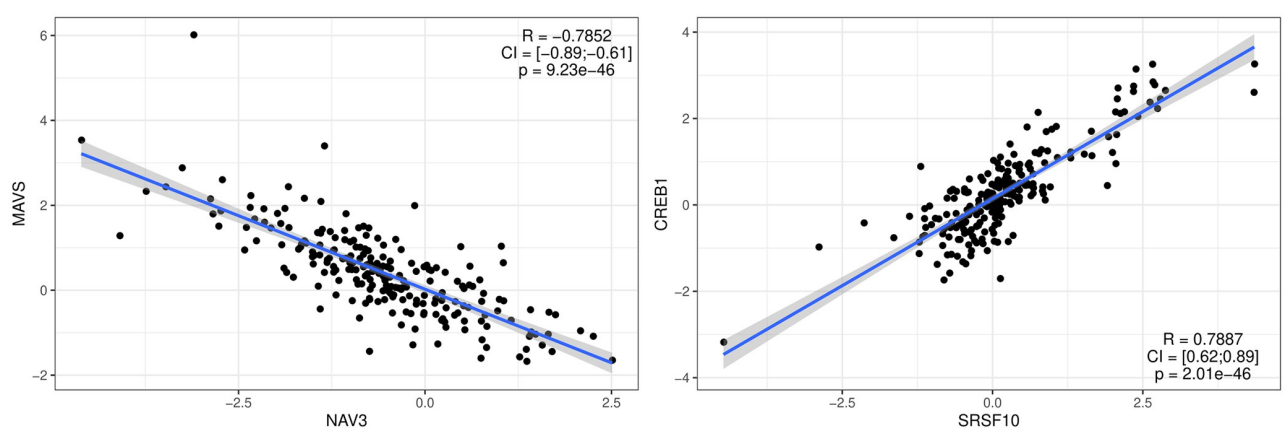


Fig 3. Quantitative relationships learned by iVAMBN. Each correlation (R) is shown along with its confidence interval (CI) and multiple testing adjusted p-value. Left: Correlation of NAV3 with TGF-Beta subgraph module member MAVS. Right: Correlation of Low density lipoprotein subgraph module member SRSF10 with CREB1, a member of the Calpastatin-calpain subgraph module. Further plots can be found in S2 and S3 Figs.

<https://doi.org/10.1371/journal.pcbi.1009894.g003>

intracellular calcium levels are a main reason for the loss of neuronal function in AD [45–47]. Changes in the Calpastatin-calpain mechanism may therefore also lead to reduced amyloid- β deposition.

External validation of iVAMBN model

We assessed the ability of the model to explain normalized gene expression data from an independent study, Mayo. Notably, all gene expression data used in this analysis was mapable to the same brain region, namely the cerebral cortex, via the Uber-anatomy ontology (UBERON) [48]. However, Mayo does not contain MMSE scores. Therefore, we first trained a modified version of our iVAMBN model on ROSMAP, which only contained the Braak score in the phenotype module, but otherwise had the edges shown in Fig 2. The full list of edges of this model together with their corresponding bootstrap confidences can be found in S3 Table. We then explored the marginal log-likelihood $\log p(\text{data} \mid \text{graph})$ of the model on the Mayo dataset and subtracted the marginal log-likelihood obtained by 1000 random permutations of the network (Fig 4), resulting in an empirical p-value. This showed that our model could explain Mayo gene expression data significantly better than randomly permuted networks ($p = 0.035$) despite the clinical differences between patients in both studies shown in Table 1.

Moreover, we evaluated the ability of our iVAMBN to predict the activity score of the prostaglandin module in the Mayo study. The prostaglandin module was chosen, because prostaglandins play a role in neuroinflammation, which is a hallmark of the disease phenotype. Moreover, the prostaglandin module was directly connected to the clinical/pathological phenotype module, in which Braak stages, however, significantly differed between Mayo and

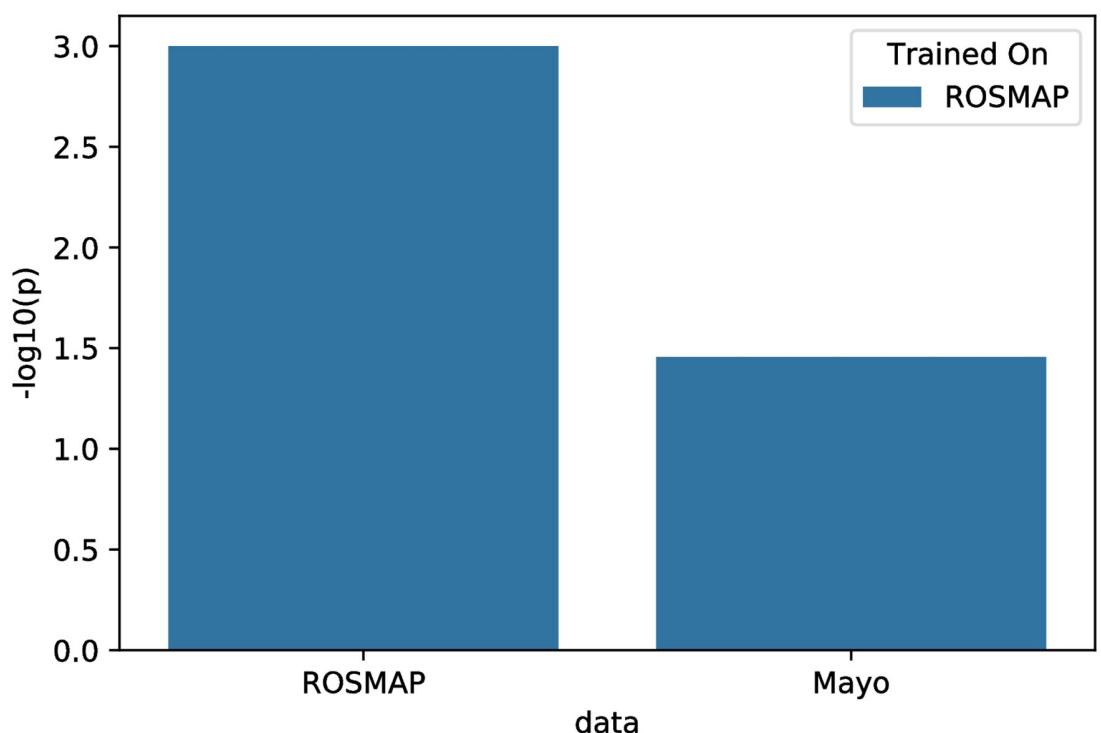


Fig 4. External model validation. Statistical significance $-\log_{10}(p)$ value of the marginal log-likelihood of the model when evaluated on the training data (ROSMAP) and external validation data (Mayo).

<https://doi.org/10.1371/journal.pcbi.1009894.g004>

ROSMAP studies. We thus regarded the activity score of the prostaglandin module as a relevant and sufficiently comparable surrogate endpoint between Mayo and ROSMAP studies. We used our iVAMBN model trained on ROSMAP to predict the activity score of prostaglandin module activity in Mayo by feeding data from genes outside the prostaglandin module to the model expression. We observed a highly significant Pearson correlation between true and predicted values in the external validation dataset ($r = 0.69$, 95% CI: [0.56;0.79]). Hence, we concluded that our iVAMBN model was predictive for the chosen endpoint.

Finally, we trained a separate iVAMBN model on Mayo data and explored the overlap with the ROSMAP model at different thresholds of the bootstrap confidence (S4 Fig). At the previously chosen 40% threshold the overlap of the newly learned edges contained in the iVAMBN models trained on ROSMAP and Mayo was statistically significant, even if edge directions were considered (hypergeometric test, $p < 1e - 38$).

Specificity and sensitivity of iVAMBN model

Specificity to brain region. We tested the ability of the model to explain normalized gene expression data from other brain regions. Therefore, we trained multiple additional iVAMBN models on patient samples belonging to the posterior cingulate cortex, the dorsolateral prefrontal cortex, and the head of caudate nucleus from the ROSMAP study, as well as on samples from the temporal cortex and the cerebellum from the Mayo study. We then investigated the overlap of each of these additional iVAMBN models with our primary one akin to the external validation described in the previous Section. Among non-cortical brain regions, the largest and statistically significant overlap on graph level was found with an iVAMBN model trained with samples from the head of caudate nucleus ($\sim 36\%$ considering edge directions). The lowest (still statistically significant) overlap was found with the cerebellum ($\sim 31\%$ considering edge directions). The primary iVAMBN model for all datasets was able to predict the activity score of the prostaglandin module, but the prediction performance was clearly lower in non-cortical brain regions (see results in Table B in S3 Note). Altogether our results thus suggest that our primary iVAMBN model is focused on cortical brain regions.

Disease specificity. Similar analyses were done for iVAMBN models trained on available healthy control samples from the posterior cingulate cortex, the dorsolateral prefrontal cortex, and the head of caudate nucleus from the ROSMAP study. The graph structures still demonstrated a significant overlap with our primary iVAMBN model but were considerably lower, see Table C in S3 Note. This suggests that our primary iVAMBN model is AD focused.

Sensitivity to knowledge graph. Finally, we explored, how sensitive our primary iVAMBN model was to the knowledge graph. For that purpose, we randomly shuffled all edges of the original knowledge graph, re-clustered this permuted graph, and re-trained a complete iVAMBN model. The iVAMBN model trained on the permuted graph demonstrated a significantly lower marginal log-likelihood $p(\text{data} \mid \text{model})$ compared to the primary iVAMBN model ($p = 4.14E - 24$), see details in Supplements (Fig B in S3 Note). Hence, we concluded that our primary iVAMBN model was sensitive to the knowledge graph structure.

CD33 down-expression simulation

To understand the potential systemic consequences of a therapeutic intervention into CD33 we used our primary iVAMBN model to simulate a down-regulation. This was achieved by a counterfactual down-expression (here: 9-fold) of CD33 in every patient (Fig 5 (top left)). Due to the fact that iVAMBN is a quantitative model, associated downstream consequences on biological mechanisms and phenotype could be predicted in every patient (see examples in Fig 5).

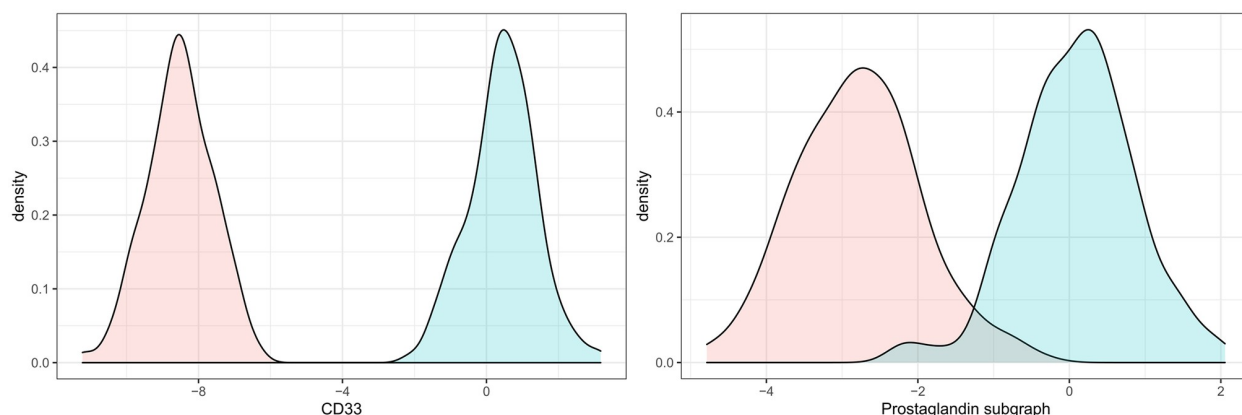


Fig 5. Module distributions in original and simulated CD33 down-expression. The blue curve describes the original distribution, while the red one describes the CD33 down-expression scenario. CD33 down-expression simulation (left) results in lower scores of the prostaglandin pathway module (right).

<https://doi.org/10.1371/journal.pcbi.1009894.g005>

CD33 down-expression simulation (left) results in higher activity scores of the prostaglandin pathway module (right).

In addition, iVAMBN predicted a significant increase of MMSE scores ($p < 0.001$, Fig 6 (left)), and also a significant decrease of Braak stages ($p < 0.001$, Fig 6 (right)). That means patients are not only predicted to improve the specific cognitive abilities tested by MMSE, but are also predicted to improve brain pathophysiology.

CD33 down-expression reveals significant changes in many mechanisms. Our iVAMBN model predicted significant effects on gene expression of 28 mechanisms and individual genes, respectively (Table 3). Significant changes were, for example, predicted for the genes CASP7 and TRAF7, and the prostaglandin and calpastatin-calpain mechanisms. But also the amyloidogenic mechanism is significantly differential expressed in a CD33 knock-down scenario.

Decreased expression of the amyloidogenic mechanism will thus result in patients with less amyloid- β deposition. While this connection of the amyloidogenic mechanism and AD is clear, others need to be further explored.

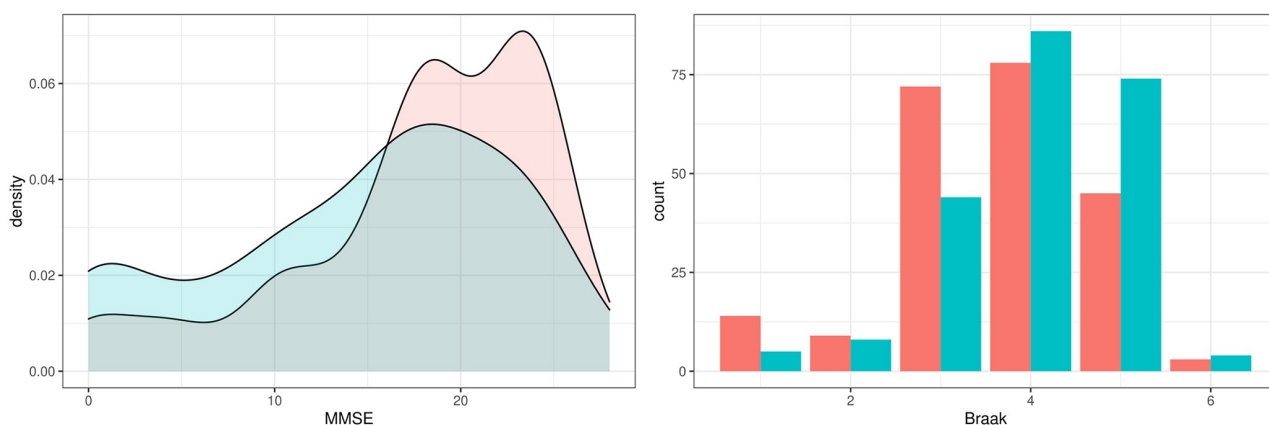


Fig 6. Predicted changes on phenotype (MMSE and Braak stages) as a consequence of CD33 down-expression. Distribution of MMSE and Braak stages in CD33 original (blue) and down-expressed (red) patients shows a significant improvement of scores and thus cognition as well as brain pathophysiology.

<https://doi.org/10.1371/journal.pcbi.1009894.g006>

Table 3. Statistical significance of gene modules. The table shows results of a global test [49], assessing the differential gene set expression of each gene module between WT and down-expression/KO of CD33. P-values of the test within simulated scenario, as well as, p-values from cell line KO are reported and corrected for multiple testing using the Benjamini-Hochberg method. The agreement of both tests is described in the last column, meaning if both tests are either significant or non-significant (+) or if they don't show same direction of significance (-). For GRIN1 no p-value could be computed, as that gene is not present in the cell line data.

Gene module	p-value simulated KD	p-value cell line KO	agreement significance
GABA subgraph	2.75e-04	3.60e-15	+
Toll like receptor subgraph	1.05e-26	1.05e-13	+
Prostaglandin subgraph	6.99e-109	1.02e-09	+
TGF-Beta subgraph	0.592	8.79e-11	-
Calpastatin-calpain subgraph	3.14e-91	5.41e-09	+
JAK-STAT signaling subgraph	0.454	2.91e-11	-
AGER / NFATC1 / CSF2	5.78e-41	0.0129	+
Chaperone subgraph	2.84e-75	2.02e-09	+
REL	4.45e-18	9.96e-11	+
Ubiquitin degradation subgraph	5.15e-20	1.06e-06	+
GRIN1	1.92e-132	NA	
PPARG	2.20e-04	1.78e-03	+
GDNF / CASP3	1.06e-17	2.98e-11	+
Gamma secretase subgraph	4.36e-10	1.93e-03	+
Epigenetic modification subgraph	6.90e-58	7.64e-03	+
TICAM1 / RALBP1	1.46e-16	0.0561	-
Amyloidogenic subgraph	4.54e-69	9.11e-10	+
Tumor necrosis factor subgraph	0.0997	0.769	+
Acetylcholine signaling subgraph	6.74e-04	0.337	-
Matrix metalloproteinase subgraph	0.0708	2.74e-10	-
NAV3	0.176	3.81e-07	-
TRAF1	1.66e-95	2.26e-08	+
CASP7	1.75e-138	0.151	-
GPR3 / ARRB2	4.87e-04	8.02e-04	+
Endoplasmic reticulum-Golgi protein export	5.19e-29	1.78e-11	+
Low density lipoprotein subgraph	0.891	8.11e-06	-
DLG4	5.85e-93	3.44e-07	+
CD33	3.33e-307	8.06e-08	+

<https://doi.org/10.1371/journal.pcbi.1009894.t003>

The link between Calpastatin-calpain mechanism and AD was already described earlier. The key aspect is its negative influence on amyloid- β deposition. PGD2, a prostaglandin mainly synthesized in neurons, was previously found to be upregulated in AD patients [41]. Prostaglandins are eicosanoids, which were found to play a role in memory learning and neuroinflammation [36, 37]. A major producer is activated microglia, which itself is activated through amyloid- β and produces inflammatory cytokines [38]. Currently, microglia and their effects on AD is a major focus in the field of research [39, 40]. Again, down-expression of the prostaglandin may result in reduced amyloid- β deposition. Altogether, the vast majority of significantly differential expressed gene sets was highly linked to AD through the amyloid- β cascade.

Experimental validation with cell line data. We checked whether our iVAMBN based predictions experimentally agreed with cell line gene expression data, specifically reflecting wild type (WT) and CD33 knock-out (KO). Our analysis (see details in [Methods](#) part) revealed significant changes of 23 AD associated mechanisms and genes in KO versus WT. Interestingly, 19 out of these 23 mechanisms overlapped with those predicted by iVAMBN ([Table 3](#)).

Likewise, iVAMBN predicted significant changes of 22 genes and gene sets, respectively, out of which only 3 were false positives at a false discovery rate threshold of 5%. Notably one of the false positive predictions (TICAM1 / RALBP1) had an adjusted p-value of 5.6% in the experimental data.

Overall, we thus observed a high degree of overlap between the dysregulated mechanisms and those predicted by the iVAMBN model, indicating that our model aligns well with the cell line data.

Simulation of the perturbation of other candidate targets. For comparison reasons, we further simulated the effect on the phenotype of a 9-fold up- or down-regulation of all other genes in our model, which showed a directed path to the phenotype module. Genes belonging to modules which were not an ancestor of the phenotype module were excluded, because they could not have any effect on the phenotype according to our model. We simulated for each candidate target an up- as well as a down-regulation.

The simulated dys-regulations showed that none of the candidate targets had a predicted effect on the phenotype stronger than CD33 (S5 Fig). Only TRAF6 and TGFB3 down-regulation as well as up-regulation of APBA2, TRAF5 and SALL1 were predicted to increase the mean MMSE score by more than two points, compared to a predicted increase by almost five points via CD33 perturbation.

APBA2 is known to interact with APP and therefore plays a role in the amyloidogenic pathway [50, 51]. TRAF6 was identified in multiple experiments as target of miR-146a which is a key regulator of innate immunity that is up regulated in AD pathology affected brain regions and might also has an effect on amyloid- β metabolism [52]. It was found that treatment with a miR-146a agomir inhibits TRAF6 expression and reduced the cognitive impairment in AD mice [53].

Discussion

The here presented work is the first to demonstrate, to our knowledge, that one can integrate gene expression and clinical data together with qualitative knowledge about cause-and-effect relationships into a quantitative, system medical model of AD. This was achieved via an AI based method, which we combined with a knowledge graph representation of AD. We could show that a simulated CD33 down-expression agrees well with experimental gene expression KO data from a THP-1 cell line. Overall, our model could thus help to understand and quantify intervention effects on a multi-scale biological system level and thus aid the identification of novel therapeutic targets, which are urgently needed in the AD field.

Our model predicted that CD33 down-regulation would yield a significant effect on cognition (MMSE) and brain pathophysiology (Braak scores) through the prostaglandin pathway. Although the role of prostaglandins is known to play a role in memory, learning and neuroinflammation [36, 37], the exact mechanism by which cognition is affected remains unknown, but seems to be coupled to amyloid- β deposition through microglia. In AD mice, a knockout of CD33 mitigated amyloid- β clearance and improved cognition [17, 18]. A positive effect on amyloid- β phagocytosis could also be observed in CD33 KO THP-1 macrophages [16].

Despite the evidence for a positive effect on cognition, we should mention that CD33 as a possible drug target has possible caveats that have been discussed in the literature [14]: i) It is not clear whether the genetic association of CD33 to AD is causal or just due to linkage disequilibrium with the true causal variant. ii) It is so far not entirely clear, how to therapeutically manipulate the expression level of CD33 in an optimal manner. iii) There might be safety issues due to the fact that CD33 is important for inhibiting immune responses and mediating self-tolerance. Systemic CD33 inhibition could potentially induce inflammatory autoimmune

diseases. We therefore see the investigation of CD33 conducted in this paper more as a showcase for our iVAMBN approach rather than making any specific recommendation regarding the therapeutic value of CD33. Integrating known side effects of approved drugs targeting specific proteins in our model's graph structure could provide hints on possible side effects and is an interesting point for further research.

Altogether we see the impact of our work two-fold: first, we have introduced a novel multi-scale, quantitative modeling approach (iVAMBN), which is widely applicable in systems medicine, specifically in situations, where only a partial mechanistic understanding of biological phenomena is given. Secondly, our developed model can be further explored by the AD field and could aid a better understanding of the disease as well as identification of novel therapeutic options.

Methods

AD knowledge graph

A major part of this study is a BEL (<https://bel.bio>) encoded, knowledge graph, which was initially compiled via text mining and later on manually curated via literature. In general, the BEL language helps to build a computer-process-able cause-and-effect relationship model. Each BEL statement consists of a subject and an object, connected through a relation. Subjects and objects could be many different entities, like genes, proteins or RNA, but also biological processes, pathologies or even chemicals. Therefore, the relations have many different facets, as well. These could be relations like *increases*, *decreases* or *association*, describing the interaction between subject and object. But there are also relationships describing something like a membership of subject and object, for example *hasComponent* and *isA*. The BEL model used here, is an enriched version of the AD cause-and-effect relationship model published in [11] and can be found in the github repository. The enrichment was done around the two genes CD33 and TREM2, such that detailed knowledge about these two genes was gathered in the context of AD.

A filtering step was necessary, in order to get only entities measured in the gene expression data. In this case only gene and protein entities from the knowledge graph can be used. Additionally, the knowledge graph was filtered for only causal interactions, such as *increases*, *decreases*, or *regulates*, resulting in a network with 431 nodes and 673 edges. From that we only took the largest connected component to reduce the dimensionality. Hence, the used graph during our study consisted of 383 nodes and 607 edges, in which any two nodes were connected through some path.

Clustering of filtered knowledge graph. One of the key aspects of iVAMBN is grouping of input features (genes, pathophysiological and clinical features) into modules in order to allow for a statistically stable identification of a Bayesian Network structure in a subsequent step. For identifying modules of genes we clustered the knowledge graph with the help of different graph clustering algorithms:

- the Markov Cluster algorithm [33, 54] implemented in the *MCL* package in R [55].
- edge betweenness [56] community detection implemented in the R package *igraph* [57]
- infomap [58] community finding method implemented in the R package *igraph* [57]

After clustering, genes being part of a single cluster were assigned to a corresponding module. Genes being not clustered but only connected to one cluster, were merged into that cluster. Genes being connected to multiple clusters were kept as single gene modules (modules consisting of a single feature) for further analysis. We selected the best clustering algorithm according

to multiple metrics described in [59] including internal density, number of edges inside clusters, average degree, expansion, cut ratio, conductance, and norm cut. Based on these metrics the average ranking of each graph clustering algorithm was computed with the rational in mind, that each cluster should have a high internal density and sparse connections across clusters. This resulted in choosing the markov clustering algorithm for further analyses. The metrics for each clustering algorithm can be found in [S4 Table](#).

Annotation of modules with AD disease mechanisms. For each module, an over-representation analysis for AD associated disease mechanisms was conducted. AD associated mechanisms were retrieved from the NeuroMMSig database [34]. For that purpose, the *enricher* function from the *clusterProfiler* package in R was used, which allows to use user-defined gene set annotations for a hypergeometric test [60]. We annotated each module with the most significant NeuroMMSig gene set after multiple testing correction via control of false discovery rate (Benjamini-Hochberg method).

Gene expression data analysis

RNAseq data from several observational clinical studies, as well as RNAseq data from a cell line knockout experiment, were used in this work. The patient data were from i) the Religious Orders Study and Memory and Aging Project (ROSMAP) [28–30], and ii) the Mayo RNAseq Study (Mayo) [31]. The last one contains two separate datasets referring to separate brain regions, namely cerebellum and temporal cortex, while ROSMAP contains samples from the dorsolateral prefrontal cortex, head of caudate nucleus, and posterior cingulate cortex. Both studies were accessed through the AMP-AD Knowledge Portal at Synapse using the data deposited in the RNAseq Harmonization Study.

Patient samples were selected based on different criteria regarding the task they were used for:

1. For training of the primary iVAMBN model only AD samples from the first ROSMAP batch were used, resulting in 221 samples from dorsolateral prefrontal cortex during the training phase.
2. For external validation we used samples of the temporal cortex of AD patients in Mayo.
3. For specificity and sensitivity analysis, samples from other batches of the ROSMAP data were used, as well as the Mayo cohort. In this step, the samples were first separated by their diagnosis, AD or healthy control, and additionally separated by their brain region, resulting in three AD and three healthy control subject subsets for ROSMAP (dorsolateral prefrontal cortex, head of caudate nucleus, and posterior cingulate cortex) and two AD subsets for Mayo (cerebellum and temporal cortex).

The Mayo study does not report Braak scores for healthy control subjects, which made us discard them from the specificity analysis, as there is no phenotype information available for them. More information about the number of samples, per brain region and study, used in each analysis step can be found in Table A in [S3 Note](#).

The used data are gene counts provided as gene count matrices that had been generated using STAR [61]. Gene counts were normalized to log counts per Million (logCPMs) and counts from AD patients were scaled against the healthy control data within each study. That means for each AD sample and gene the corresponding mean expression value of the same gene in cognitively normal subjects was subtracted. Subsequently we divided the values by the standard deviation of the gene in healthy controls. That means raw expression values were converted into abnormality scores. For making the expression data across studies comparable,

a batch correction with ComBat [62] was applied to the scaled AD data. This normalized, scaled, and batch corrected data was then used for further analysis steps.

The cell line RNAseq data used during this study is from a THP-1 monocyte cell line with two different genetic backgrounds and two treatments. It can be found under GEO accession GSE155567. A sample could have either wild-type CD33 or a knocked out CD33 gene, plus either a control vector or a SHP-1 knock-down vector, resulting in four different conditions: i) *wild-type with control*, ii) *wild-type with SHP-1 knock-down vector*, iii) *CD33 knockout with control vector*, and iv) *CD33 knockout with SHP-1 knock-down vector*. There were 6 biological replicates per condition. Within the here presented study, only samples containing the control vector were used, resulting in twelve used samples. Therefore samples from condition 1 were called as *wild-type (WT)* samples and samples from condition 3 as *knockout (KO)* samples. Reads were aligned with STAR and gene counts were generated via the *featureCounts* function of the *Rsubread* package [63]. More details about the data can be found in [16] and under GEO accession GSE155567.

Variational Autoencoders (VAE)

Variational autoencoders [26] are one of the most frequently used type of unsupervised neural network techniques. They can be interpreted as a special type of probabilistic graphical model, which has the form $Z \rightarrow X$, where Z is a latent, usually multivariate standard Gaussian, and X a multivariate random variable describing the input data. Moreover, for any sample (x, z) , we have $p(x | z) = N(\mu(z), \sigma(z))$. One of the key ideas behind VAEs is to variationally approximate

$$\log q(z|x) = \log N(z | \mu(x), \sigma(x)) \quad (1)$$

This means that $\mu(x)$ and $\sigma(x)$ are the multivariate mean and standard deviation of the approximate posterior $q(z | x)$ and are outputs of a multi-layer perceptron neural network (the encoder) that is trained to minimize for each data point x the criterion

$$\log(x) \geq \frac{1}{2} \sum_{j=1}^D \left(1 + \log \sigma_j(x)^2 - \mu_j(x)^2 - \sigma_j(x)^2 \right) + \frac{1}{L} \sum_l \log p(x|z^{(l)}) \quad (2)$$

Here the index j runs over the D dimensions of the input x , and $z = \mu(x) + \sigma(x) \odot \epsilon^{(l)}$ with $\epsilon^{(l)} \sim N(0, I)$ being the l th random sample drawn from a standard multivariate Gaussian, and $\odot$ denotes an element-wise multiplication. Notably, the right summand corresponds to the reconstruction error of data point x by the model, whereas the first term imposes a regularization. We refer to [26] for more details.

Heterogeneous Incomplete Variational Autoencoders (HI-VAE)

Variational autoencoders were originally developed for homogeneous, continuous data. However, in our case variables grouped into the phenotype module do not fulfill this assumption, because Braak stages and MMSE scores are discrete ordinal. In agreement to our earlier work [25] we thus employed the HI-VAE [35], which is an extension of variational autoencoders and allows for various heterogeneous data types, even within the same module. More specifically, the authors suggest to parameterize the decoder distribution as

$$p(x_j | z) = p(x_j | \gamma_j = h_j(z)) \quad (3)$$

where $h_j(\cdot)$ is a function learned by the neural network, and γ_j accordingly models data modality specific parameters. For example, for real-valued data we have $\gamma_j = (\mu(z), \sigma_j(z)^2)$, while for ordinal discrete data we use a thermometer encoding, where the probability of each ordinal

category can be computed as

$$p(x_j = r \mid \gamma_j) = p(x_j \leq r \mid \gamma_j) - p(x_j \leq r - 1 \mid \gamma_j) \quad (4)$$

with

$$p(x_j \leq r \mid z) = \frac{1}{1 + \exp(-(\theta_j(z) - h_j(z)))} \quad (5)$$

The thresholds $\theta_j(z)$ divide the real line into R regions, and $h_j(z)$ indicates, in which region z falls. The data modality specific parameters are thus $\gamma_j = \{h_j(z), \theta_1(z), \dots, \theta_{R-1}(z)\}$ and are modeled as output of a feed forward neural network.

According to [35] we use batch normalization to account for differences in numerical ranges between different data modalities.

For multi-modal data and in particular discrete data a single Gaussian distribution may not be a sufficiently rich representation in latent space. Hence, the authors propose to replace the standard Gaussian prior distribution imposed for z in VAEs by a Gaussian mixture prior with K components:

$$s \sim \text{Categorical}(\pi) \quad (6)$$

$$z \mid s \sim N(\mu(s), I_K) \quad (7)$$

where $\pi_k = \frac{1}{K}$ for $k = 1, 2, \dots, K$ and s is a one-hot vector encoding of the mixture component. We evaluated different choices of K using a 3-fold cross-validation, while employing the reconstruction error $\frac{1}{L} \sum_l \log p(x|z^{(l)})$ as an objective. In conclusion it turned out that $K = 1$ component was an optimal choice for all modules in our iVAMBN model.

Modular bayesian networks

Let $X = (X_v)_{v \in V}$ be a set of random variables indexed by nodes V in a directed acyclic graph (DAG) $G = (V, E)$. In our case each of these nodes corresponds either to lower dimensional embedding of a group of variables (i.e. module) in the original data, or to an original features (e.g. biological sex) in the dataset. According to the definition of a Bayesian Network (BN), the joint distribution $p(X_1, X_2, \dots, X_n)$ factorizes according to

$$p(X_1, X_2, \dots, X_n) = \prod_{v \in V} p(X_v \mid X_{pa(v)}) \quad (8)$$

where $pa(v)$ denotes the parent set of node v [27]. In our case random variables follow either a Gaussian or a multinomial distribution, i.e. the BN is hybrid. Notably, no discrete random variable was allowed to be a child of a Gaussian one.

Since the BN in our case is defined over low dimensional representations of groups of variables, we call the structure Modular Bayesian Network (MBN). Notably, a MBN is a special instance of a hierarchical BN over a structured input domain [64–67].

A typical assumption in (M)BNs is that the set of parameters $(\theta_v)_{v \in V}$ associated to nodes V are statistically independent. For a Gaussian node v parameters can thus be estimated by fitting a linear regression function with parents of v being predictor variables [27]. Similarly, for a discrete node $\tilde{v}$ having only discrete parents, parameters can be estimated by counting relative frequencies of variable configurations, resulting into a conditional probability table.

Quantitative modeling across biological scales via iVAMBN

Model training. The here presented *Integrative Variational Autoencoder Modular Bayesian Network (iVAMBN)* approach (Fig 1), integrates different biological scales together with a knowledge graph into the previously published Variational Autoencoder Modular Bayesian Network (VAMBN) approach [25]. More precisely, there are four steps to build an iVAMBN model: i) Definition of modules of variables, ii) Training of a HI-VAE for each module, iii) Definition of logical constraints for possible edges in the MBN, and iv) Structure and parameter learning of the MBN using encoded values for each module. These four steps result from the fact that HI-VAEs (as well as any other variants of variational autoencoders) themselves can be interpreted as specific types of BNs and thus the overall log-likelihood of an iVAMBN model can be decomposed accordingly. That means the overall iVAMBN model can be interpreted as a special type of Bayesian Network, see [25] for details.

The four model building steps were followed in the application of the iVAMBN approach in this work as well. The modules of variables were mainly defined through the previously explained Markov clustering of the knowledge graph, plus an additional module summarizing MMSE (Mini-Mental State Examination) and Braak stages into one *phenotype* module. MMSE measures cognitive impairment by testing the orientation in time and space, recall, language, and attention, while Braak stages refer to the degree of biological brain pathology [68]. Some non-assigned genes, were directly treated as nodes in the MBN construction and thus also called gene modules. The same was done for demographic features, like sex, age, years of education and the APOE genotype.

For training the HI-VAEs for each module a hyperparameter optimization (grid search) was implemented over learning rate (learning rate $\in \{0.001, 0.01\}$) and minibatch size (minibatch size $\in \{16, 32\}$) as in [25]. Each parameter combination was evaluated with the reconstruction loss as objective function in a 3-fold cross-validation scenario.

In general the number of possible MBN DAG structures for n nodes grows super-exponentially with n [24], making identification of the true graph structure highly challenging. Therefore, our aim was to restrict the set of possible DAGs a priori as much as possible via knowledge based logical constraints. More specifically we imposed the following causal restrictions:

- Nodes defined by demographic or clinical features (like age, gender, APOE genotype, and brain region) can only have outgoing edges.
- The phenotype module (= clinical outcome measures) can only have incoming edges.
- Genes and gene modules can not influence demographic or clinical features, except the age.

To additionally integrate prior knowledge defined through the knowledge graph, we tested three different strategies while building a MBN:

1. **Completely data driven:** The knowledge graph is completely ignored for structure learning.
2. **Knowledge informed:** The knowledge graph is used in the greedy hill climbing algorithm for structure learning i) as starting point, ii) as white list (intending that those edges were defined as pre-existing), or iii) as both.
3. **Completely knowledge driven:** The knowledge graph provides the structure of the MBN and additional connections are only allowed for demographics or the phenotype module.

Structure learning of the MBN was always performed via a greedy hill climber using the Bayesian Information Criterion for model selection. We employed the implementation provided in R-package *bnlearn* [69].

Evaluating the model fit. To evaluate the fit of the overall iVAMBN model we employed the generative nature of our model: Following a topological sorting of the nodes of the DAG of the MBN we first sampled from the distribution of each node conditional on its parent. Notably, for MBN nodes representing modules this amounted to sample from the posterior of the according HI-VAE, which in practice can be realized via injection of normally distributed noise, see Section Variational Autoencoders, Eq (2). Subsequently, the random sample was then decoded via the HI-VAE. Altogether we thus generated as many synthetic subjects as real ones. We then compared the marginal distribution of each variable based on the synthetic and the real data. Results, including summary statistics and Kullback-Leibler divergences are shown in the supplementary material (Fig A and B in S4 Note). Furthermore, we compared the correlation matrices of synthetic and real data.

CD33 down-expression simulation and analysis

To be able to simulate a down-expression of CD33, we first shifted the distribution of CD33 such that it reflects a 9-fold down-expression of CD33. In agreement to the theory of Bayesian Networks this operation made CD33 conditionally independent of its parents in the MBN, which amounts to deleting any of its incoming edges and resulted into a mutilated MBN. Afterwards we exploited the fact that iVAMBN is a generative model. That means we first drew samples from the conditional densities of the mutilated MBN. Practically this amounted to first topologically sort the nodes in the MBN, hence exploiting the fact that the underlying graph structure cannot have cycles. Subsequently, samples were drawn from the statistical distribution of each node while conditioning on the value of its parents. The result was a per-sample module activity score, which we then decoded through our HI-VAE models into single gene scores.

Differences between the wild-type and simulated down-expression samples were investigated afterwards via multiple statistical hypothesis tests: First, a linear regression was used to model the down-expression effect on gene expression and on the different phenotype scores. Second, the *globaltest* package in R was used to test the differential expression of specific gene sets between the wild-type and simulated down-expression group [49]. Those tested gene sets were here defined through the modules' genes used in the MBN, meaning that we tested for differential expression of MBN's gene modules. P-values were adjusted for multiple test scenario with the help of the *subsets* option of *globaltest* and via calculating the false discovery rate. The *globaltest* for gene sets, as well as the fold change analysis, was also applied to the cell line WT and KO data to be able to validate the results.

Effects of the perturbation of other candidate targets were simulated similarly as the CD33 knock-down. Again, the distribution of the respective target was shifted such that it reflected a 9-fold down- or up-regulation. The module was identified to which the candidate target had been assigned, and all variables (including the perturbed target) mapping to that module were encoded via the previously trained HI-VAE for the module. Subsequently, the effects on the phenotype could be predicted in the same way as described for CD33.

Supporting information

S1 Table. Module enrichment analysis. If the genes in a module do not enrich NeuroMMSig terms significantly (adjusted $p < 0.05$), individual genes are reported. If significant enriched terms could be found, all significant pathways are reported.
(XLSX)

S2 Table. Module assignment. For every gene the corresponding module number from the Markov clustering is given. Module 0 refers to all standalone genes.
(XLSX)

S3 Table. Bootstrap confidence results. This is the full list of the bootstrap confidence of each possible edge in the Bayesian Network. For every edge the corresponding start and end node, as well as, the bootstrap strength and the direction is given.
(XLSX)

S4 Table. Graph Clustering Metrics. The three clustering algorithms: 1) markov clustering, 2) edge betweenness, and 3) infomap were applied on the knowledge graph. For every cluster algorithm the corresponding metrics are given, as well as, the average rank. Printed in bold is the best algorithm according to the respective metric. The algorithms were ranked per metric and the average rank per algorithm was calculated.
(XLSX)

S1 Note. iVAMBNs Module Definition.
(PDF)

S2 Note. iVAMBNs Knowledge Integration.
(PDF)

S3 Note. Specificity and sensitivity analysis.
(PDF)

S4 Note. Evaluating the model fit.
(PDF)

S1 Fig. Clustered knowledge graph. Knowledge graph modules (clusters) are annotated with significantly enriched (adjusted $p < 0.05$) NeuroMMSig mechanisms. If the genes in a module do not enrich NeuroMMSig terms significantly, symbols of contained genes are reported. If multiple significant enriched terms could be found, the most significant pathway was used for naming the corresponding node. In case that a module contains a single gene, the gene symbol is reported. CD33 is marked in red, while other single genes are displayed in green, and non-single gene modules in purple.
(PNG)

S2 Fig. Quantitative effect between modules of shortest path. Each correlation (R) is shown along with its confidence interval (CI) and multiple testing adjusted p-value. Left: Correlation of CD33 with prostaglandin pathway module. Right: Correlation of prostaglandin pathway module with the phenotype module.
(PNG)

S3 Fig. Quantitative effect between modules of newly trained edges with confidence 1. Each correlation (R) is shown along with its confidence interval (CI) and multiple testing adjusted p-value. The *from* module is always shown on x-axis while the *to* module is shown on y-axis.
(PNG)

S4 Fig. Overlap of ROSMAP and Mayo network structures. The overlap of the independent bootstrap structure learning for ROSMAP data and Mayo data is shown for different threshold values. The black line represents the overlap when considering the direction of the edge, the dashed line the overlap of the network skeletons.
(PNG)

S5 Fig. Effects on phenotype scores of up- and down-regulation simulations. The bar plots show the difference between the mean score in the original data and the mean score in the simulated data for each target and each phenotype score, namely MMSE (upper two rows) and Braak score (bottom two rows). First and third row shows the results of under-expression, while second and forth rows shows the results of over-expression.
(PNG)

Acknowledgments

The results published here are in whole or in part based on data obtained from the AD Knowledge Portal (<https://adknowledgeportal.org>). Data generation was supported by the following NIH grants: P30AG10161, P30AG72975, R01AG15819, R01AG17917, R01AG036836, U01AG46152, U01AG61356, U01AG046139, P50 AG016574, R01 AG032990, U01AG046139, R01AG018023, U01AG006576, U01AG006786, R01AG025711, R01AG017216, R01AG003949, R01NS080820, U24NS072026, P30AG19610, U01AG046170, RF1AG057440, and U24AG061340, and the Cure PSP, Mayo and Michael J Fox foundations, Arizona Department of Health Services and the Arizona Biomedical Research Commission. We thank the participants of the Religious Order Study and Memory and Aging projects for the generous donation, the Sun Health Research Institute Brain and Body Donation Program, the Mayo Clinic Brain Bank, and the Mount Sinai/JJ Peters VA Medical Center NIH Brain and Tissue Repository. Data and analysis contributing investigators include Nilüfer Ertekin-Taner, Steven Younkin (Mayo Clinic, Jacksonville, FL), Todd Golde (University of Florida), Nathan Price (Institute for Systems Biology), David Bennett, Christopher Gaiteri (Rush University), Philip De Jager (Columbia University), Bin Zhang, Eric Schadt, Michelle Ehrlich, Vahram Haroutunian, Sam Gandy (Icahn School of Medicine at Mount Sinai), Koichi Iijima (National Center for Geriatrics and Gerontology, Japan), Scott Noggle (New York Stem Cell Foundation), Lara Mangravite (Sage Bionetworks).

Author Contributions

Data curation: Tamara Raschka, Bruce Schultz, Christian Ebeling.

Formal analysis: Tamara Raschka, Meemansa Sood.

Funding acquisition: Holger Fröhlich.

Methodology: Tamara Raschka, Meemansa Sood, Bruce Schultz, Aybuge Altay, Christian Ebeling, Holger Fröhlich.

Software: Tamara Raschka, Meemansa Sood, Bruce Schultz, Christian Ebeling.

Supervision: Holger Fröhlich.

Visualization: Tamara Raschka.

Writing – original draft: Tamara Raschka, Holger Fröhlich.

Writing – review & editing: Holger Fröhlich.

References

1. Hodson R. Alzheimer's disease. *Nature*. 2018; 559:S1–S1. <https://doi.org/10.1038/d41586-018-05717-6> PMID: 30046078
2. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*. 2017; 45. <https://doi.org/10.1093/nar/gkw1092> PMID: 27899662

3. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, et al. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Research*. 2011; 39. <https://doi.org/10.1093/nar/gkq1039> PMID: 21071392
4. Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Research*. 2018; 46. <https://doi.org/10.1093/nar/gkx1064> PMID: 29136241
5. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Research*. 2018; 46. <https://doi.org/10.1093/nar/gkx1132> PMID: 29145629
6. Carvalho-Silva D, Pierleoni A, Pignatelli M, Ong C, Fumis L, Karamanis N, et al. Open Targets Platform: new developments and updates two years on. *Nucleic Acids Research*. 2019; 47. <https://doi.org/10.1093/nar/gky1133> PMID: 30462303
7. Wang Y, Zhang S, Li F, Zhou Y, Zhang Y, Wang Z, et al. Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Research*. 2019. <https://doi.org/10.1093/nar/gkz981>
8. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research*. 2019.
9. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*. 2019; 47. <https://doi.org/10.1093/nar/gky1131> PMID: 30476243
10. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, et al. The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Research*. 2014; 42. <https://doi.org/10.1093/nar/gkt1115> PMID: 24234451
11. Kodamullil AT, Younesi E, Naz M, Bagewadi S, Hofmann-Apitius M. Computable cause-and-effect models of healthy and Alzheimer's disease states and their mechanistic differential analysis. *Alzheimer's and Dementia*. 2015; 11:1329–1339. <https://doi.org/10.1016/j.jalz.2015.02.006> PMID: 25849034
12. Estus S, Shaw BC, Devanney N, Katsumata Y, Press EE, Fardo DW. Evaluation of CD33 as a genetic risk factor for Alzheimer's disease. *Acta Neuropathologica*. 2019; 138. <https://doi.org/10.1007/s00401-019-02000-4> PMID: 30949760
13. Jiang T, Yu JT, Hu N, Tan MS, Zhu XC, Tan L. CD33 in Alzheimer's Disease. *Molecular Neurobiology*. 2014; 49. <https://doi.org/10.1007/s12035-013-8536-1> PMID: 23982747
14. Zhao L. CD33 in Alzheimer's Disease—Biology, Pathogenesis, and Therapeutics: A Mini-Review. *Gerontology*. 2019; 65. <https://doi.org/10.1159/000492596> PMID: 30541012
15. Siddiqui SS, Springer SA, Verhagen A, Sundaramurthy V, Alisson-Silva F, Jiang W, et al. The Alzheimer's disease–protective CD33 splice variant mediates adaptive loss of function via diversion to an intracellular pool. *Journal of Biological Chemistry*. 2017; 292. <https://doi.org/10.1074/jbc.M117.799346>
16. Wißfeld J, Nozaki I, Mathews M, Raschka T, Ebeling C, Hornung V, et al. Deletion of Alzheimer's disease-associated CD33 results in an inflammatory human microglia phenotype. *GLIA*. 2021.
17. Griciuc A, Patel S, Federico AN, Choi SH, Innes BJ, Oram MK, et al. TREM2 Acts Downstream of CD33 in Modulating Microglial Pathology in Alzheimer's Disease. *Neuron*. 2019; 103. <https://doi.org/10.1016/j.neuron.2019.06.010> PMID: 31301936
18. Griciuc A, Serrano-Pozo A, Parrado A, Lesinski A, Asselin C, Mullin K, et al. Alzheimer's Disease Risk Gene CD33 Inhibits Microglial Uptake of Amyloid Beta. *Neuron*. 2013; 78. <https://doi.org/10.1016/j.neuron.2013.04.014> PMID: 23623698
19. Steckmann T, Awan Z, Gerstman BS, Chapagain PP. Kinetics of peptide secondary structure conversion during amyloid B-protein fibrillogenesis. *Journal of Theoretical Biology*. 2012; 301:95–102. <https://doi.org/10.1016/j.jtbi.2012.02.012> PMID: 22586726
20. Proctor CJ, Pienaar IS, Elson JL, Kirkwood TB. Aggregation, impaired degradation and immunization targeting of amyloid-beta dimers in Alzheimer's disease: A stochastic modelling approach. *Molecular Neurodegeneration*. 2012; 7:32. <https://doi.org/10.1186/1750-1326-7-32> PMID: 22748062
21. Oblak AL, Forner S, Territo PR, Sasner M, Carter GW, Howell GR, et al. Model organism development and evaluation for late-onset Alzheimer's disease: MODEL-AD. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*. 2020; 6. <https://doi.org/10.1002/trc2.12110> PMID: 33283040
22. Jankowsky JL, Zheng H. Practical considerations for choosing a mouse model of Alzheimer's disease. *Molecular Neurodegeneration*. 2017; 12:1–22. <https://doi.org/10.1186/s13024-017-0231-7> PMID: 29273078
23. Arber C, Lovejoy C, Wray S. Stem cell models of Alzheimer's disease: Progress and challenges. *Alzheimer's Research and Therapy*. 2017; 9:1–17. <https://doi.org/10.1186/s13195-017-0268-4> PMID: 28610595

24. Chickering DM, Meek C, Heckerman D. Large-Sample Learning of Bayesian Networks is NP-Hard. 2012;.
25. Gootjes-Dreesbach L, Sood M, Sahay A, Hofmann-Apitius M, Fröhlich H. Variational Autoencoder Modular Bayesian Networks for Simulation of Heterogeneous Clinical Study Data. *Frontiers in Big Data*. 2020; 3:16. <https://doi.org/10.3389/fdata.2020.00016> PMID: 33693390
26. Kingma DP, Welling M. Auto-Encoding Variational Bayes. 2013;.
27. Heckerman D. In: Holmes DE, Jain LC, editors. *A Tutorial on Learning with Bayesian Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 33–82.
28. Mostafavi S, Gaiteri C, Sullivan SE, White CC, Tasaki S, Xu J, et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nature Neuroscience*. 2018; 21:811–819. <https://doi.org/10.1038/s41593-018-0154-9> PMID: 29802388
29. Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious Orders Study and Rush Memory and Aging Project. *Journal of Alzheimer's Disease*. 2018; 64. <https://doi.org/10.3233/JAD-179939> PMID: 29865057
30. Bennett DA, Schneider JA, Arvanitakis Z, Wilson RS. Overview and Findings from the Religious Orders Study. *Current Alzheimer Research*. 2012; 9. <https://doi.org/10.2174/156720512801322573> PMID: 22471860
31. Allen M, Carrasquillo MM, Funk C, Heavner BD, Zou F, Younkin CS, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Scientific Data*. 2016; 3:160089. <https://doi.org/10.1038/sdata.2016.89> PMID: 27727239
32. Sonawane AR, Platig J, Fagny M, Chen CY, Paulson JN, Lopes-Ramos CM, et al. Understanding Tissue-Specific Gene Regulation. *Cell Reports*. 2017 10; 21:1077–88. <https://doi.org/10.1016/j.celrep.2017.10.001> PMID: 29069589
33. Dongen SV. Graph Clustering Via a Discrete Uncoupling Process. *SIAM Journal on Matrix Analysis and Applications*. 2008; 30.
34. Domingo-Fernández D, Kodamullil AT, Iyappan A, Naz M, Emon MA, Raschka T, et al. Multimodal mechanistic signatures for neurodegenerative diseases (NeuroMMSig): A web server for mechanism enrichment. *Bioinformatics*. 2017; 33. <https://doi.org/10.1093/bioinformatics/btx399> PMID: 28651363
35. Nazábal A, Olmos PM, Ghahramani Z, Valera I. Handling incomplete heterogeneous data using VAEs. *Pattern Recognition*. 2020; 107:107501. <https://doi.org/10.1016/j.patcog.2020.107501>
36. Bsci DT, Msc GK, Weisinger RS, Sinclair AJ. The role of eicosanoids in the brain. *Asia Pac J Clin Nutr*. 2008; 17:220–228.
37. Ardura-Fabregat A, Boddeke EWGM, Boza-Serrano A, Brioschi S, Castro-Gomez S, Ceyzériat K, et al. Targeting Neuroinflammation to Treat Alzheimer's Disease. *CNS Drugs*. 2017; 31:1057–1082. <https://doi.org/10.1007/s40263-017-0483-3> PMID: 29260466
38. Biringer RG. The role of eicosanoids in alzheimer's disease. *International Journal of Environmental Research and Public Health*. 2019; 16. <https://doi.org/10.3390/ijerph16142560> PMID: 31323750
39. Candlish M, Hefendehl JK. Microglia Phenotypes Converge in Aging and Neurodegenerative Disease. *Frontiers in Neurology*. 2021; 12. <https://doi.org/10.3389/fneur.2021.660720> PMID: 34025562
40. Schwabe T, Srinivasan K, Rhinn H. Shifting paradigms: The central role of microglia in Alzheimer's disease. *Neurobiology of Disease*. 2020; 143:104962. <https://doi.org/10.1016/j.nbd.2020.104962> PMID: 32535152
41. Iwamoto N, Kobayashi K, Kosaka K. The formation of prostaglandins in the postmortem cerebral cortex of Alzheimer-type dementia patients. *Journal of Neurology*. 1989; 236:80–84. <https://doi.org/10.1007/BF00314401> PMID: 2709057
42. Shioya M, Obayashi S, Tabunoki H, Arima K, Saito Y, Ishida T, et al. Aberrant microRNA expression in the brains of neurodegenerative diseases: miR-29a decreased in Alzheimer disease brains targets neuron navigator 3. *Neuropathology and Applied Neurobiology*. 2010; 36(4):320–330. <https://doi.org/10.1111/j.1365-2990.2010.01076.x> PMID: 20202123
43. Seth RB, Sun L, Ea CK, Chen ZJ. Identification and Characterization of MAVS, a Mitochondrial Antiviral Signaling Protein that Activates NF- κ B and IRF3. *Cell*. 2005; 122(5):669–682. <https://doi.org/10.1016/j.cell.2005.08.012> PMID: 16125763
44. Serrano-Pozo A, Das S, Hyman BT. APOE and Alzheimer's disease: advances in genetics, pathophysiology, and therapeutic approaches. *The Lancet Neurology*. 2021; 20(1):68–80. [https://doi.org/10.1016/S1474-4422\(20\)30412-9](https://doi.org/10.1016/S1474-4422(20)30412-9) PMID: 33340485
45. Ferreira A. Calpain Dysregulation in Alzheimer's Disease. *ISRN Biochemistry*. 2012; 2012:1–12. <https://doi.org/10.5402/2012/728571> PMID: 25969760

46. Vosler PS, Brennan CS, Chen J. Calpain-mediated signaling mechanisms in neuronal injury and neurodegeneration. *Molecular Neurobiology*. 2008; 38:78–100. <https://doi.org/10.1007/s12035-008-8036-x> PMID: 18686046
47. McCarty MF, Dinicolantonio JJ, Lerner A. A fundamental role for oxidants and intracellular calcium signals in Alzheimer's pathogenesis—and how a comprehensive antioxidant strategy may aid prevention of this disorder. *International Journal of Molecular Sciences*. 2021; 22:1–27. <https://doi.org/10.3390/ijms22042140> PMID: 33669995
48. Mungall CJ, Torniai C, Gkoutos GV, Lewis SE, Haendel MA. Uberon, an integrative multi-species anatomy ontology. *Genome Biology*. 2012; 13. <https://doi.org/10.1186/gb-2012-13-1-r5> PMID: 22293552
49. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics*. 2004; 20. <https://doi.org/10.1093/bioinformatics/btg382> PMID: 14693814
50. Abou-Fadel J, Vasquez M, Grajeda B, Ellis C, Zhang J. Systems-wide analysis unravels the new roles of CCM signal complex (CSC). *Heliyon*. 2019; 5:e02899. <https://doi.org/10.1016/j.heliyon.2019.e02899> PMID: 31872111
51. Jensen TMT, Albertsen L, Bartling CRO, Haugaard-Kedström LM, Strømgaard K. Probing the Mint2 Protein-Protein Interaction Network Relevant to the Pathophysiology of Alzheimer's Disease. *Chembiochem: a European journal of chemical biology*. 2018. <https://doi.org/10.1002/cbic.201800004>
52. Goodall EF, Heath PR, Bandmann O, Kirby J, Shaw PJ. Neuronal dark matter: the emerging role of microRNAs in neurodegeneration. *Frontiers in cellular neuroscience*. 2013; 7:178. <https://doi.org/10.3389/fncel.2013.00178> PMID: 24133413
53. Mai H, Fan W, Wang Y, Cai Y, Li X, Chen F, et al. Intranasal Administration of miR-146a Agomir Rescued the Pathological Process and Cognitive Impairment in an AD Mouse Model. *Molecular therapy Nucleic acids*. 2019; 18:681–695. <https://doi.org/10.1016/j.omtn.2019.10.002> PMID: 31707205
54. Dongen SV. Graph clustering by flow simulation. Proefschrift Universiteit Utrecht; 2000.
55. Jäger ML. MCL: Markov Cluster Algorithm.; 2015.
56. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Physical Review E*. 2004; 69:026113. PMID: 14995526
57. Csardi G, Nepusz T, et al. The igraph software package for complex network research. *InterJournal, complex systems*. 2006; 1695(5):1–9.
58. Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*. 2008; 105:1118–1123. <https://doi.org/10.1073/pnas.0706851105> PMID: 18216267
59. Arratia A, Mirambell MR. Clustering assessment in weighted networks. *PeerJ Computer Science*. 2021; 7:1–27. <https://doi.org/10.7717/peerj-cs.600> PMID: 34239979
60. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*. 2012; 16. <https://doi.org/10.1089/omi.2011.0118>
61. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886
62. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007; 8. <https://doi.org/10.1093/biostatistics/kxj037> PMID: 16632515
63. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*. 2019; 47. <https://doi.org/10.1093/nar/gkz114>
64. Parviainen P, Kaski S. Bayesian Networks for Variable Groups; 2016.
65. Parviainen P, Kaski S. Learning structures of Bayesian networks for variable groups. *International Journal of Approximate Reasoning*. 2017; 88:110–127. <https://doi.org/10.1016/j.ijar.2017.05.006>
66. Becker AK, Dörr M, Felix SB, Frost F, Grabe HJ, Lerch MM, et al. From heterogeneous healthcare data to disease-specific biomarker networks: A hierarchical Bayesian network approach. *PLOS Computational Biology*. 2021; 17:e1008735. <https://doi.org/10.1371/journal.pcbi.1008735> PMID: 33577591
67. Gyftodimos E, Flach PA. Hierarchical Bayesian Networks: An Approach to Classification and Learning for Structured Data. In: Vouros GA, Panayiotopoulos T, editors. *Methods and Applications of Artificial Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2004. p. 291–300.
68. Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*. 1991; 82:239–259. <https://doi.org/10.1007/BF00308809> PMID: 1759558
69. Scutari M. Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software*. 2010; 35. <https://doi.org/10.18637/jss.v035.i03>

A.2 Unraveling progression subtypes in people with Huntington's Disease

Reprinted with permission from “**Raschka, T.**, Li, Z., Gassner, H., Kohl, Z., Jukic, J., Marxreiter, F., Fröhlich, H. (2024). Unraveling progression subtypes in people with Huntington's Disease. *EPMA Journal*, 15, 275-287. doi:10.1007/s13167-024-00368-2”

Copyright © Raschka, T., *et al.*, 2024



Unraveling progression subtypes in people with Huntington's disease

Tamara Raschka¹ · Zexin Li^{1,2} · Heiko Gaßner^{3,4} · Zacharias Kohl⁵ · Jelena Jukic^{3,7} · Franz Marxreiter^{3,6,7} · Holger Fröhlich^{1,2}

Received: 4 March 2024 / Accepted: 9 May 2024 / Published online: 28 May 2024
© The Author(s) 2024

Abstract

Background Huntington's disease (HD) is a progressive neurodegenerative disease caused by a CAG trinucleotide expansion in the huntingtin gene. The length of the CAG repeat is inversely correlated with disease onset. HD is characterized by hyperkinetic movement disorder, psychiatric symptoms, and cognitive deficits, which greatly impact patient's quality of life. Despite this clear genetic course, high variability of HD patients' symptoms can be observed. Current clinical diagnosis of HD solely relies on the presence of motor signs, disregarding the other important aspects of the disease. By incorporating a broader approach that encompasses motor as well as non-motor aspects of HD, predictive, preventive, and personalized (3P) medicine can enhance diagnostic accuracy and improve patient care.

Methods Multisymptom disease trajectories of HD patients collected from the Enroll-HD study were first aligned on a common disease timescale to account for heterogeneity in disease symptom onset and diagnosis. Following this, the aligned disease trajectories were clustered using the previously published Variational Deep Embedding with Recurrence (VaDER) algorithm and resulting progression subtypes were clinically characterized. Lastly, an AI/ML model was learned to predict the progression subtype from only first visit data or with data from additional follow-up visits.

Results Results demonstrate two distinct subtypes, one large cluster ($n = 7122$) showing a relative stable disease progression and a second, smaller cluster ($n = 411$) showing a dramatically more progressive disease trajectory. Clinical characterization of the two subtypes correlates with CAG repeat length, as well as several neurobehavioral, psychiatric, and cognitive scores. In fact, cognitive impairment was found to be the major difference between the two subtypes. Additionally, a prognostic model shows the ability to predict HD subtypes from patients' first visit only.

Conclusion In summary, this study aims towards the paradigm shift from reactive to preventive and personalized medicine by showing that non-motor symptoms are of vital importance for predicting and categorizing each patients' disease progression pattern, as cognitive decline is oftentimes more reflective of HD progression than its motor aspects. Considering these aspects while counseling and therapy definition will personalize each individuals' treatment. The ability to provide patients with an objective assessment of their disease progression and thus a perspective for their life with HD is the key to improving their quality of life. By conducting additional analysis on biological data from both subtypes, it is possible to gain a deeper understanding of these subtypes and uncover the underlying biological factors of the disease. This greatly aligns with the goal of shifting towards 3P medicine.

Keywords Huntington's disease · Progression · Artificial intelligence · Cognition · Non-motor symptoms · Predictive preventive personalized medicine · Patient stratification · Precision medicine

Introduction

The role of CAG repeat length in predicting disease onset

Huntington's disease (HD) is a progressive, autosomal-dominantly inherited neurodegenerative disease, caused by a CAG trinucleotide expansion in the huntingtin gene [1].

Tamara Raschka and Zexin Li, as well as, Franz Marxreiter and Holger Fröhlich contributed equally to this work.
Tamara Raschka and Zexin Li contributed equally to this work as shared first author.
Holger Fröhlich and Franz Marxreiter contributed equally to this work as shared last author.

Extended author information available on the last page of the article

CAG repeat lengths > 39 show full penetrance, and length of the CAG repeat is inversely correlated with disease onset [2, 3]. Core feature of HD is a hyperkinetic movement disorder, called chorea. In addition, psychiatric symptoms and cognitive deficits are an essential part [4–6], predominantly influencing patient's quality of life [7–10], and are often early signs of HD [4, 11]. This gives great potential for predictive, preventive, and personalized medicine (3PM) to improve diagnostic and treatment of patients.

The “CAG age product” (CAP) that is, as the name intends, based on age and CAG repeat length can predict the age of onset of the disease [12, 13]. The CAP is commonly collected and used in clinical studies to determine the pre-manifest status of the patients or to assess the disease stage in terms of being close to or far away from predicted onset at study entry [12, 14–16]. Nevertheless, although CAG repeat length is often measured in a predictive manner in HD patient's relatives, no preventive actions are taken, as no preventive treatment is available.

Taking into account the genetic component and the clear psychiatric symptoms and cognitive deficits in HD patients, there is great potential for the aspects of 3P medicine, as the clinical diagnosis of manifest HD is only based on the presence of unequivocal motor signs in CAG repeat expansion carriers and ignores the other aspects [17–19].

The need for a 3PM approach

Although there is such a clear genetic background in HD, the multifactorial and progressive nature of HD, in combination with a long pre-manifest phase again strengthens the need for 3PM concepts in HD, as there is a high variability of symptoms in HD patients that challenges counseling [6, 20, 21].

Commonly, clinical diagnostic criteria feature a pre-manifest phase (prior to motor diagnosis), followed by the motor-manifest period, divided into five stages [17]. More recently, a biological classification system has been established to characterize individuals for research purposes [22] and longitudinal data-driven machine-learning algorithms identified even nine distinct disease stages [23].

These results have tremendously helped to better understand the natural course of HD on a population level and paves the way for significantly earlier and correct diagnosis, enabling the paradigm shift from delayed intervention towards predictive and personalized medicine. Earlier diagnosis in the end helps to identify and apply effective early interventions for HD patients in early disease stages [19, 24, 25].

Therefore, the urgent need for personalized medicine has to be addressed. Here, identifying patient subtypes with a similar disease trajectory could help to untangle disease-modifying factors and define patient stratification [21, 24,

26–29]. Such an approach aims, on the one hand, to identify the best treatment strategy for each individual disease subtype, but, on the other hand, also helps to understand the disease characteristics itself, which in the end will contribute to a better prediction of disease progression, as well [24, 29].

State of the art in subtype identification in neurodegenerative diseases

Several models have been developed that cluster and predict patient subtypes based on their longitudinal trajectories, but most of them do not account for variations in the dynamics of the disease progression or identified markers discriminating pre-defined progression groups [21, 23]. In fact, a fully multivariate data-driven approach to unravel potential underlying disease progression groups, that otherwise can be overseen, is still missing. In this regard, methods from the field of data science and artificial intelligence (AI) pose a great opportunity to come closer to the vision of a predictive, personalized, and preventive medicine (3PM).

In the past, data-driven clustering of multisymptom disease trajectories has shown promising results in other neurodegenerative disorders, such as Alzheimer's or Parkinson's [28, 30, 31]. The rate of progression is typically variable across the disease trajectory in HD, with the steepest decline in function being seen in early to mid-disease stages [32]. Also, patients may be initially diagnosed at different stages of their disease. Thus, there is a need for alignment of trajectories observed for an individual patient along a population average disease trajectory [33, 34]. This can be done by modeling a continuous common disease timescale, e.g., with the help of non-linear mixed effect models. These models enable a removal of the effect of the actual time point of symptom onset on the overall trajectories by assessing a latent time estimate and adjust all patients' trajectories across a common disease timescale [33, 35, 36].

Novelty beyond the state of the art

After diagnosis, accurate and early prognosis of the progression of the disease is of core relevance for those living with HD. For a patient, it is important to know her prospective symptom development in order to adapt life accordingly. In addition, a concrete prognosis could help doctors to better organize and manage therapies. This is where our work contributes to. For the first time, to our knowledge, HD subtypes based on multisymptom disease trajectories were identified and validated. Here, not only motor symptoms were evaluated but also cognitive symptoms are taken into account for subtype identification, as psychiatric symptoms and cognitive deficit are an essential part of the disease and also early signs of HD that are ignored in the diagnosis of manifest HD [4–6, 11, 17–19].

With the help of an AI/ML model that allows for predicting the disease progression subtypes based on only the first visit or by including additional follow-up visits, we addressed important 3PM aspects, because such a model could support a better individualized disease management via optimized counseling as well as support patient in their own life decisions.

Working hypothesis

In the current study, we hypothesize that manifest HD patients define heterogeneous progression subtypes that can be identified by advanced AI methods. Subtype identification will allow the classification of patients into one of the multisymptom trajectory clusters based on their first visit data only or by including additional follow-up visits. This will in the end result into a better personalized and preventive treatment for HD patients as patient counseling can be optimized. Additionally, patients can be provided with a better indication of their prognosis.

Therefore, we clustered HD patients on the basis of their longitudinal trajectories using our previously introduced Variational Deep Embedding with Recurrence (VaDER) neural network algorithm [30]. As such trajectories are often temporally related to the study baseline, this study aims to adjust for this confounding effect using a non-linear mixed effect model, originally introduced by Raket [36], which assumes that the longitudinal trajectories of pre-manifest and manifest HD patients are adjustable along a common disease timescale. An early prognosis of disease progression for each patient, reflecting the predictive and personalized aspect of 3PM, will be achieved by developing, evaluating, and validating a machine learning-based prediction model.

Methods

Dataset and patient selection criteria

Data used in this work were generously provided by the participants in the Enroll-HD study [37] (ClinicalTrials.gov Identifier: NCT01574053) and made available by CHDI Foundation, Inc. Enroll-HD is a global clinical research platform designed to facilitate clinical research in Huntington's disease. Core datasets are collected annually from all research participants as part of this multicenter longitudinal observational study. Data are monitored for quality and accuracy using a risk-based monitoring approach. All sites are required to obtain and maintain local ethical approval. Overall, up to now, over 20,000 patients from 21 countries at 159 sites were recruited [38]. Enroll-HD provides data from HD patients, including manifest and pre-manifest patients. The study has

collected clinical score data and demographics during annual visits. During these visits, patients undergo a core battery of test, including demographics, medical history, and clinical assessments of the four HD domains: motor, cognitive, behavioral, and functional. A CAG genotyping was done for each participant to get their CAG repeat length.

During this study, we used the fifth release of the Enroll-HD data from October 2020 that was released in December 2020. Within this release, we selected the information from scheduled visits of manifest and pre-manifest patients that had at least one follow-up visit after the baseline screening. Control participants were excluded for the purpose of this study. Participant category was assessed at each visit by the study examiner. Details about the definition of pre-manifest and manifest patient group can be found in the Supplementary Material.

Used features during modeling and clustering of trajectories are the Unified Huntington's Disease Rating Scale (UHDRS) [39] and the mini-mental state examination (MMSE). The UHDRS consists of four different domains: motor, cognitive, behavioral, and functional, where each of the domains relates to different fields of possible symptoms. Additionally, the MMSE was used as a cognitive test. In total, we used three different features, coming from two different assessments, namely, UHDRS total motor score (TMS), UHDRS symbol digit modality test (SDMT), and the total MMSE score (MMSE). Furthermore, age and sex were used as covariates. Characteristics of all used patients can be found in Table 1.

Shown are statistics of patients used for the subtype analysis as training (manifest) and validation (pre-manifest) set. Statistics are shown at patients' baseline visit for manifest and first manifest visit for pre-manifest patients after conversion. The age, CAG repeat length, and clinical scores are described by their mean and standard deviation.

Table 1 Characteristics of manifest and pre-manifest patients

	Manifest	Pre-manifest
Number of patients	7533	372
Age	52.9 ± 12.53	47.01 ± 12.06
Sex		
Male	3659	170
Female	3874	202
CAG length	43.96 ± 3.86	43.3 ± 3.32
UHDRS		
Total motor score	36.92 ± 21.07	14.19 ± 8.44
Symbol digit modality test	23.39 ± 13.06	38.95 ± 12.76
MMSE	25.00 ± 4.39	28.01 ± 2.12

Model building methods

In the following analysis, both manifest and pre-manifest patients were used for modeling common disease time trajectories with a non-linear mixed effect (NLME) model. VaDER clustering and prediction models were then trained on manifest patients only, while validating these models afterwards on pre-manifest patients. Common disease time-scale trajectories of pre-manifest patients were therefore shortened to their manifestation phase, meaning that the first visit when manifesting the disease was considered the start of their used trajectory.

Non-linear mixed effect model

The here-used model follows the rational described earlier [33, 36]. Shortly, it is a non-linear mixed effect model with a mean curve defined by the fixed effects and random effects describing the deviation of each patient from that mean curve. During the here-described analysis, the mean curve is formulated as a generalized logistic function:

$$\mu(t) = A + \frac{K - A}{(1 + e^{(-B(t+s))^v})} + c$$

where A is the left and K the right asymptotic value which reflects the minimal and maximal possible value of a specific clinical measure. Parameters B and v define the curvature of the function, with B as time scaling parameter and v as an asymmetry parameter. A shift in time is modelled by s as horizontal shift and a vertical shift can be modelled by c .

During the modeling, the values of A and K are fixed according to the actual modelled clinical score. Fixed effects can be modelled for specific covariates, e.g., the manifestation status of the patients at baseline, such that the parameter estimate of s describes the mean difference in time between the pre-manifest and manifest patient groups over the continuous common disease timescale. This difference is modelled relative to the pre-manifest patient group, meaning that $t = 0$ on the common disease timescale corresponds to the average status of the pre-manifest patients at the time of conversion into manifest disease status.

The model was built with the help of the *proglmod* R-package [40] which is based on the *nlme* package [41]. More details on the model formulation can be found in the Supplements.

Multivariate clustering of clinical trajectories

The previously published Variational Deep Embedding with Recurrence (VaDER) [30] was used to cluster the shifted time series that are the output of the *nlme* model. Based on a variational deep embedding framework for learning

low-dimensional representation of data points, two long short-term memory networks that model the multivariate time series, and an implicit imputation layer, the VaDER method allows to model and cluster short time series with large amount of missingness. More details can be found in the original publication [30] and in the Supplements. Training of the VaDER was based on manifest patients only, but pre-manifest patients were used as validation cohort in the later phase of the study.

Machine learning classifiers

Random Forest [42] and XGBoost [43] classifiers were trained on baseline data only (BL), as well as baseline and follow-up data (BLtoFU1 and BLtoFU2) from manifest patients. We here used the labels of the VaDER approach, the cluster assignment, as the classes that need to be predicted. As predictors, multiple cognitive, motor, functional, and neurobehavioral scores were used. Those can be found in Supplementary Table 3. A hyperparameter optimization with randomized search with 100 parameter settings preceded a tenfold nested cross-validation implemented with *scikit-learn* [44] and XGBoost [43] packages. Hyperparameter spaces and optimal hyperparameters, found based on the best AU-ROC, can be found in the Supplementary Material. Cross-validation results of both Random Forest and XGBoost are shown in the Supplementary Material, along with the decision to choose XGBoost for final modeling due to higher prediction performance. These trained XGBoost models were then applied on the pre-manifest patients as a validation set in the later phase of the study.

Feature importance analysis

For feature importance analysis, SHAP values [45] for each feature in the model were calculated with the implementation of the python package *shap*. For interpretation, the natural additive behavior of SHAP values was used to calculate aggregated SHAP values for specific domains of clinical tests, such as functional, cognitive, or neurobehavioral tests. This was done by summarizing SHAP values from clinical tests belonging to each of the domains. The scores belonging to each of the domains are listed in Supplementary Table S3.

Statistical testing

Statistical testing was following the SHAP analysis for all single features underlying the top aggregated features. Depending on the type of variable, either Kruskal–Wallis (numerical), Fisher's exact (bi-categorical), or chi-square independence test (multicategorical) was used to test the distribution of the respective features within the found subtypes. Age and sex were included as possible confounders

in all tests and p -values were corrected for multiple testing using the Benjamini–Hochberg procedure.

The analysis of medications for specific indications across progression subtypes was performed using Fisher’s exact test, and p -values were corrected for multiple testing using Benjamini–Hochberg procedure.

Results

Data

Data used in this study comes from the Enroll-HD study (ClinicalTrials.gov Identifier: NCT01574053) [37] (Data Cut 10/2020). In total, 11,093 patients (7548 manifest and 3545 pre-manifest) were used in this work for the longitudinal modeling over a common disease timescale, of which 7905 patients (7533 manifest and 372 pre-manifest) were eligible for the later analysis regarding the subtypes. For pre-manifest patients, only their manifest phase was used in the later analysis as independent validation set, explaining the large drop in the number of patients. An overview of the characteristics of patients used for the subtype analysis at their first included visit is presented in Table 1.

Enroll-HD is a longitudinal study, collecting data from patients at multiple annually scheduled follow-up visits. The number of available patients per follow-up visit is shown in Table 2.

Shown is the number of manifest and pre-manifest patients used in the subtype analysis having the respective visits available; e.g., 7066 manifest patients had a first follow-up visit, while 4817 manifest patients also had a second follow-up visit. For manifest patients, available visits over 6 years were summarized for visual reasons. The maximum number of visits a manifest patient had was 14. Baseline visit for pre-manifest patients relates to their first manifest visit, after conversion from pre-manifest status.

Non-linear mixed effect model

A non-linear mixed effect (NLME) model was fitted to longitudinal data of 11,093 manifest and pre-manifest patients in a multivariate manner, modeling the Unified Huntington’s Disease Rating Scale (UHDRS) total motor score (TMS), UHDRS symbol digit modality test (SDMT), and mini-mental state examination (MMSE). These scores represent current gold standard scores addressing motor and cognitive aspects of the disease [46, 47]. The

NLME model gives the opportunity to align trajectories on a common, potentially unobserved (i.e. latent), disease timescale. Figure 1a–c shows the aligned trajectories of pre-manifest and manifest patients of the three modeled clinical scores along the common disease timescale. As expected, the original trajectories of pre-manifest patients are mostly shifted towards the left (back in time). Vice versa, manifest patients’ trajectories are mostly shifted to the right, as patients naturally are in pre-manifest state before their disease manifests. Resulting latent time estimates, indicating the difference between the actual time axis and the common disease timescale, are validated by correlating the predicted age at time 0 (i.e. manifestation time) for each individual against the observed age at first diagnosis and first motor symptoms. Hereby, a linear regression with a slope of 1 would indicate that observed diagnosis and predicted symptom onset are consistent and therefore, trajectories would be perfectly aligned. Results (Fig. 1d, e) in this study show slopes of 0.999 (diagnosis) and 1.002 (first motor symptoms), demonstrating a good fit of the NLME model to the observed data. Hence, estimates of latent time seem reliable.

VaDER clustering results in two subtypes

Aligned trajectories of 7533 manifest HD patients were used to train a VaDER model for clustering patients into subtypes [30]. This resulted in two clusters, of which one cluster was a large one with 7122 patients included, while the second cluster only contained 411 patients (Fig. 2). While the second cluster shows a steep decline of SDMT and MMSE and rise of TMS starting from around day 200 on the common disease timescale, the first cluster demonstrates almost no impairment of patients over all three outcome scores and the whole common disease timescale, meaning that the large cluster and thus most HD patients show a stable pattern in the clinical scores after a minimal worsening in the beginning of their disease. However, a smaller subset of particularly vulnerable patients represents faster disease progression than most HD patients. Additionally, SDMT progression patterns of the two clusters are strictly distinguished of each other. Here, the baseline levels are already different, whereas both clusters are not separable in the beginning when looking at the MMSE or TMS.

Table 2 Number of baseline and follow-up visits

	BL	FU1	FU2	FU3	FU4	FU5	FU6	> 6FUs
Manifest	7533	7066	4817	3256	1922	774	211	62
Pre-manifest	372	372	213	94	34	9	-	-

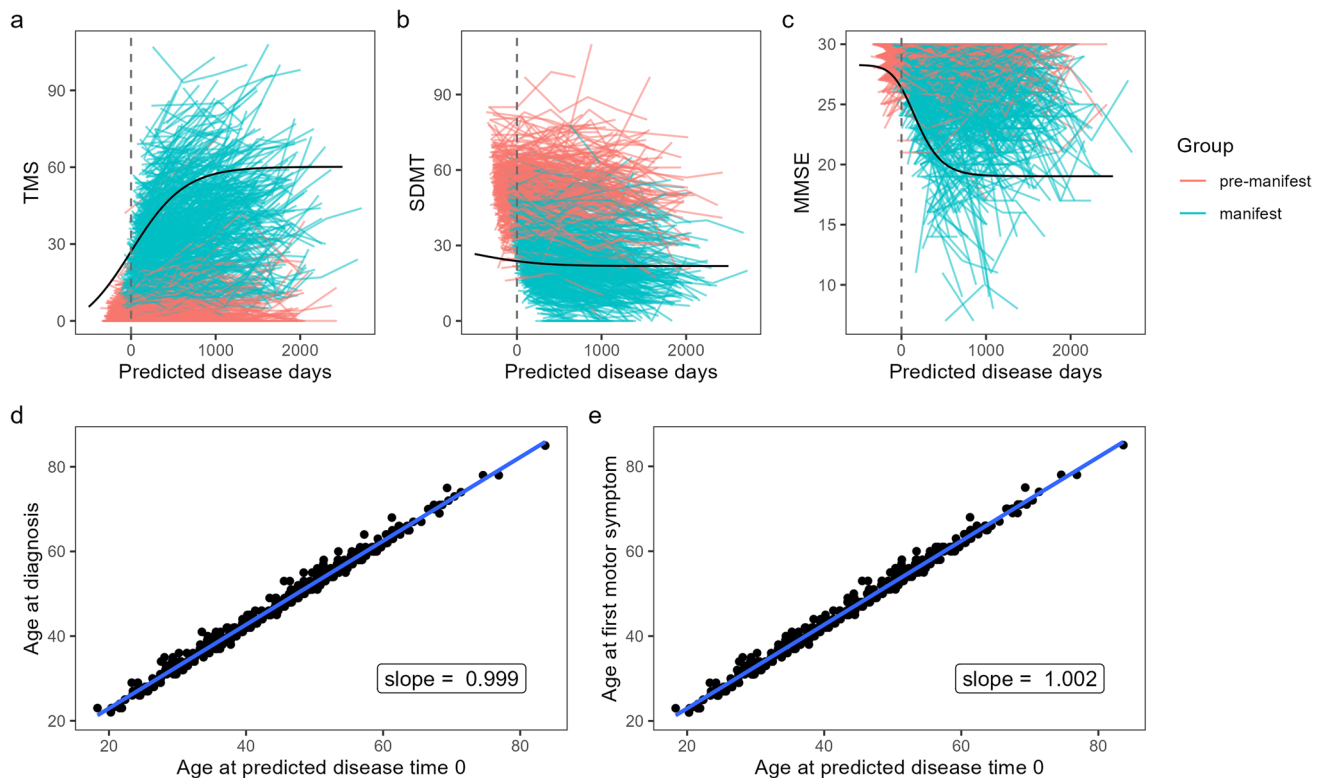


Fig. 1 Multivariate NLME modeling. **a–c** Disease trajectories of pre-manifest (red) and manifest (blue) patients aligned along a common disease timescale based on the estimated random effects representing the latent time. Three different outcomes are shown: Unified Huntington’s Disease Rating Scale (UHDRS) total motor score (TMS), symbol digit modality test (SDMT), and mini-mental state examination (MMSE). For visualization reasons, 1000 trajectories were randomly

selected for plotting. The black curve shows the underlying mean curve estimated by the NLME model. The grey dashed line marks time equals 0 on the common disease timescale. **d, e** Validation plots of time alignment, where age at diagnosis (**d**) and first motor symptom (**e**) are plotted against the age at predicted disease time 0 fitted with linear models resulting from the NLME approach

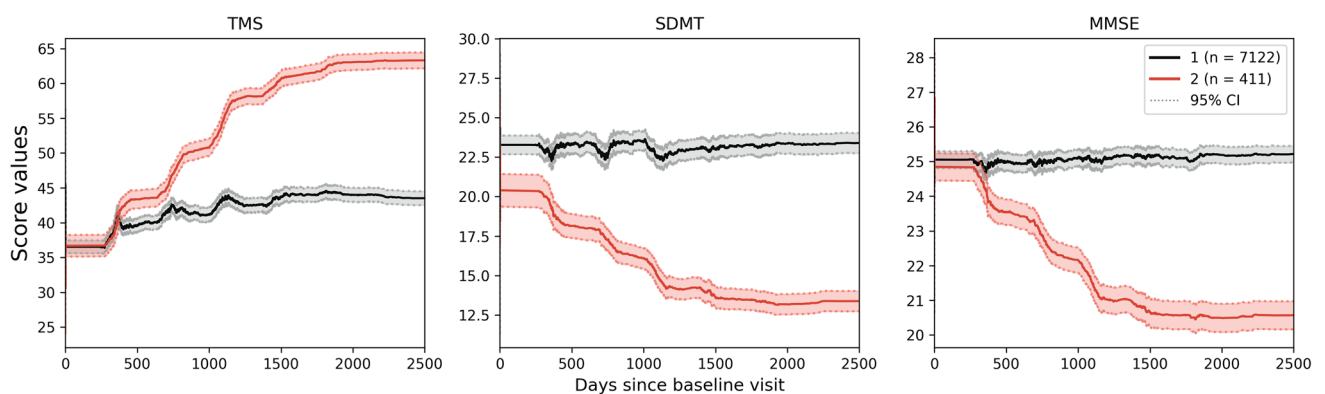


Fig. 2 Mean progression cluster trajectories of manifest HD patients. Two clusters are found as the result of training VaDER on the aligned multisymptom (TMS, SDMT, MMSE) trajectories of manifest HD patients. Cluster 1 is shown in black and contains 7122 patients,

whereas Cluster 2 (red) only contains 411 patients. Dashed lines indicate the 95% confidence interval of the mean trajectories. Overall, the first cluster shows more stable patterns over time compared to the second cluster

Predicting the HD progression subtype from clinical data

We trained an XGBoost classifier to predict the progression subtype of a patient only using baseline data (BL), or additional data from the next (BLtoFU1) or next two visits (BLtoFU2). We used multiple cognitive, motor, functional, and neurobehavioral scores, as well as demographic and medical history features, as predictors. Those can be found in Supplementary Table S3. All classifiers were trained in a tenfold cross-validation scheme. That means we systematically and sequentially held 10% of the patients out for testing our classifier, while the training was performed on the rest of the data. Resulting receiver operator characteristic (ROC) curves are shown in Fig. 3 and indicate a high area under ROC curve (AUC) of 95% (BL), 99% (BLtoFU1), and 99% (BLtoFU2), respectively. This shows that classification performance grows with including additional follow-up visits, although the accuracy based on only the baseline visit is already high. Hence, first visit data alone already contains sufficient signal to make an accurate prognosis regarding subsequent progression pattern of the disease. A comparison to ground-truth models, containing either only CAG repeat, age plus sex, or all three confounders together, shows that the models using additional clinical data are significantly better (CAG 63%; age and sex 64%; age, sex, and CAG 71%).

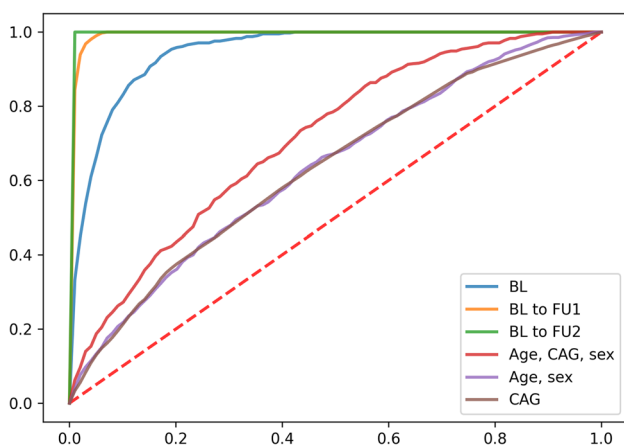


Fig. 3 XGBoost classifier ROC curves. Mean ROC curves from tenfold cross-validation setting, are shown for multiple cases of prediction from clinical data, based on baseline visit only (BL, blue), baseline plus first follow-up visit (BLtoFU1, orange), or baseline plus first and second follow-up visit (BLtoFU2, green). The training performance of the three classifiers is summarized via the area under the ROC curve (AUC) of the mean ROC curve and is 95%, 99%, and 99%, respectively. Additionally, ROC curves from ground-truth classifiers based on only age, sex, and CAG repeat (red); age and sex (purple); or only CAG repeat (brown) are shown. Here, AUCs are 71%, 64%, and 63%)

Feature importance analysis

We conducted an analysis using Shapley Additive Explanations (SHAP) to better understand the contribution of individual features in the machine learning classifiers. Here, multiple motor, cognitive, functional, and neurobehavioral scores were aggregated within their domains. A complete list of which scores were included in each domain can be found in Supplementary Table S3. The SHAP analysis demonstrated that cognitive scores were very important for predicting the correct HD progression subtype in all classifiers, most influencing in the BL classifier (Fig. 4). However, also motor scores played a relevant role, specifically in the BLtoFU2 classifier, which employs information of the second follow-up visit. Additionally, the number of CAG repeats was among the most influential features for the BL and BLtoFU1 classifiers. Moreover, neurobehavioral scores had a strong impact on all models. The amount of cognitive impairment, observable apathy, and a history of perseverative obsessive behaviors were among the top important features in the BL classifier, but were becoming less influencing when including more follow-up visits. Interestingly, functional scores became more important for the subtype prediction when including more follow-up visits.

Clinical characterization of progression subtypes

In addition to the SHAP analysis, statistical tests regarding the differences in the features, aggregated in the top 10 most important features for the prediction of progression subtypes, were conducted (see details in the “Methods” section). All tests were adjusted for age and sex as possible confounders and corrected for multiple testing. For the BL classifier, among the five most significantly different single features were the cognitive impairment ($p < 3.3e^{-13}$) and multiple cognitive assessments, like categorical verbal fluency ($p < 3.3e^{-14}$), letter verbal fluency ($p < 3.3e^{-14}$), and Trail Making Test Part A ($p < 3.4e^{-9}$) and B ($p < 4.5e^{-16}$). The amount of patients with cognitive impairment is higher in the second, smaller cluster, which is in concordance with the result, that the time needed for completion of Trail Making Test Part A and B is higher, and the number of correct answers in the categorical and letter verbal fluency tests is lower in this cluster than in the first, large cluster. Visualization of the distribution of these features can be found in the Supplementary Material. In the case of BLtoFU1 and BLtoFU2 classifiers, cognitive assessments were among the top significantly different features as well. Here, the SDMT (BLtoFU1: $p < 1.4e^{-32}$; BLtoFU2: $p < 2.1e^{-35}$ (FU2)) and Stroop tests (color naming: $p < 1.6e^{-23}$ (BLtoFU1), $p < 4.4e^{-31}$ (BLtoFU2); inference: $p < 5.7e^{-23}$ (BLtoFU1)) are present. The number of correct answers in the Stroop and SDMT is lower in the second cluster, as well. Also,

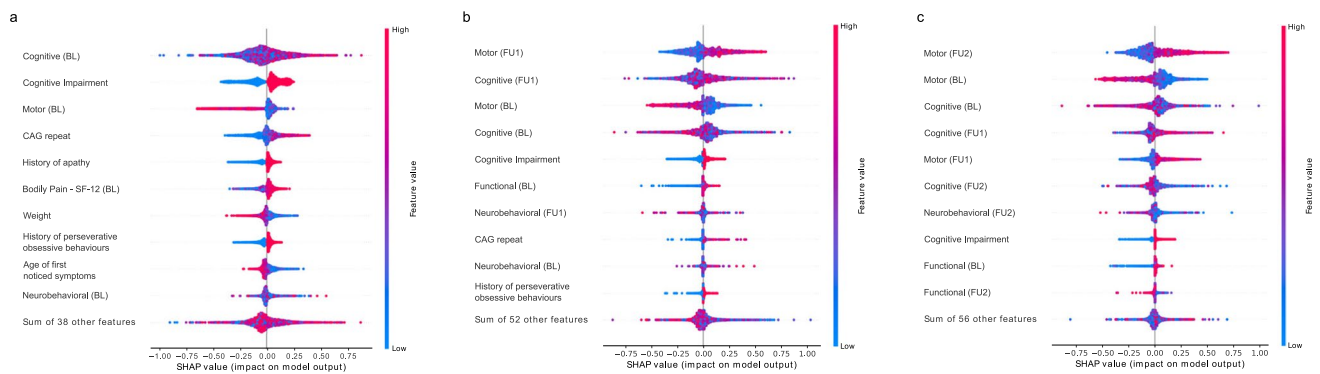


Fig. 4 SHAP values of TOP10 aggregated features. The impact of the features based on aggregated SHAP values is shown here for BL (a), BLtoFU1 (b), and BLtoFU2 (c) XGBoost classifier. Multiple features were aggregated for the SHAP analysis based on the same domain of clinical tests, namely, cognitive, motor, functional, and neurobehav-

ioral tests. A higher positive SHAP value indicates a higher influence of a feature to predict a patient as fast progressing. A more negative value indicates a higher tendency towards predicting the patient as slow progressing. The actual value of the feature is shown in a color code. The darker the red color, the higher the feature value

the MMSE scores are significantly (BLtoFU1: $p < 3.0e^{-30}$; BLtoFU2: $p < 1.2e^{-30}$) lower in the second cluster. Therefore, subtypes were mostly distinguished by the amount of cognitive impairment of patients at baseline. Additionally, the TMS is significantly higher (BLtoFU1: $p < 3.8e^{-26}$, BLtoFU2: $p < 1.1e^{-48}$) in the second cluster. Furthermore, CAG repeat length, which is known to be a highly predictive feature of HD onset [3, 48], was also a predictive feature for subtype assignment and thus for HD progression. A complete list of the statistical test results can be found in Supplementary Table S7.

With respect to medication, intake of medication for five different indications, particularly relevant to HD, was compared between subtypes: chorea, depression, anxiety, irritability, and cognitive decline. Here, significant differences could be observed across the progression subtypes for chorea ($p = 0.0002$), irritability ($p = 0.0037$), and anxiety ($p = 0.0079$) and while no significant difference between the progression groups were found for depression ($p = 0.0892$) or cognitive disorder ($p = 0.0981$) medication. Distributions of each medication in each subtype can be found in Supplementary Fig. S3.

Application of model on pre-manifest patients for validation

To further validate the generalization ability of the classification models, 372 pre-manifest at baseline patients were used as a further validation set. First, their conversion point from pre-manifest to manifest status was identified, as the visit at which the diagnostic confidence level altered from “motor abnormalities are likely signs of HD” to “motor abnormalities are unequivocal signs of HD”. Then, common disease time trajectories beyond this conversion time were clustered into two subtypes by applying the previously

trained VaDER model. By only using the manifest part of the disease trajectory of an originally pre-manifest patient, we ensure the applicability of the model on this independent validation set, as it had been trained on manifest patients only. As can be seen in Fig. 5, Cluster 1 (black) contains, similar to the training set, the majority of patients, 341 in total, compared to only 31 patients in Cluster 2. Additionally, the trajectories of first cluster show more stable patterns over time in contrast to the second cluster. The larger confidence band of the second cluster in comparison to the training result, is due to the small number of patients within this group.

The assigned subtype of each patient was then used to validate the previously trained prediction models, BL, BLtoFU1, and BLtoFU2. ROC curves for the three different classifiers are shown in Fig. 6. The performance of the three different classifiers, BL only, BLtoFU1, and BLtoFU2, showed AUCs of 75%, 79%, and 88%. This demonstrated the generalization ability of the classification models and also strengthened the fact that classification performance increased with inclusion of additional follow-up visits, as expected.

Analysis of medication for five different indications showed no significance between the two subtypes (chorea: 0.1725, depression: $p = 0.8520$, anxiety: $p = 0.1725$, irritability: $p = 0.1987$, cognitive decline: $p = 0.3048$). Distributions of each medication in each subtype can be found in Supplementary Fig. S4.

Discussion

The vision of a predictive, personalized, and preventive medicine demands to better tailor disease management and treatment to the needs of an individual patient by

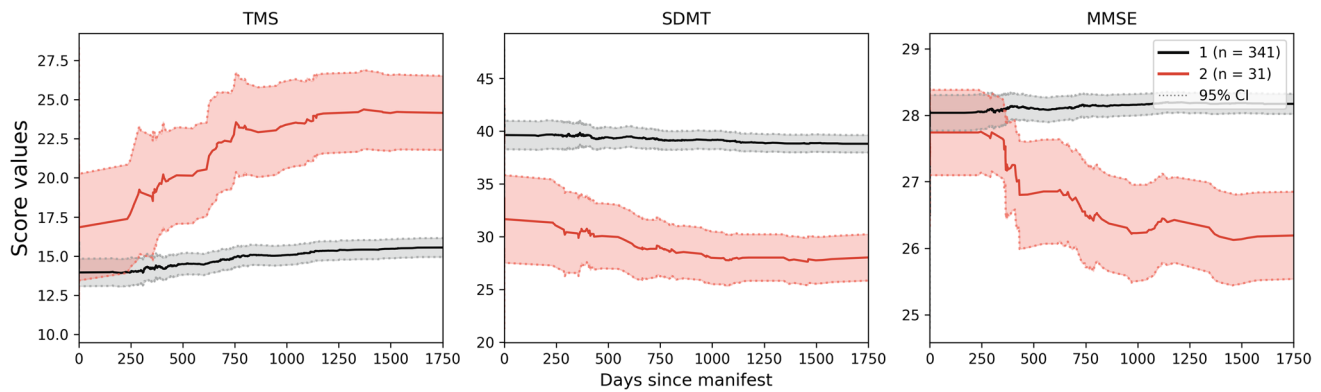


Fig. 5 Mean progression cluster trajectories of pre-manifest patients. Applying the previously trained VaDER model on the multisymptom (TMS, SDMT, MMSE) trajectories of pre-manifest patients as a validation set results in two main trajectory clusters. Similar to the training scenario, Cluster 1 (black) contains the majority of patients (341),

while Cluster 2 (red) only contains 31 patients. Again, the trajectories of first cluster show more stable patterns over time compared to the second cluster. Confidence intervals, especially for the second cluster, are larger than in the training case because of the significantly smaller number of patients

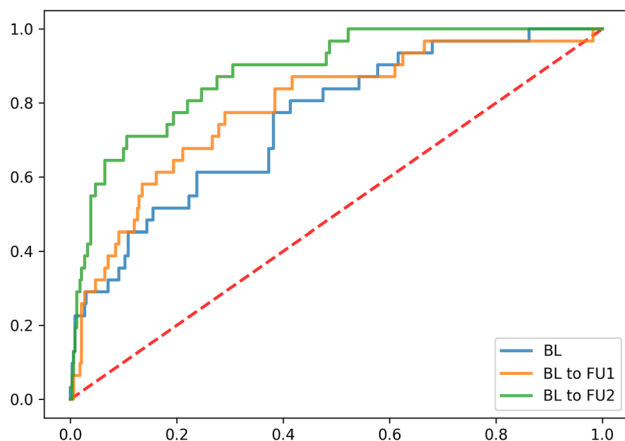


Fig. 6 ROC for XGBoost classifiers applied on pre-manifest patients. The performance of the three RF classifiers BL only (blue), BLtoFU1 (orange), and BLtoFU2 (green) including different numbers of follow-up visits additional to the BL visit is presented. The area under the ROC curve (AUC) of the classifiers is 75%, 79%, and 88%, validating that classification performance increases with including additional visits

considering a large set of patient specific characteristics. Specifically, in the context of HD, there is an urgent need for a better personalized projection of symptom development to optimize counseling and planning of a patient's future. In this regard, our work first identified two distinct HD subtypes within the Enroll-HD dataset, based on an AI-based clustering of patients' disease trajectories that were aligned on a common disease timescale. From the two identified subtypes, surprisingly, one subtype relates to a large cluster showing a relatively stable disease progression, and the second, albeit smaller cluster, shows a

dramatically more progressive disease trajectory. While this clustering was based on the UHDRS total motor score (TMS), UHDRS symbol digit modality test (SDMT), and the MMSE, the derived progressive subtype correlated with the CAG repeat length [49], underlining the robustness and validity of our clustering approach. In addition, the found subtypes could be replicated on an independent set of pre-manifest patients from the same cohort, further substantiating the validity of our approach.

Our approach is advantageous over other methods used and able to capture the heterogeneity in the disease course of HD, because integration of covariates into the modeling of each outcome allows to model effects of these covariates not only on disease stages but also on progression rate. Thus, we are able to adjust for potential covariate effects, such as age and gender on motor and cognitive symptoms. This approach also prevents bias in the data caused by disease diagnosis time, which is automatically considered when shifting patients to an earlier disease time on the common disease timescale, and thus, this strategy accounts for heterogeneity and uncertainty in the data. The subsequent use of VaDER for clustering the aligned disease trajectories allows then for simultaneous multivariate clustering rather than univariate clustering [30]. Furthermore, it allows for non-linear interactions across multiple scales for identification of the subtypes and not only captures a snapshot at a very specific time point in a patients' medical history, but integrates the whole available disease course of each patient for specific clinical variables that are suggested for measuring the disease severity [30].

In addition to the above-mentioned differences in subtypes, further neurobehavioral, psychiatric, and cognitive scores were correlated with the found subtypes. In particular, the amount of cognitive impairment was the major difference between groups. Cognitive decline is correlated

with the age at onset in HD patients [49]. Here, not only the scores used for clustering itself, namely, the MMSE and the SDMT, two already established scores reflecting morbidity and reduction in quality of life in HD, but also other cognitive scales, especially frontal-executive tests, as the letter verbal fluency, and Trail Making A and B, are distinguishing the subtypes. Thus, deficits in the overall cognitive performance, particularly the executive function, are associated with a drastically more progressive disease course, underlining previous data [50].

One potential confounder in our analysis may be the different intake of antichoreatic medication between groups. As our clustering is based on the TMS among others, intake of antichoreatic medication, leading to a lower TMS, may bias towards the slow progression group. However, a significantly higher amount within the fast progression cluster received antichoreatic medication. Thus, antichoreatic medication intake, likely reducing TMS motor scores, does not appear to influence clustering. Moreover, intake of antidepressive and anxiolytic medication was significantly more frequent in the fast-progressing group. These data suggest that the higher intake reflects the need to treat more advanced symptoms and does not cause a treatment bias in our clustering approach. In addition, the above-mentioned fact that cognitive scores are not used for clustering and CAG repeat length are among the strongest separating factors suggests a minor role of medication intake associated biases.

Conclusion and expert recommendation in the framework of 3PM

The study identified two distinct HD progression subtypes, one relating to a larger cluster showing relatively stable disease progression and the second, smaller cluster, showing a more progressive disease trajectory. Characterization of the two subtypes showed a major difference in the amount of cognitive impairment between the groups. Multiple cognitive scales, especially frontal-executive tests, are distinguishing the subtypes. Thus, deficits in the overall cognitive performance are associated with a drastically more progressive disease course. Therefore, cognitive tests need to be taken into account, as they are more reflective of the disease progression than motor tests in many cases.

In relation to 3PM in HD, we see the contributions of our work as follows:

- (i) *Predictive approach*: The subtype classification AI model can categorize HD patients based on their motor, as well as non-motor symptoms. Each individual patient can be assessed based on their personal profile and a clear and objectively measured disease

progression is predicted. This offers HD patients a better perspective on their disease progression and allows them to organize their lives accordingly, which is the key to improving patient's quality of life. In addition, a concrete prognosis could help doctors to optimize counseling and treatment of symptoms.

- (ii) *Targeted prevention*: Cognitive deficits commonly appear in HD patients and a clear correlation with the disease progression is also known, but until now, they have been ignored in diagnosis, prevention, and prediction. This study clearly shows that non-motor symptoms, especially cognitive decline, are of major importance and need to be addressed in an optimized patient counseling and treatment (e.g., via cognitive training) by taking the results of cognitive test into account. But further work is needed for concrete advancements towards targeted prevention based on the two HD progression subtypes.
- (iii) *Personalization of medical services*: A personalized projection of symptom development via the here-developed AI model helps to optimize counseling and thus the planning of each individual patient's future. The individual prediction enables doctors to initiate the appropriate personalized therapies. With a reassessment at the next clinic visit and the resulting possible adjustment or refinement and concretization of the prognosis, the treatment and therapy can be adapted. Such refinements could improve the quality of patient's life.

Conclusion

In summary, we show that non-motor symptoms are of major importance for predicting and categorizing each patients' disease progression pattern, even though they alone are not diagnostic. Our results substantiate a clinically well-known aspect of HD: the fact that cognitive decline is oftentimes more reflective of the progression of the disease than motor aspects. As a consequence, our analysis suggests that patient counseling should take into account results of the cognitive test found to be relevant in this study.

Limitations and outlook

To further explore on our results, future work should focus on the correlation of our clustering with biological data such as MRI volumetry [51] or CSF biomarkers like neurofilament light protein [52]. This would help to delineate how far our clustering approach reflects underlying biological aspects of the disease and addresses the multimodal diagnostic concept to provide the maximum of clinically relevant information. In addition, neurobehavioral or psychiatric scores are currently underrepresented in our clustering approach, as well as in other

studies [11, 46]. Yet, to include these scores in future work, we need more stable and easier to estimate non-linear mixed models allowing for the alignment of multivariate trajectories. Although, there is a multivariate version of the NLME available [33], estimating parameters of such a model is time-consuming and needs many manual steps, e.g., for finding or even tuning the start parameters for the estimation, which is inappropriate when integrating even more desired outcomes.

A further prospect regarding 3PM is to find the optimal treatment for each subtype. With the basis of our identified subtypes, clinical trial populations could now be adopted towards this vision. Enrichment of a clinical trial population with the rapid progressive subtype may help to show disease-modifying aspects of novel compounds, as this group is more likely to decline within typical trial periods of 1–2 years.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13167-024-00368-2>.

Acknowledgements Biosamples and data used in this work were generously provided by the participants in the Enroll-HD study and made available by CHDI Foundation, Inc. Enroll-HD is a clinical research platform and longitudinal observational study for Huntington's disease families intended to accelerate progress towards therapeutics; it is sponsored by CHDI Foundation, a non-profit biomedical research organization exclusively dedicated to collaboratively developing therapeutics for HD. Enroll-HD would not be possible without the vital contribution of the research participants and their families. We would like to thank all individuals who contributed to the collections of the Enroll-HD data as listed here.

Author contribution F.M., H.G., and H.F. were responsible for the study conception and design. Data analysis was performed by T.R. and Z.L. T.R., F.M., and H.F. contributed to the interpretation and discussion of the results. The first draft of the manuscript was written by T.R. F.M. and H.F. substantively revised it and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Tamara Raschka and Holger Fröhlich were supported by the ERA PerMed EU-wide project DIGIPD (01KU2110). Heiko Gaßner was supported by the Fraunhofer Internal Programs under Grant No. Attract 044-602140 and 044-602150. Franz Marxreiter is funded by the European Huntington's Disease Network (EHDN) Seed Fund program. The Department of Molecular Neurology at University Hospital Erlangen was supported by the ERA PerMed EU-wide project DIGIPD (01KU2110). The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Data availability The Enroll-HD database is available upon request from <https://www.enroll-hd.org>.

Declarations

Competing interests Holger Fröhlich has received grants from UCB and AbbVie outside this work. All other authors declare no competing interests.

Disclaimer The funders played no role in the study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. MacDonald ME, et al. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*. 1993;72:971–83. [https://doi.org/10.1016/0092-8674\(93\)90585-E](https://doi.org/10.1016/0092-8674(93)90585-E).
2. Lee J-M, et al. Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell*. 2015;162:516–26. <https://doi.org/10.1016/j.cell.2015.07.003>.
3. Langbehn DR, et al. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. *Clin Genet*. 2004;65:267–77. <https://doi.org/10.1111/j.1399-0004.2004.00241.x>.
4. Medina A, Mahjoub Y, Shaver L, Pringsheim T. Prevalence and incidence of Huntington's disease: an updated systematic review and meta-analysis. *Mov Disord Off J Mov Disord Soc*. 2022;37:2327–35. <https://doi.org/10.1002/mds.29228>.
5. Horta-Barba A, et al. Measuring cognitive impairment and monitoring cognitive decline in Huntington's disease: a comparison of assessment instruments. *J Neurol*. 2023;270:5408–17. <https://doi.org/10.1007/s00415-023-11804-0>.
6. Snowden JS. The neuropsychology of Huntington's disease. *Arch Clin Neuropsychol*. 2017;32:876–87. <https://doi.org/10.1093/arclin/acx086>.
7. Ho AK, Gilbert AS, Mason SL, Goodman AO, Barker RA. Health-related quality of life in Huntington's disease: which factors matter most? *Mov Disord*. 2009;24:574–8. <https://doi.org/10.1002/mds.22412>.
8. Banaszkiewicz K, et al. Huntington's disease from the patient, caregiver and physician's perspectives: three sides of the same coin? *J Neural Transm*. 2012;119:1361–5. <https://doi.org/10.1007/s00702-012-0787-x>.
9. Read J, et al. Quality of life in Huntington's disease: a comparative study investigating the impact for those with pre-manifest and early manifest disease, and their partners. *J Huntingt Dis*. 2013;2:159–75. <https://doi.org/10.3233/JHD-130051>.
10. Van Walsem MR, Howe EI, Ruud GA, Frich JC, Andelic N. Health-related quality of life and unmet healthcare needs in Huntington's disease. *Health Qual Life Outcomes*. 2017;15:6. <https://doi.org/10.1186/s12955-016-0575-7>.
11. Ross CA, Pantelyat A, Kogan J, Brandt J. Determinants of functional disability in Huntington's disease: role of cognitive and motor dysfunction. *Mov Disord Off J Mov Disord Soc*. 2014;29:1351–8. <https://doi.org/10.1002/mds.26012>.

12. Ross CA, et al. Huntington disease: natural history, biomarkers and prospects for therapeutics. *Nat Rev Neurol*. 2014;10:204–16. <https://doi.org/10.1038/nrneurol.2014.24>.
13. Penney JB, Vonsattel J, Macdonald ME, Gusella JF, Myers RH. CAG repeat number governs the development rate of pathology in Huntington's disease. *Ann Neurol*. 1997;41:689–92. <https://doi.org/10.1002/ana.410410521>.
14. Zhang Y, et al. Indexing disease progression at study entry with individuals at-risk for Huntington disease. *Am J Med Genet B Neuropsychiatr Genet*. 2011;156:751–63. <https://doi.org/10.1002/ajmg.b.31232>.
15. Orth M, et al. Observing Huntington's disease: the European Huntington's Disease Network's REGISTRY. *PLoS Curr*. 2011;2:RRN1184. <https://doi.org/10.1371/currents.RRN1184>.
16. Dorsey ER, the Huntington Study Group COHORT Investigators. Characterization of a large group of individuals with Huntington disease and their relatives enrolled in the COHORT study. *PLoS ONE*. 2012;7:e29522. <https://doi.org/10.1371/journal.pone.0029522>.
17. Hogarth P, et al. Interrater agreement in the assessment of motor manifestations of Huntington's disease. *Mov Disord*. 2005;20:293–7. <https://doi.org/10.1002/mds.20332>.
18. Reilmann R, Leavitt BR, Ross CA. Diagnostic criteria for Huntington's disease based on natural history. *Mov Disord*. 2014;29:1335–41. <https://doi.org/10.1002/mds.26011>.
19. Mandel S, Kerczyn AD. In *Neurodegener. Dis. Integr. PPPM Approach Med. Future* (ed. Mandel, S.) 2013;2:95–140. Springer Netherlands. https://doi.org/10.1007/978-94-007-5866-7_6
20. Tabrizi SJ, et al. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *Lancet Neurol*. 2009;8:791–801. [https://doi.org/10.1016/S1474-4422\(09\)70170-X](https://doi.org/10.1016/S1474-4422(09)70170-X).
21. Abeyasinghe PM, et al. Tracking Huntington's disease progression using motor, functional, cognitive, and imaging markers. *Mov Disord*. 2021;36:2282–92. <https://doi.org/10.1002/mds.28650>.
22. Tabrizi SJ, et al. A biological classification of Huntington's disease: the Integrated Staging System. *Lancet Neurol*. 2022;21:632–44. [https://doi.org/10.1016/S1474-4422\(22\)00120-X](https://doi.org/10.1016/S1474-4422(22)00120-X).
23. Mohan A, et al. A machine-learning derived Huntington's disease progression model: insights for clinical trial design. *Mov Disord*. 2022;37:553–62. <https://doi.org/10.1002/mds.28866>.
24. EPMA, Golubnitschaja O, Costigliola V. General report & recommendations in predictive, preventive and personalised medicine 2012: white paper of the European Association for Predictive, Preventive and Personalised Medicine. *EPMA J*. 2012;3:14. <https://doi.org/10.1186/1878-5085-3-14>
25. Golubnitschaja O, Kinkorova J, Costigliola V. Predictive, preventive and personalised medicine as the hardcore of 'Horizon 2020': EPMA position paper. *EPMA J*. 2014;5:6. <https://doi.org/10.1186/1878-5085-5-6>.
26. Wang M, et al. Statistical methods for studying disease subtype heterogeneity. *Stat Med*. 2016;35:782–800. <https://doi.org/10.1002/sim.6793>.
27. Giannoula A, Gutierrez-Sacristán A, Bravo Á, Sanz F, Furlong LI. Identifying temporal patterns in patient disease trajectories using dynamic time warping: A population-based study. *Sci Rep*. 2018;8:4216. <https://doi.org/10.1038/s41598-018-22578-1>.
28. Birkenbihl C, et al. Artificial intelligence-based clustering and characterization of Parkinson's disease trajectories. *Sci Rep*. 2023;13:2897. <https://doi.org/10.1038/s41598-023-30038-8>.
29. Golubnitschaja O, et al. Medicine in the early twenty-first century: paradigm and anticipation - EPMA position paper 2016. *EPMA J*. 2016;7:23. <https://doi.org/10.1186/s13167-016-0072-4>.
30. de Jong J, et al. Deep learning for clustering of multivariate clinical patient trajectories with missing values. *GigaScience*. 2019;8:giz134. <https://doi.org/10.1093/gigascience/giz134>.
31. Krishnagopal S, Coelln RV, Shulman LM, Girvan M. Identifying and predicting Parkinson's disease subtypes through trajectory clustering via bipartite networks. *PLoS ONE*. 2020;15: e0233296. <https://doi.org/10.1371/journal.pone.0233296>.
32. Marder K, et al. Rate of functional decline in Huntington's disease. *Neurology*. 2000;54:452–452. <https://doi.org/10.1212/WNL.54.2.452>.
33. Kühnel L, Berger A-K, Markussen B, Raket LL. Simultaneous modeling of Alzheimer's disease progression via multiple cognitive scales. *Stat Med*. 2021;40:3251–66. <https://doi.org/10.1002/sim.8932>.
34. Koval I, et al. Forecasting individual progression trajectories in Huntington disease enables more powered clinical trials. *Sci Rep*. 2022;12:18928. <https://doi.org/10.1038/s41598-022-18848-8>.
35. Li D, Iddi S, Thompson WK, Donohue MC, Alzheimer's Disease Neuroimaging Initiative. Bayesian latent time joint mixed effect models for multicohort longitudinal data. *Stat Methods Med Res*. 2019;28:835–45. <https://doi.org/10.1177/0962280217737566>.
36. Raket LL. Statistical disease progression modeling in Alzheimer disease. *Front Big Data*. 2020;3:24. <https://doi.org/10.3389/fdata.2020.00024>.
37. Landwehrmeyer GB, et al. Data analytics from Enroll-HD, a global clinical research platform for Huntington's disease. *Mov Disord Clin Pract*. 2017;4:212–24. <https://doi.org/10.1002/mdc3.12388>.
38. Sathe S, et al. Enroll-HD: an integrated clinical research platform and worldwide observational study for Huntington's disease. *Front Neurol*. 2021;12: 667420. <https://doi.org/10.3389/fneur.2021.667420>.
39. Kiebertz K, Penney JB, Como P, Ranen N, Shoulson I. Unified Huntington's Disease Rating Scale: reliability and consistency. Huntington Study Group. *Mov Disord Off J Mov Disord Soc*. 1996;11:136–42. <https://doi.org/10.1002/mds.870110204>.
40. Raket LL. progmod. 2020. at <<https://github.com/larslau/progmod>>.
41. Pinheiro J, Bornkamp B, Glimm E, Bretz F. Model-based dose finding under model uncertainty using general parametric models. *Stat Med*. 2014;33:1646–61. <https://doi.org/10.1002/sim.6052>.
42. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32. <https://doi.org/10.1023/A:1010933404324>.
43. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 785–794 (ACM, 2016). <https://doi.org/10.1145/2939672.2939785>
44. Pedregosa F, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825–30.
45. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. Preprint at <https://doi.org/10.48550/arXiv.1705.07874> (2017) <https://doi.org/10.48550/arXiv.1705.07874>
46. Mestre TA, et al. Rating scales for cognition in Huntington's disease: critique and recommendations. *Mov Disord Off J Mov Disord Soc*. 2018;33:187–95. <https://doi.org/10.1002/mds.27227>.
47. Mestre TA, et al. Rating scales for motor symptoms and signs in Huntington's disease: critique and recommendations. *Mov Disord Clin Pract*. 2018;5:111–7. <https://doi.org/10.1002/mdc3.12571>.
48. Chao T-K, Hu J, Pringsheim T. Risk factors for the onset and progression of Huntington disease. *Neurotoxicology*. 2017;61:79–99. <https://doi.org/10.1016/j.neuro.2017.01.005>.
49. McAllister B, et al. Timing and impact of psychiatric, cognitive, and motor abnormalities in Huntington disease. *Neurology*. 2021;96:e2395–406. <https://doi.org/10.1212/WNL.0000000000001893>.
50. Tabrizi SJ, et al. Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *Lancet Neurol*. 2013;12:637–49. [https://doi.org/10.1016/S1474-4422\(13\)70088-7](https://doi.org/10.1016/S1474-4422(13)70088-7).

51. Georgiou-Karistianis N, Scahill R, Tabrizi SJ, Squitieri F, Aylward E. Structural MRI in Huntington's disease and recommendations for its potential use in clinical trials. *Neurosci Biobehav Rev.* 2013;37:480–90. <https://doi.org/10.1016/j.neubiorev.2013.01.022>.
52. Byrne LM, et al. Neurofilament light protein in blood as a potential biomarker of neurodegeneration in Huntington's disease: a retrospective cohort analysis. *Lancet Neurol.* 2017;16:601–9. [https://doi.org/10.1016/S1474-4422\(17\)30124-2](https://doi.org/10.1016/S1474-4422(17)30124-2).

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

Authors and Affiliations

Tamara Raschka¹  · Zexin Li^{1,2} · Heiko Gaßner^{3,4}  · Zacharias Kohl⁵ · Jelena Jukic^{3,7} · Franz Marxreiter^{3,6,7}  · Holger Fröhlich^{1,2} 

✉ Holger Fröhlich
holger.froehlich@scai.fraunhofer.de

¹ Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53757 Sankt Augustin, Germany

² Bonn-Aachen International Center for IT, University of Bonn, Friedrich-Hirzebruch-Allee 6, 53115 Bonn, Germany

³ Department of Molecular Neurology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany

⁴ Fraunhofer IIS, Fraunhofer Institute for Integrated Circuits IIS, Am Wolfsmantel 33, 91058 Erlangen, Germany

⁵ Department of Neurology, University of Regensburg, Regensburg, Germany

⁶ Center for Movement Disorders, Passauer Wolf, 93333 Bad Gögging, Germany

⁷ Center for Rare Diseases Erlangen (ZSEER), University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany