A novel machine learning based framework for developing composite digital biomarkers of disease progression

Song Zhai[1], Andy Liaw[1], Judong Shen[1], Yuting Xu[1], Vladimir Svetnik[1], James J. FitzGerald[2,3], Chrystalina A. Antoniades[2], Dan Holder[1], Marissa F. Dockendorf[4], Jie Ren[4]* and Richard Baumgartner[1]*

[1]Biostatistics and Research Decision Sciences, Merck & Co., Inc., Rahway, NJ, United States, [2]NeuroMetrology Lab, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom, [3]Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom, [4]Digital Clinical Measures, Translational Medicine, Merck & Co., Inc., Rahway, NJ, United States

Background: Current methods of measuring disease progression of neurodegenerative disorders, including Parkinson's disease (PD), largely rely on composite clinical rating scales, which are prone to subjective biases and lack the sensitivity to detect progression signals in a timely manner. Digital health technology (DHT)-derived measures offer potential solutions to provide objective, precise, and sensitive measures that address these limitations. However, the complexity of DHT datasets and the potential to derive numerous digital features that were not previously possible to measure pose challenges, including in selection of the most important digital features and construction of composite digital biomarkers.

Methods: We present a comprehensive machine learning based framework to construct composite digital biomarkers for progression tracking. This framework consists of a marginal (univariate) digital feature screening, a univariate association test, digital feature selection, and subsequent construction of composite (multivariate) digital disease progression biomarkers using Penalized Generalized Estimating Equations (PGEE). As an illustrative example, we applied this framework to data collected from a PD longitudinal observational study. The data consisted of Opal^{TM} sensor-based movement measurements and MDS-UPDRS Part III scores collected at 3-month intervals for 2 years in 30 PD and 10 healthy control participants.

Results: In our illustrative example, 77 out of 235 digital features from the study passed univariate feature screening, with 11 features selected by PGEE to include in construction of the composite digital measure. Compared to MDS-UPDRS Part III, the composite digital measure exhibited a smoother and more significant increasing trend over time in PD groups with less variability, indicating improved ability for tracking disease progression. This composite digital measure also demonstrated the ability to classify between *de novo* PD and healthy control groups.

Conclusion: Measures from DHTs show promise in tracking neurodegenerative disease progression with increased sensitivity and reduced variability as compared to traditional clinical scores. Herein, we present a novel framework and methodology to construct composite digital measure of disease progression from high-dimensional DHT datasets, which may have utility in accelerating the development and application of composite digital biomarkers in drug development.

1 Introduction

Neurodegenerative diseases, including Parkinson's Disease (PD), are an area of vast unmet medical need. Drug development efforts in this area have increasingly focused on the search for disease-modifying therapies that slow down the underlying disease progression mechanisms. However, a lack of validated measures that allow for disease progression to be monitored objectively, relatively rapidly, and with high precision makes it challenging to effectively demonstrate therapeutic efficacy and hinders drug development efforts. PD clinical trials generally use the Movement Disorder Society—Unified Parkinson's Disease Rating Scale (MDS-UPDRS) to track disease progression longitudinally. However, MDS-UPDRS is subjective in nature, relies on patient and caregiver-reported symptoms and clinician's qualitative ratings (1), is slow to change, and has low measurement precision, resulting in large and lengthy clinical trials to test efficacy for potential disease modifying therapies (2).

Recent advances in digital health technologies (DHTs) offer unprecedented opportunities to collect more objective, precise, and sensitive measures, both in the clinic and remotely, that were out of reach in the past. Such measures could provide new insights into neurogenerative disease progression, including for Parkinson's disease. There are many studies that have investigated using measures from sensor-based digital health technologies in neurodegenerative diseases (3–11). These studies have collectively demonstrated that many neurodegenerative disease symptoms can be quantified by DHTs. Moreover, multiple longitudinal observational studies have shown that digital measures can pick up changes over time that are indicative of disease progression (12–18). It is further thought that the objective measures enabled by DHTs could offer improved sensitivity and reduced variability (12, 19), which could translate to smaller and shorter clinical trial designs (20) and, in turn, potential for accelerated drug development. Despite promising results, the longitudinal studies published to date have used different DHTs and analysis methodologies to identify the digital features of importance and to derive respective digital clinical measures, making it difficult to compare across studies or create consensus among the research community. Open discussions on the methodology of digital clinical measure development and evaluation are critically needed to move the field forward.

It has been increasingly recognized that composite digital measures, rather than reliance on individual digital features, are needed for more effective measurement of disease progression as compared to traditional clinical composite scores. Adams et al. (21) showed that no individual digital feature (from gait, tremor, turns, speech, and cognition) outperformed MDS-UPDRS Part III (a composite clinical score) in terms of the standardized change from baseline after 12 months in a PD observational study (WATCH-PD). Furthermore, Czech et al. (22) demonstrated individual sensor-based digital features of upper and lower extremity bradykinesia often lacked strong sensitivity to longitudinal changes, whereas digital composite scores showed significant differences over 12 months in WATCH-PD.

There have been several examples where composite digital measures were developed for disease classification and/or tracking symptom progression (22–30); however, the approach taken has varied, and there have been limited discussions on the methodologies to effectively select informative digital features and construct the most performant composite measures. For example, Perumal and Sankar (23) developed a Linear Discriminant Analysis (LDA) classifier using multiple gait features collected from wearable sensors to distinguish between PD patients and healthy control (HC) subjects. Czech et al. (22) constructed composite digital scores using pre-defined combinations of features from single tasks (pronation-supination and toe-tapping) and used them to measure longitudinal progression of bradykinesia after 1 year. Sotirakis et al. (30) developed a Random Forest model to estimate the MDS-UPDRS III values using gait and sway features and used the model to detect progression of motor symptoms longitudinally. These efforts vary in terms of the measure construction (pre-defined vs. supervised ML, choice of models), the clinical label selection (MDS-UPDRS III total score or single item), the selection of digital tasks (single task e.g., toe-tapping or a combination of tasks), as well as the selection of input features (e.g., whether features are pre-screened). Overall, the field has not adopted consistent and systematic methods and/or analysis frameworks. Therefore, there is an urgent need to develop methodologies and analysis pipelines for the construction of composite digital measures for disease progression tracking, tailored for high-dimensional, longitudinal data with digital features collected from sensor technologies.

The types of data generated by DHTs are often longitudinal and high dimensional, which differs from traditional clinical measures, calling for novel analytical strategies to handle such data for the construction of composite digital measures. Unlike traditional clinical measures that collect a defined set of measures at each time point, DHTs leverage various sensors to generate large amounts of time-series data (e.g., acceleration, screen touch, audio/video, keyboard press), either collected from defined active task-based assessments or from passive monitoring. Such data are often not readily analysable statistically and need to be aggregated and transformed into digital features first. For example, for measurement of physical activity, continuous accelerometer signals are often converted to epoch level activity counts and then aggregated over time into features such as daily total activity count, total steps, non-sedentary time, etc., for further statistical analysis. There can be large numbers of features derived from the high-frequency sensor signals; such features may have various data types (i.e., categorical, continuous, duration, etc.) and clinimetric properties, many of which may not yet have been fully explored as it was not previously possible to measure them without use of DHTs. These features could have intrinsic skewness in distribution, floor/ceiling effects, as well as unknown redundancies and covariances. In addition, the high frequency nature of DHT data collection and potential for remote data acquisition can also lend itself to higher levels of data missingness. Furthermore, not all digital features that can be generated from sensor data may have clinical significance or be valuable for creating composite digital measures. These attributes of DHT data make it a unique

challenge in the development of composite digital measures to track longitudinal disease progression.

Machine learning (ML) methods offer a valuable tool for selecting the most informative digital features to reflect disease progression and to construct clinically meaningful composite digital measures. ML-based techniques can often improve prediction performance in analysing digital data in neurodegenerative diseases; however, existing ML methodologies for longitudinal data analysis are also challenged by the high dimensionality of DHT data. For example, although the generalized estimating equations (GEE) method (31) incorporating different patterns of working correlation matrix across multiple timepoints has been widely used in longitudinal data analysis, the direct use of classical unpenalized GEE in high-dimensional longitudinal data analysis may lead to misleading results (32). To address this, an ML-based penalized GEE (PGEE) method (32) could be used to improve upon the GEE method in handling DHT data. PGEE performs simultaneous coefficient estimation and variable selection for longitudinal data analysis with high-dimensional covariates by including a penalty term in the GEE model, which can be better-suited to handle high-dimensional feature sets.

In this paper, we propose a principled, scalable, and comprehensive methodology framework for the development of novel composite digital biomarkers, derived from DHT data and anchored to the MDS-UPDRS score, to measure neurodegenerative disease progression. This framework includes data processing, univariate digital feature screening, multivariate (composite) digital biomarker construction (using PGEE methods), and composite biomarker performance evaluation.

We further demonstrate the utility of this framework by applying it to a sample dataset containing high-dimensional, longitudinal movement data collected by a body-worn accelerometer system from a PD longitudinal observation study. The current analytical challenges of high-dimensional and longitudinal digital data and path forward for the application of composite digital biomarkers in measurement of neurodegenerative disease progression are also discussed.

2 Materials and methods

2.1 Study overview

To illustrate our proposed methodology to construct composite digital measures for tracking longitudinal disease progression, we applied the framework to data from 30 PD patients (10 *de novo* PD patients, 10 mild-to-moderate PD patients on levodopa, and 10 advanced PD patients) and 10 healthy control subjects from a PD longitudinal observational study conducted at John Radcliffe Hospital in Oxford, UK (11, 30, 33). The participants visited the clinic once every 3 months for 2 years. At each visit, they wore six synchronized inertial measurement units (IMUs) ("Opal" sensors, APDM Wearable Technologies, a Clario Company) across their body and performed two-minute walk, postural sway, and timed up-and-go (TUG) tasks. The Mobility Lab^{TM} software (APDM Wearable Technologies, a Clario Company) was then

used to process these raw sensor signals and generate epoch-level digital features at each instance of a time period or physical movement (e.g., per minute, per step, per turn, or per sit-to-stand event). The MDS-UPDRS Part III assessments were also conducted at these clinic visits. The MDS-UPDRS Part III score and subscales (including Bradykinesia, Postural and Gait, Rigidity, and Tremor, defined in Supplementary Table S1) were calculated. Demographic data including age and sex of the participants were also collected at the beginning of the study.

2.2 Statistical analysis

The workflow of our proposed comprehensive machine learning based framework is illustrated in Figure 1, which comprises five main steps: (1) data collection and processing; (2) univariate feature screening; (3) univariate association testing; (4) multivariate analysis (using PGEE) to construct a composite digital measure for longitudinal disease progression; (5) performance evaluation. The specifics of each step are described below.
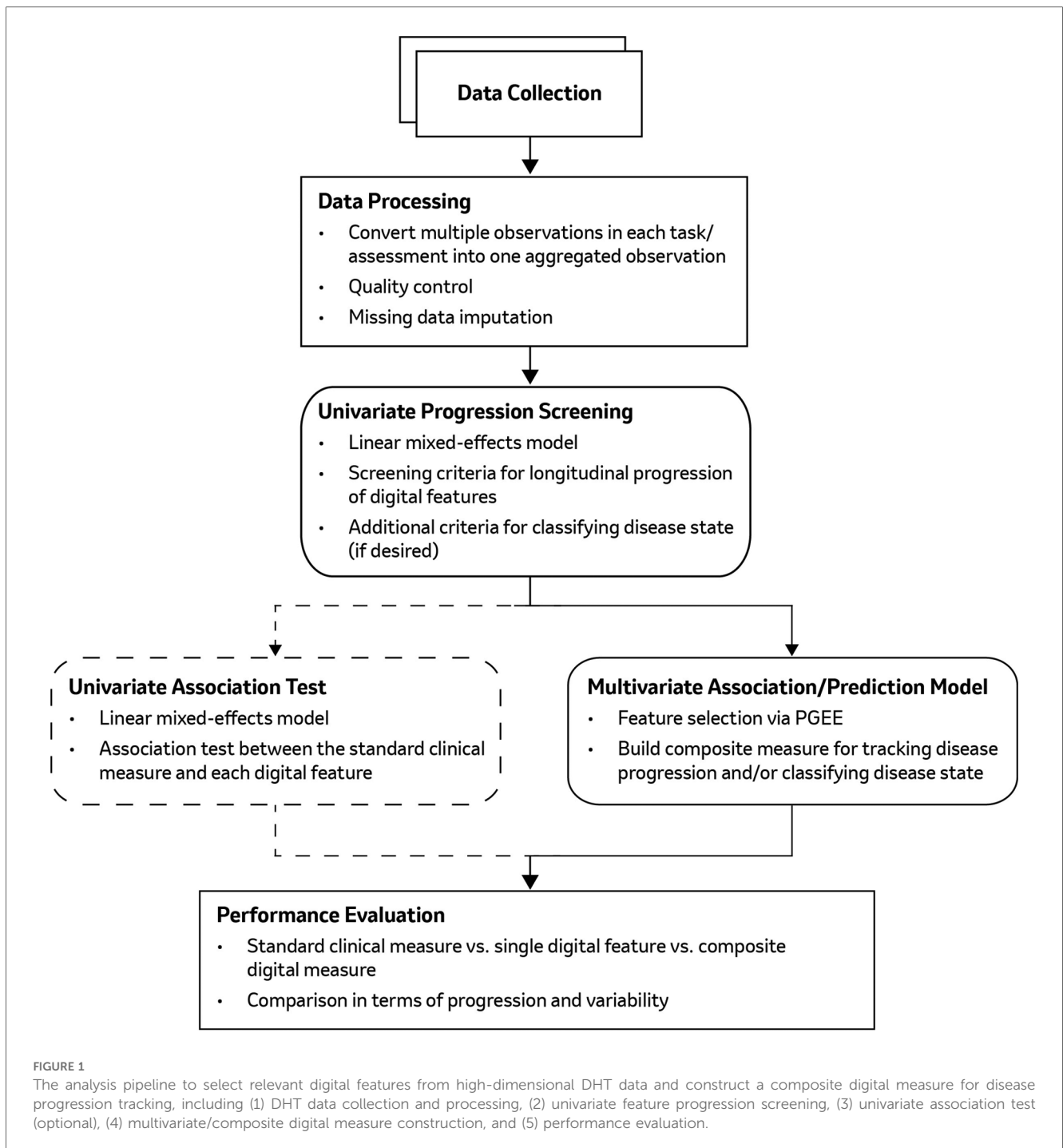
2.2.1 Data processing and quality control

In this first step, data aggregation and pre-processing are performed to convert high-frequency, epoch-level data into a set of aggregated digital features for each task. The movement data collected from DHTs often include epoch-level features (e.g., per second, per minute, or per walking step) that are collected repeatedly during an active task (e.g., two-minute walk). This step simplifies such data and produces a clean, high-dimensional feature set for each participant at each clinical time point, in order to facilitate subsequent longitudinal analyses.

In our illustrative PD example, summary statistics (mean, median, standard deviation, and mean absolution deviation) were calculated to represent the repeated measurements across the entire task for features that had repeated measurements during the task. For example, during the two-minute walk task, step lengths of every step that the participant took were recorded; these were aggregated into task-level features such as mean step length during the two-minute walk task period. After that, we had 256 digital features generated in total. Then, distributions of all features were examined, and the non-informative features that had few distinct values, included a large amount of data missingness, or contained extreme values were removed. For the remaining features, missing data imputation was performed using the mean of available data in each feature. Finally, additional feature quality control steps were implemented, which included removing highly correlated features, log-transforming skewed features, and removing outliers. 141 unique digital features were left for univariate progression screening in the next step.

2.2.2 Univariate progression screening

In the second step of our framework, univariate progression screening is recommended to identify whether each digital feature detected disease progression during the study duration. In this step, a linear mixed effects model (LMM) is used to screen the univariate features against a set of pre-determined criteria. Each

FIGURE 1
The analysis pipeline to select relevant digital features from high-dimensional DHT data and construct a composite digital measure for disease progression tracking, including (1) DHT data collection and processing, (2) univariate feature progression screening, (3) univariate association test (optional), (4) multivariate/composite digital measure construction, and (5) performance evaluation.

digital feature is used as the response variable for the screening separately. Independent variables are added to the model as fixed effects, including covariates to be adjusted, group membership, visit, group-by-visit interaction, and covariate-by-visit interactions. Random intercept and slope are added to the model as random effects.

In our illustrative PD example, we applied relatively relaxed screening criteria to select digital features for downstream analysis. We considered a digital feature as a "candidate" if (1) its longitudinal trend was flat in the HC group (i.e., the LMM slope p-value of HC group was larger than 0.05) and (2) it

demonstrated a progression trend with time in PD groups (i.e., the LMM group-by-visit interaction p-value was <0.1 or the p-value of the differential slope between *de novo*/mild-to-moderate/advanced PD and HC was <0.1).

2.2.3 Univariate association test

To gain additional insights on the univariate associations between the standard clinical measure (i.e., MDS-UPRDS Part III) and the candidate digital features that passed the univariate progression screening, our framework employs a univariate association test step. In this step, a linear mixed effects model is

employed, with the clinical measure as the dependent variable and each individual digital feature as the independent variable. Covariates to be adjusted are also included in the model. Random intercepts for each subject are allowed in the model and p-values are calculated to assess the significance of the association between the clinical measure and digital features.

An optional procedure is to further filter the candidate digital features based on their associations to the standard reference measure (i.e., MDS-UPDRS Part III and its subscales in our example) and exclude non-significant features. In our example, we chose to implement relatively relaxed screening criteria to retain more features for the subsequent feature selection, and therefore, we did not exclude features that did not show association with MDS-UPDRS Part III in our downstream analysis.

2.2.4 Multivariate prediction model

In the final step of our framework, a multivariate prediction model is developed to select a subset of digital features from those that passed the univariate progression screening and combine them into a composite digital biomarker of disease progression.

For feature selection, we used the longitudinal MDS-UPDRS Part III data as the training endpoint in our illustrative example. Additionally, we included features that were important for classifying the *de novo* PD cohort from healthy controls in the feature selection process since patient identification could also be an important attribute for the composite digital biomarker. Importantly, depending on the intended context-of-use of the developed measure, one could use our proposed framework to optimize the measure for disease progression tracking, or patient identification, or both, by adjusting the screening criteria and the training endpoints used.

To model the high-dimensional longitudinal data, our framework includes a ML-based Penalized Generalized Estimating Equations (PGEE) method (32), which performs simultaneous coefficient estimation and variable selection. Compared to the traditional GEE method, PGEE introduces a penalty term to the estimating function of GEE (details of PGEE is provided in Supplementary Method S1).

To determine the optimal number of digital features (P) to be included into the final multivariate prediction model, a cross-validation (CV) strategy is implemented into the framework to avoid overfitting (34). Specifically, all digital features are first ranked by their PGEE estimates from the training set, then a series of PGEE models with different numbers of top features are built and evaluated in the testing set. The optimal number of features is then determined to be the number of features from the model with the smallest Root Mean Squared Error (RMSE). The approach is further described in Supplementary Method S2.

Once P is determined, the PGEE estimates of the digital features that pass the univariate screening are calculated again using the whole dataset, and the top P features with the largest PGEE estimates were selected. Two sets of digital features were identified based on a PGEE model for MDS-UPDRS Part III progression and another PGEE model for *de novo* PD vs. Healthy Control classification, respectively. These two feature sets

were merged into a final feature set for the composite digital measure construction.

A GEE model is then fitted, with this final feature set plus the covariates as independent variables, and MDS-UPDRS Part III as a continuous dependent variable. This generates our final composite digital measure for performance evaluations.

3 Results

3.1 Patient demographics and baseline characteristics

The baseline demographic characteristics for the participants included in our illustrative analysis are shown in Table 1 and Supplementary Figure S1. The mean ages of four groups (*de novo* PD, mild-to-moderate PD, advanced PD, and HC) were 66.2, 61.6, 71.2, and 65.6 years, respectively. The ratios of male-to-female subjects in the four groups were 5:5, 9:1, 5:5, and 3:7, respectively.

To determine if age and sex needed be considered covariates to be adjusted for in our models, we calculated the age-by-visit and sex-by-visit interaction p-values in linear mixed effects models with MDS-UPDRS Part III as the response in the pooled PD group. The results, summarized in Supplementary Table S2, suggested that age would affect the slope of MDS-UPDRS Part III progression (with p-value = 0.04) while sex would not (with p-value = 0.19). We therefore considered only age as a covariate to be adjusted in our data analysis models.

3.2 Univariate progression screening results

In our illustrative example, our univariate progression screening criteria were such that a digital feature would "pass" if the LMM model for that digital feature showed (1) no progression in the control group and (2) a progression in at least one of the three PD groups. 77 digital features out of 141 screened passed these criteria, including 15 features from postural sway task, 5 features from timed up-and-go (TUG) task, and 57 features from two-minute walk task. Among these, Walk GLLGS (Gait—Lower Limb—Gait Speed) had the smallest group-by-visit interaction p-value (6.0×10^{-7}) and the smallest *de novo* PD vs. HC progression slope p-value (4.7×10^{-4}); Walk GLLDS

TABLE 1 Patient baseline characteristics (age and sex) for the three PD groups and healthy control group.

	de novo PD	Mild-to-moderate PD (on-therapy)	Advanced PD	Healthy Control
N	10	10	10	10
Age (years) [mean (SD)]	66.2 (6.46)	61.6 (10.76)	71.2 (4.78)	65.6 (6.98)
Sex [Male (%)/Female (%)]	5 (50)/5 (50)	9 (90)/1 (10)	5 (50)/5 (50)	3 (30)/7 (70)

(Gait—Lower Limb—Double Support) had the smallest mild-to-moderate PD vs. HC progression slope p-value (0.01); and Walk GLLSD (Gait—Lower Limb—Step Duration) had the smallest advanced PD vs. HC progression p-value (1.2×10^{-6}). P-values of TUG TPV (Timed Up and Go—Turn Peak Velocity) for group-by-visit interaction, *de novo* PD vs. HC progression slope, mild-to-moderate PD vs. HC progression slope, and advanced PD vs. HC progression slope were 0.008, 0.001, 0.147, and 0.013, respectively. A summary heatmap of all 77 digital features that met the screening criteria is displayed in Figure 2, and the heatmap of all the digital features that were screened is displayed Supplementary Table S3.

3.3 Univariate association analysis results

Figure 3 shows the univariate association testing results between the 77 digital features that passed the univariate screening in our illustrative example and MDS-UPDRS Part III scores (and its subscales). 37 of these 77 digital features (48.1%) showed significant associations (i.e., p-value < 0.05) with MDS-UPDRS Part III scores (including 32 features from the Walk task, 3 features from the TUG task, and 2 features from the Sway task). The associations of digital features with the MDS-UPDRS Part III scores were generally consistent with their associations with the Bradykinesia (BK) subscale within MDS-UPDRS Part III. Specifically, 40 of the 77 digital features were associated with the BK subscale (including 31 features from the Walk task, 3 features from the TUG task, and 6 features from the Sway task). In addition, 59 of the 77 digital features were associated with the Postural Instability and Gait (PIGD) subscale (including 54 features from the Walk task, 4 features from the TUG task, and 1 feature from the Sway task), while only 3 of the 77 features (TUG TPV, TUG TA, and Walk GULMV) were associated with the Tremor Dominant (TD) subscale.

Turn Peak Velocity (TPV), obtained from the Timed Up and Go (TUG) test (35), demonstrated the most significant association with MDS-UPDRS Part III. TUG TPV is defined as the maximum achieved angular velocity of trunk rotation in the y-axis during 180-degree turns (deg/sec) and has been found to be related to PD progression in multiple studies (12, 36–38). The progression characteristics of TUG TPV are shown in Figure 4, where the group-wise and subject-wise lines were obtained from the linear mixed effect model and the points represented the observed data. In terms of TUG TPV, the mild-to-moderate, on therapy PD and HC groups were stable, while the *de novo* and advanced PD groups showed progression.

In general, the univariate association observations were consistent with the progression patterns seen in the MDS-UPDRS Part III and its subscales, which is shown in Supplementary Figure S2. Specifically, compared to the HC group, the BK subscale progressed across all PD groups (at $\alpha = 0.1$ level). The PIGD subscale progressed in *de novo* and advanced PD groups while staying stable in the mild-to-moderate, on-therapy PD group. This pattern was similar to most of the digital features included in the analysis, as indicated

in Figure 2. In contrast, the TD subscale progressed in the mild-to-moderate, on-therapy PD group, while remaining unchanged in *de novo* and advanced PD groups.

3.4 Multivariate feature selection and prediction results

3.4.1 Feature selection

We first conducted multivariate feature selection to determine the optimal number of features to be selected for inclusion into the composite score and prediction model in our illustrative example analysis. Supplementary Figure S3 indicated that for developing a composite digital measure for disease progression tracking, using 9 top features (ranked by their PGEE estimates in training sets during cross-validation) overall yielded the smallest RMSE; and for classifying disease status, using 3 top features resulted in the largest AUC via internal cross-validation.

We then ranked all pre-screened features (i.e., digital features) according to their PGEE estimates in the whole dataset supervised by the continuous endpoint (i.e., MDS-UPDRS Part III) and the binary endpoint (i.e., *de novo* PD vs. HC), respectively. Nine digital features (TUG TD, TUG TPV, TUG STSD, Walk TA, Walk GLLC, Walk GLLSW, Walk GLLLSM, Walk APAMAA, and Sway PSAN95ESA) were selected for disease progression tracking; additionally, three digital features (TUG TPV, Walk GLLTOA, and Walk GULMV) were selected for PD vs. Control classification. Table 2 lists the description of these selected features. The two sets of digital features were further merged; since one of the features (TUG TPV) was in both feature sets, 11 unique digital features were included in the final feature set. This feature set was then used to create the composite digital biomarker by fitting a GEE model.

3.4.2 Composite digital biomarker for tracking MDS-UPDRS part III

The performance of the composite digital biomarker was evaluated using 10-fold cross-validation in PD and HC groups, respectively. As shown in Figure 5, the composite digital measure showed a pattern of no change vs. time in the HC group as expected (with RMSE in HC group = 2.8). On the other hand, it had a smoother increasing trend in the overall PD group, as well as each PD subgroup (with RMSE in PD group = 12.7).

We further compared performances among MDS-UPDRS Part III, the composite digital measure, and each of the univariate digital features included in the composite digital measure (e.g., TUG TPV) quantitatively in terms of both progression and variability. Detailed results are summarized in Table 3. Overall, the group-by-visit interaction p-value of the composite digital measure was close to that of MDS-UPDRS Part III (7.65×10^{-3} vs. 6.22×10^{-3}). The increasing trend of the composite digital measure was much more significant compared to MDS-UPDRS Part III and individual digital feature TUG TPV in *de novo* and advanced PD groups. Specifically, for *de novo* PD vs. HC, the effect sizes of progression slope were 1.41, 2.14, and 1.37 for MDS-UPDRS Part III, composite digital measure, and TUG TPV,
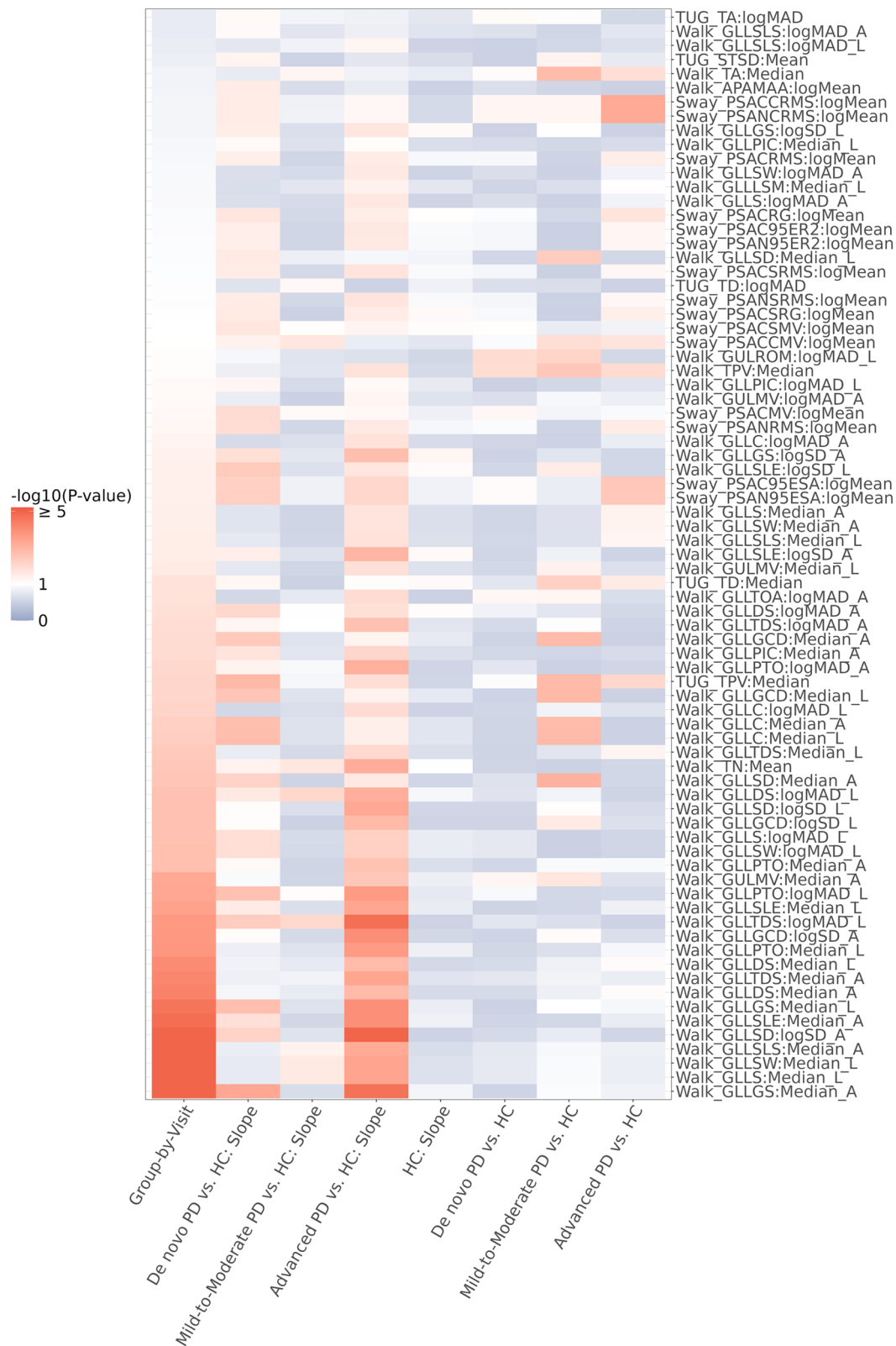
FIGURE 2

Heatmap representation of the *p*-values of the 77 digital features that passed the progression screening. The screening criteria applied were (1) no time progression in the HC group (i.e., LMM slope *p*-value of HC > 0.05), and (2) time progression in at least one of the three PD groups (i.e., LMM group-by-visit interaction *p*-value < 0.1 or *p*-value of differential slope between *de novo*/mild-to-moderate/advanced PD and HC < 0.1).
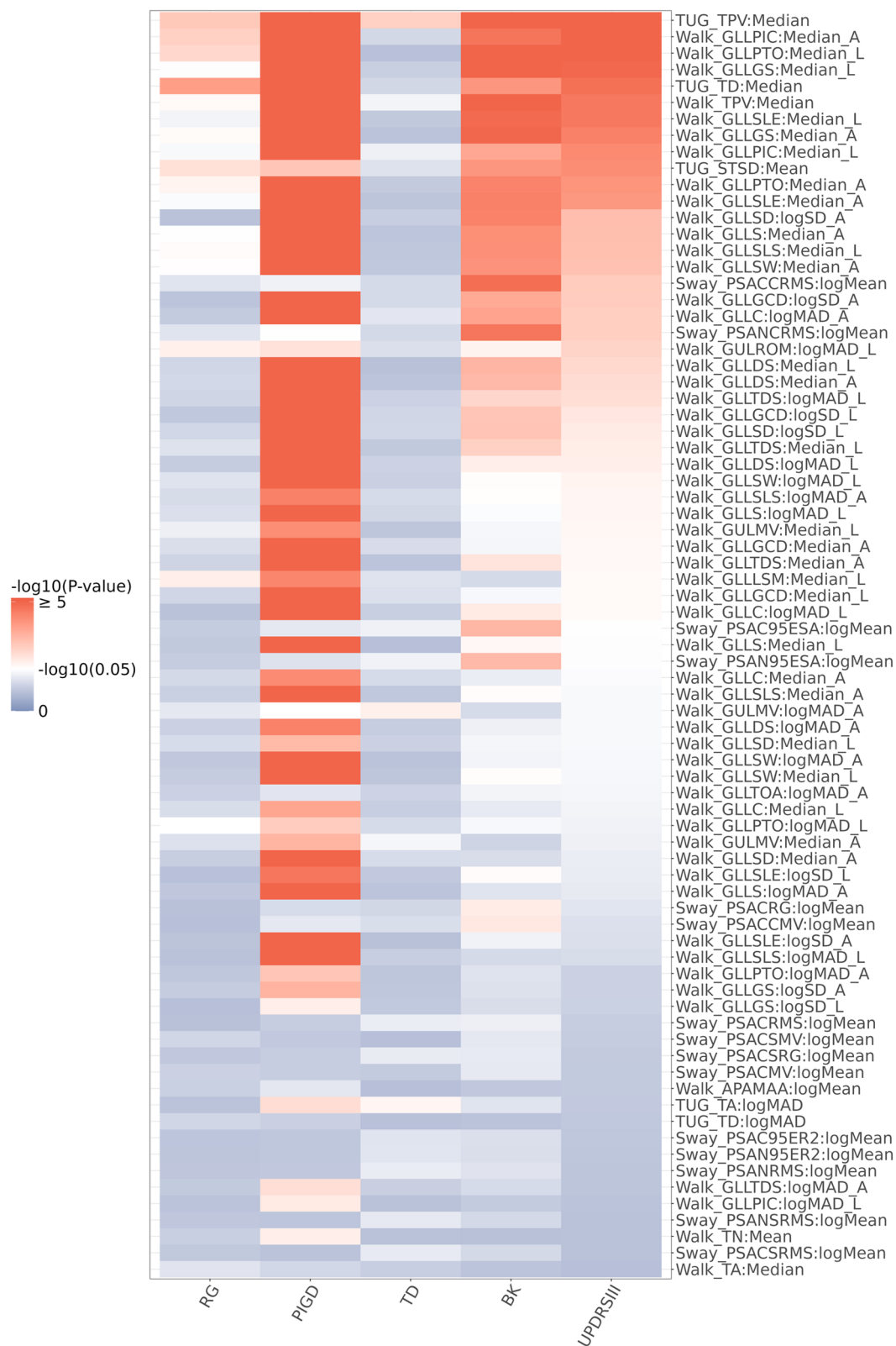
FIGURE 3
Heatmap of the univariate association testing *p*-values between MDS-UPDRS part III (and its subscales: BK, TD, PIGD, RG) and the 77 digital features that passed the univariate screening. *P*-values were calculated from a linear mixed effects model with MDS-UPDRS Part III or its subscales as the outcome variable. The 77 features were ranked based on their association *p*-values from the analysis with the MDS-UPDRS Part III score. Each digital feature and age were included as independent variables. Random intercept was added as a random effect.
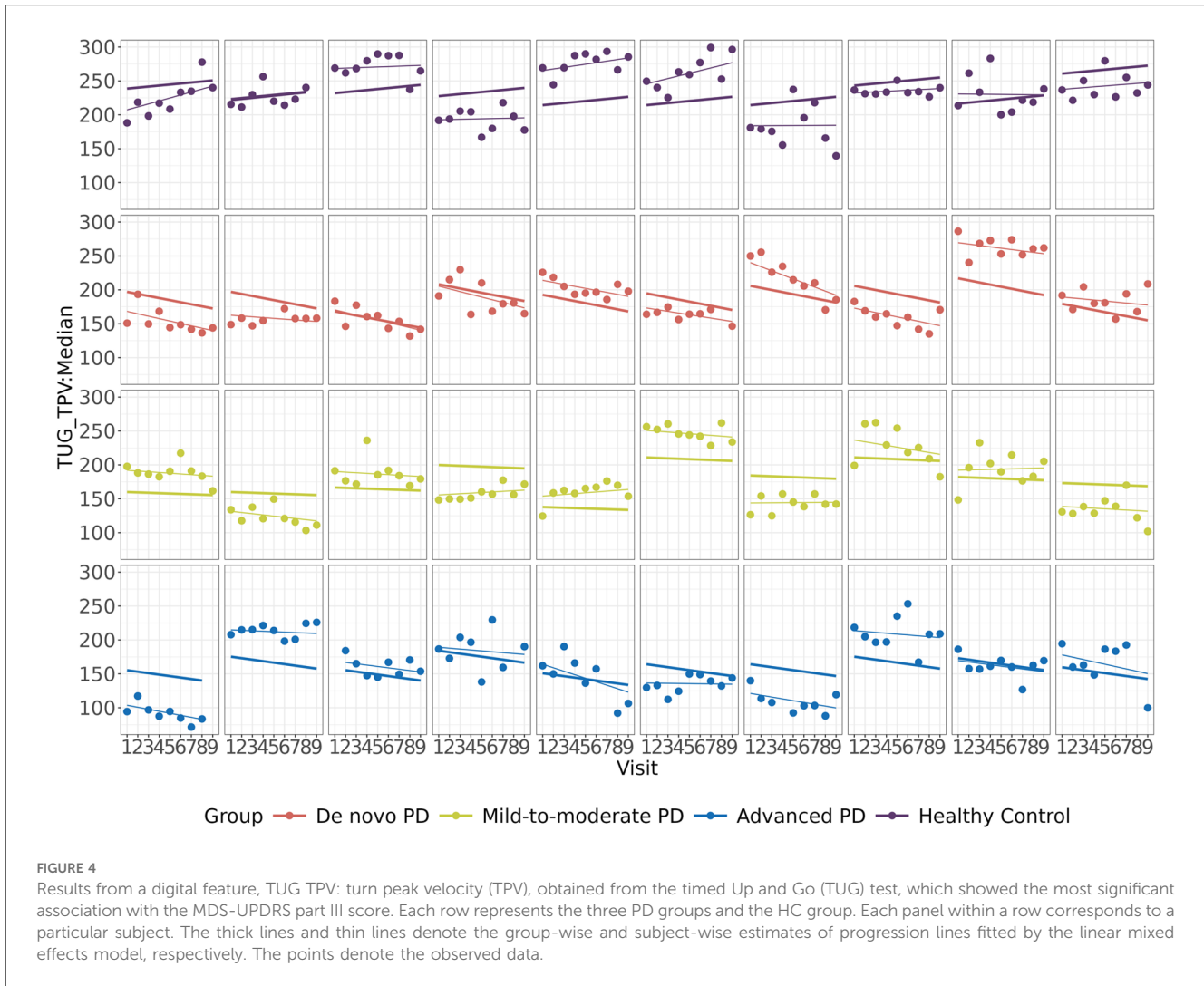
FIGURE 4

Results from a digital feature, TUG TPV: turn peak velocity (TPV), obtained from the timed Up and Go (TUG) test, which showed the most significant association with the MDS-UPDRS part III score. Each row represents the three PD groups and the HC group. Each panel within a row corresponds to a particular subject. The thick lines and thin lines denote the group-wise and subject-wise estimates of progression lines fitted by the linear mixed effects model, respectively. The points denote the observed data.

TABLE 2 Description of the selected features: 9 features selected for longitudinally disease progression tracking, and 3 features selected for *de novo* PD vs. HC classification.

Objective for feature selection	Feature	Statistic	Side	Description	PGEE Estimate
Disease progression tracking	TUG TD	Median		Turns—Duration	0.39
	TUG TPV	Median		Turns—Turn Velocity	−0.38
	TUG STSD	Mean		Stand to Sit—Duration	0.34
	Walk TA	Median		Turns—Angle	0.26
	Walk GLLC	MAD	A	Gait/Lower Limb—Cadence	0.16
	Walk GLLSW	MAD	L	Gait/Lower Limb—Swing	0.07
	Walk GLLLSM	Median	L	Gait/Lower Limb—Circumduction	−0.06
	Walk APAMAA	Mean		Anticipatory Postural Adjustment—Forward APA Peak	−0.06
	Sway PSAN95ESA	Mean		Postural Sway/Angles—Sway Area	0.05
de novo PD vs. HC classification	TUG TPV	Median		Turns—Turn Velocity	−0.60
	Walk GLLTOA	MAD	A	Gait/Lower Limb—Toe Out Angle	−0.43
	Walk GULMV	Median	A	Gait/Upper Limb—Arm Swing Velocity	−0.27

MAD, Mean Absolute Deviation; A, Affected side; L, Less affected side.

respectively. For advanced PD vs. HC, the effect sizes of progression slope were 0.76, 0.90, and 0.76 for MDS-UPDRS Part III, composite digital measure, and TUG TPV, respectively. On the other hand, the composite digital measure didn't show significant progression in the mild-to-moderate, on-therapy PD group, which was consistent with what is observed in Figure 5. Recall that none of the 11 selected digital features had significant univariate progression in the mild-to-moderate, on-therapy PD group (for example, the mild-to-moderate PD vs. HC slope p-value of TUG TPV was not significant, $p = 0.15$). Thus, it was not surprising that the
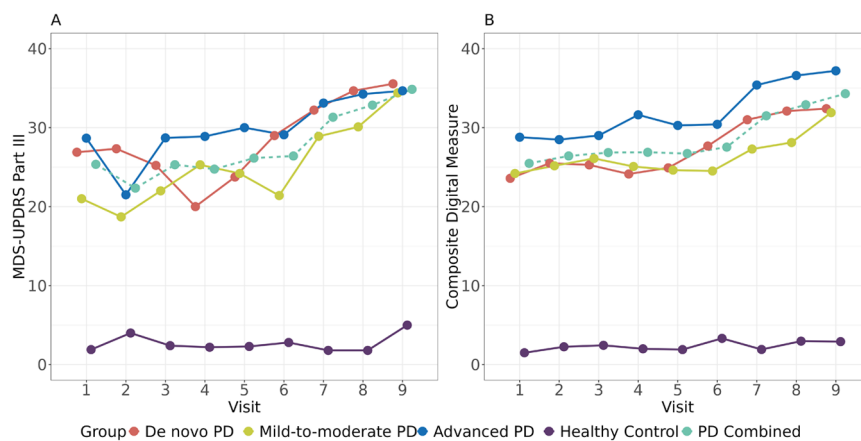
FIGURE 5
Tracking PD progression via (A) the MDS-UPDRS part III and (B) the composite digital measure based on the 11 selected digital features. The dashed lines represented the observed MDS-UPDRS Part III scores and the predicted composite digital measures in the combined PD group, respectively.

TABLE 3 Performance comparison among MDS-UPDRS part III, the composite digital measure, and TUG TPV in terms of both progression and variability.

	MDS-UPDRS Part III	Composite Digital Measure	TUG TPV
Group-by-visit p-value	6.22×10^{-3}	7.65×10^{-3}	8.05×10^{-3}
de novo PD vs. HC: slope p-value	0.02	8.28×10^{-3}	1.35×10^{-3}
Mild-to-moderate PD vs. HC: slope p-value	2.25×10^{-4}	0.16	0.15
Advanced PD vs. HC: slope p-value	0.07	4.22×10^{-3}	0.01
HC slope p-value	0.01	0.58	0.78
Between-subject coefficient of variation	39.0%	20.4%	17.2%
Within-subject coefficient of variation	34.1%	16.9%	9.7%
Effect size in progression slope between *de novo* PD and HC	1.41 (0.51, 2.31)	2.14 (1.24, 3.04)	1.37 (0.47, 2.27)
Effect size in progression slope between mild-to-moderate PD and HC	2.28 (1.40, 3.16)	0.86 (−0.02, 1.73)	0.60 (−0.28, 1.47)
Effect size in progression slope between advanced PD and HC	0.76 (−0.11, 1.64)	0.90 (0.02, 1.77)	0.76 (−0.12, 1.64)

composite digital measure preserved the same pattern. Moreover, the composite digital measure showed smaller between-/within-subject coefficient of variation than MDS-UPDRS Part III. In summary, the results from Figure 5 and Table 3 indicate that the composite digital measure is an attractive aggregated measure for tracking PD progression compared to MDS-UPDRS Part III and to individual digital features.

3.4.3 Performance in classifying *de novo* PD and HC

We further examined if the composite digital measure developed above (for tracking PD progression longitudinally) was also effective in classifying between *de novo* PD and HC subjects. Results are shown in Figure 6, where the boxplot of composite digital measures in the *de novo* PD subgroup is clearly higher than the boxplot in the HC subgroup. The composite digital measure had an AUC of 0.992 in such classification, which was very similar to that achieved for the classification model based on MDS-UPDRS Part III (AUC of 0.991). This demonstrated that the composite digital measure was able to preserve the ability to differentiate PD from HC groups and was effective in classifying *de novo* PD and HC.

4 Discussion

DHT-derived measures have shown great promise in both tracking disease progression and disease classification. However, it remains challenging to identify digital features for predicting disease progression longitudinally in a high dimensional space. Furthermore, methodologies for combining individual digital features into composite digital measures have not been fully explored and standardized in the field of DHTs. Although there have been several examples where composite digital measures were developed for tracking symptom progression, many of these prior efforts used simple sums of pre-defined, unweighted features without optimizing for performance (22, 29). In other cases where digital composite measures were trained/optimized to predict clinical scores, machine learning methodologies were often used without consideration of the longitudinal nature of the features (15, 30). The field has not adopted consistent and systematic methods and/or analysis frameworks that use statistical or machine learning methods capable of handling high-dimensional longitudinal data for feature selection and prediction while considering the within-subject correlation across visits. Therefore, there is an urgent need to develop this kind of new methodologies
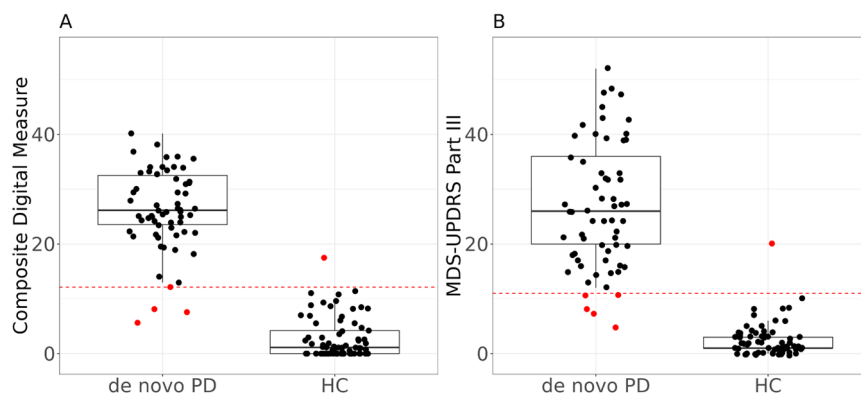
FIGURE 6
Boxplots of (A) the composite digital measure vs. (B) the MDS-UPDRS Part III in *de novo* PD and HC groups, respectively. Each dot indicates the composite digital measure or clinical score of each subject at each visit; the dotted line indicates the optimized threshold for classification: 12.14 for composite digital measure and 11 for MDS-UPDRS Part III. Dots above the line were classified into *de novo* PD, and dots below the line were classified into HC. Values that lead to false classifications are shown in red. The composite digital measure has AUC of 0.992 and MDS-UPDRS Part III has AUC of 0.991.

and analysis pipelines for the construction of composite digital measures for disease progression tracking, tailored for high-dimensional, longitudinal data with digital features collected from sensor technologies. In this paper, we propose a principled, scalable, and comprehensive methodology for the identification of relevant digital features of disease progression from large DHT data sets, and subsequent construction of a composite digital measure for disease progression tracking. Specifically, in Step 1, data is collected and processed for aggregated observation and quality control. In Step 2, we apply a linear mixed effects model for univariate screening for longitudinal progression of digital features. In Step 3, a univariate association test is conducted between candidate digital features (i.e., features that pass the univariate screening) and clinical scores, for example the MDS-UPDRS Part III and/or its subscales. In Step 4, the candidate digital features are ranked via a ML-based method, PGEE, for high-dimensional longitudinal data analysis. The optimal number of top features to be included into the composite digital measure is further determined using a cross-validation based algorithm to avoid overfitting. Note that PGEE (penalized GEE) method is designed for longitudinal data analysis with high-dimensional covariates by including a penalty term in the traditional GEE model. PGEE is particularly useful in handling high-dimensional feature sets, applicable for data from DHTs.

To demonstrate the utility of our methodology, we applied it to the data collected from a PD longitudinal observational study, which consisted of Opal^{TM} sensor-based movement measurements and MDS-UPDRS Part III scores collected from PD patients at a range of disease stages and healthy controls over a 2-year duration. Our primary interest in developing a composite digital measure is to track disease progression. The composite digital measure developed from this illustrative example generally showed a smoother and more significant increasing trend in PD groups and smaller between-/within-subject coefficients of variation than MDS-UPDRS Part III in

this small dataset ($N = 40$), indicating potential utility for the composite digital measures to be used to track disease progression more sensitively and with less variability vs. standard clinical measures. It should be noted that the dataset in our illustrative example was small ($N = 40$), and therefore, results of our analysis should be interpreted with caution. The analysis reported here was presented as an illustration of our proposed methodology and framework and was not intended as a proposed composite measure for use in future studies. We also note that the composite digital measure shows less significant progression trending in mild-to-moderate, on-therapy PD patients compared to in *de novo* and advanced PD patients. This outcome is consistent with the trends observed by Brzezicki et al. (11) using data derived from the OxQUIP study. We further evaluated the classification performance between *de novo* PD and HC using the composite digital measure built from our methodology (primarily for tracking PD progression). The measure had an AUC ROC of 0.992 for classification (vs. AUC = 0.991 when using MDS-UPDRS Part III), indicating that the composite digital measure also had a good performance in classifying between *de novo* PD and HC subjects, comparable to MDS-UPDRS Part III.

Note that in our analysis, the top digital features (i.e., those with the largest PGEE estimates from the multivariate penalized regression model) were selected for both tracking MDS-UPDRS Part III progression and classifying between *de novo* PD and HC. While the digital feature TUG TPV ranked high in both subsets of selected features, we observe that the digital features that are important for disease progression tracking are not necessarily the same as digital features important for patient identification. We constructed the composite digital biomarker with the merged feature list in this example.

Regarding performance, this composite digital biomarker keeps the main characteristics of individual digital features but exhibits a more significant increasing trend indicative of disease progression.

On one hand, this composite digital biomarker shows progression in *de novo* and advanced PD groups but remains flat in mild-to-moderate, on-therapy PD and HC groups which is consistent with the general behaviours of individual digital features (Figure 2). On the other hand, it exhibits a more significant increasing longitudinal trend compared to individual digital features (including TUG TPV). It is worth noting that the features selected using the PGEE model tend to have diverse measurement properties. The final feature set is not a combination of the best-performing individual features in terms of individual progression signal (i.e., neither features with the most progressions in PD groups nor features with the most significant univariate association with MDS-UPDRS Part III). A possible explanation is that combining top features with high correlations doesn't necessarily add additional information to the composite; there could be redundancy among digital features. It also suggests opportunities to further improve the performance of the composite digital measure by enriching the feature set with different assessments/tasks and measures.

The superior performance observed in the multivariate analysis, albeit from a small pilot dataset, suggests promises for use of composite digital measures for progression tracking in future studies. Recent modelling efforts have shown that an increased precision made possible by more objective and frequent composite digital measures could lead to smaller and shorter proof-of-concept studies to demonstrate disease-modifying treatment effect (20), which is critical in enabling and accelerating drug development. Open discussions on methodologies to identify the relevant digital features (from the multitude of digital measure possible with DHTs) and construct composite digital measures are critical to enable the adoption of such digital measures, and we present a methodology for this herein.

We see broad applicability of our proposed framework in handling high-dimensional, longitudinal DHT datasets and developing novel digital biomarkers for disease progression. To gain confidence in the use of such biomarkers for decision-making in clinical development, we anticipate that further efforts in technical validation and clinical validation will also be needed to build confidence in the constructed composite digital measures. Additionally, operational feasibility and user acceptance are critical to ensure that the measure can be successfully collected in clinical trials. All these elements will be part of the evidence package to support the fit-for-purpose use of a new digital biomarker and will be important to both clinical implementation and the interpretation of results.

Lastly, although we propose here a machine learning-based approach to develop composite digital biomarkers as indicators of traditional clinical endpoints, it is also valuable to further explore the clinical and biological relevance of the identified features. For example, one could examine the univariate associations between individual digital features and the clinical scores or domain sub-scores (as included in our pipeline and illustrated in Figure 3). Further, the relevance of many symptom features to the underlying disease mechanism have also been reported in the literature. In our illustrative example, several turning features, including peak velocity, duration, and angle, were identified to be

valuable for progression tracking; coincidentally, turning has been highlighted in many prior publications as a common challenge in PD (39–42) and is also included in a Phase 2 interventional study as a key digital feature (38).

There are several limitations of our work. First, a major caveat of the results reported from the illustrative example herein is that this analysis only used a small number of participants. Our proposed analysis workflow for digital biomarker development needs to be applied to additional studies with larger N to further demonstrate utility. The identified individual digital features of Parkinson's disease progression and the composite digital measure presented herein is solely for purposes of illustrating the methodology approach. They would need to be validated and verified in an independent dataset in further research before they can be used as digital biomarkers of disease progression and treatment response. Second, the digital features in our study were obtained from sensor-based movement measurements using one DHT system used during supervised, in-clinic tasks. Different or expanded digital features may be available with different DHTs, different task-based assessments, use of passive monitoring approaches, technology evolution, and further algorithm development. It is worth noting that we mainly use this feature set to demonstrate the methodology, and our proposed high-dimensional longitudinal data analysis framework (including feature selection and predictive modelling) is adaptive for different feature sets collected from different sensor technologies. Third, in our illustrative example, we examined potential confounders (i.e., age and sex) in the pooled PD group to identify factors that might significantly impact the progression trending. As a result, age was identified and included into our model as a covariate to be adjusted. It would have been preferred to assess potential confounders in each PD subgroup (i.e., *de novo* PD, mild-to-moderate PD, and advanced PD) rather than the pooled PD group; however, the small sample size and imbalanced datasets within the subgroups posed challenges to doing so.

Notably, the current dataset is longitudinal but only contains in-clinic visit data. One advantage of DHTs is that they may offer the ability to capture data outside of the clinic much more frequently. Other studies, including the Phase 2 Trial of Anti α-Synuclein Antibody in Early Parkinson's Disease (PASADENA) study (10) (daily tasks) and the Personalized Parkinson Project (PPP) study (43) (bi-weekly tasks), have shown utility in capturing remotely collected DHT data with increased measurement frequency. Increased measurement frequency could further enhance the performance of digital measures in quantifying disease progression, as it could address the day-to-day symptom fluctuations and reduce the measurement variability. Such remotely acquired digital features could also be applied to the methodology and framework we've reported here.

In addition, there is emerging research into characterization of the neurodegenerative disease progression directly from raw sensor signals recorded by DHTs (e.g., wearable sensors, environmental sensors, smartphone sensors) using deep neural networks and other black box algorithms (44, 45). Germane to these efforts is an important question about the interpretability of the ensuing

models and results (46, 47). In our work, we identified candidate digital features of disease progression using inherently interpretable linear models. We did not explore deep learning of the raw sensor data directly; such an approach is an interesting future direction of research.

In summary, with the rapid development of DHTs, digital measures are playing an increasingly important role in not only neurodegenerative disease detection, but also longitudinally tracking disease progression over time and detection of therapeutic response. Our proposed ML-based framework for identifying digital features of progression and constructing composite digital measures adds to the existing body of literature on digital measure analysis methodologies and may help accelerate the translation of digital measures to utility for drug development and clinical practice.

Data availability statement

The datasets presented in this article are not readily available because the original data presented in this paper is from the ongoing OxQUIP study and cannot be shared until completion of the whole study and full dissemination of results. This is expected to become possible within 24 months from the end of the study. Qualified researchers will be able to contact the Principal Investigator at the University of Oxford. Requests to access these datasets should be directed to Chrystalina A. Antoniades, chrystalina.antoniades@ndcn.ox.ac.uk.

Ethics statement

The studies involving humans were approved by a research ethics committee and the Health Research Authority (REC16/SW/0262). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SZ: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Formal Analysis, Methodology, Software, Visualization. AL: Conceptualization, Formal Analysis, Methodology, Writing – review & editing, Visualization. JS: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision. YX: Data curation, Formal Analysis, Writing – review & editing. VS: Conceptualization, Methodology, Writing – review & editing. JF: Funding acquisition, Investigation, Writing – review & editing. CA: Funding acquisition, Investigation, Writing – review & editing. DH: Methodology, Writing – review & editing. MD: Supervision, Writing – review & editing. JR: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Methodology, Supervision. RB: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Methodology, Supervision.

Funding

Conflict of interest

SZ, AL, JS, YX, VS, DH, MD, JR and RB were employed by Merck & Co., Inc. JS is a review editor for Statistical Genetics and Methodology in Frontiers in Genetics. JF is an Associate Editor for Neuroprosthetics in Frontiers in Neuroscience. CA is a review editor for Perception Science in Frontiers in Neuroscience and Frontiers in Psychology. JR is a review editor for Soft Matter Physics in Frontiers in Physics. RB is a review editor for Medical Physics and Imaging in Frontiers in Physiology and Frontiers in Physics.

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

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2024.1500811/full#supplementary-material

References

1. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement Disord*. (2008) 23(15):2129–70. doi: 10.1002/mds.22340

2. Committee for Medicinal Products for Human Use. Qualification Opinion on Dopamine Transporter Imaging as an Enrichment Biomarker for Parkinson's Disease Clinical Trials in Patients with Early Parkinsonian Symptoms. European Medicines Agency:EMA/CHMP SAWP/765041/2017 (2018).

3. Horak FB, Mancini M. Objective features of balance and gait for Parkinson's disease using body-worn sensors. *Mov Disord*. (2013) 28(11):1544–51. doi: 10.1002/mds.25684

4. Heldman DA, Espay AJ, LeWitt PA, Giuffrida JP. Clinician versus machine: reliability and responsiveness of motor endpoints in Parkinson's disease. *Parkinsonism Relat D*. (2014) 20(6):590–5. doi: 10.1016/j.parkreldis.2014.02.022

5. LeBaron V, Hayes J, Gordon K, Alam R, Homdee N, Martinez Y, et al. Leveraging smart health technology to empower patients and family caregivers in managing cancer pain: protocol for a feasibility study. *Jmir Res Protoc*. (2019) 8(12):e16178. doi: 10.2196/16178

6. Evers LJ, Raykov YP, Krijthe JH, de Lima ALS, Badawy R, Claes K, et al. Real-life gait performance as a digital biomarker for motor fluctuations: the Parkinson@home validation study. *J Med Internet Res*. (2020) 22(10):e19068. doi: 10.2196/19068

7. Erb MK, Karlin DR, Ho BK, Thomas KC, Parisi F, Vergara-Diaz GP, et al. Mhealth and wearable technology should replace motor diaries to track motor fluctuations in Parkinson's disease. *NPJ Digitl Med*. (2020) 3(1):6. doi: 10.1038/s41746-019-0214-x

8. Mahadevan N, Demanuele C, Zhang H, Volfson D, Ho B, Erb MK, et al. Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device. *NPJ Digit Med*. (2020) 3(1):5. doi: 10.1038/s41746-019-0217-7

9. Burq M, Rainaldi E, Ho KC, Chen C, Bloem BR, Evers LJW, et al. Virtual exam for Parkinson's disease enables frequent and reliable remote measurements of motor function. *NPJ Digit Med*. (2022) 5(1):65. doi: 10.1038/s41746-022-00607-8

10. Pagano G, Boess FG, Taylor KR, Ricci B, Mollenhauer B, Poewe W, et al. A phase II study to evaluate the safety and efficacy of prasinezumab in early Parkinson's disease (PASADENA): rationale, design, and baseline data. *Front Neurol*. (2021) 12:705407. doi: 10.3389/fneur.2021.705407

11. Brzezicki MA, Conway N, Sotirakis C, FitzGerald JJ, Antoniades CA. Antiparkinsonian medication masks motor signal progression in *de novo* patients. *Heliyon*. (2023) 9(11):e16415. doi: 10.1016/j.heliyon.2023.e16415

12. Pagano G, Taylor KI, Anzures-Cabrera J, Marchesi M, Simuni T, Marek K, et al. Trial of prasinezumab in early-stage Parkinson's disease. *New Engl J Med*. (2022) 387 (5):421–32. doi: 10.1056/NEJMoa2202867

13. Robin J, Xu M, Detke M, Simpson W. Validation of an objective, SpeechBased object content score for measuring disease progression in AD. *J Prev Alzheimer's Dis*. (2022) 9:S190. doi: 10.14283/jpad.2022.97 (P143).

14. Liu Y, Zhang G, Tarolli CG, Hristov R, Jensen-Roberts S, Waddell EM, et al. Monitoring gait at home with radio waves in Parkinson's disease: a marker of severity, progression, and medication response. *Sci Transl Med*. (2022) 14(663): eadc9669. doi: 10.1126/scitranslmed.adc9669

15. Yang Y, Yuan Y, Zhang G, Wang H, Chen YC, Liu Y, et al. Artificial intelligence-enabled detection and assessment of Parkinson's disease using nocturnal breathing signals. *Nat Med*. (2022) 28(10):2207–15. doi: 10.1038/s41591-022-01932-x

16. Kadirvelu B, Gavriel C, Nageshwaran S, Chan JPK, Nethisinghe S, Athanasopoulos S, et al. A wearable motion capture suit and machine learning predict disease progression in Friedreich's ataxia. *Nat Med*. (2023) 29:86–94. doi: 10.1038/s41591-022-02159-6

17. Ricotti V, Kadirvelu B, Selby V, Festenstein R, Mercuri E, Voit T, et al. Wearable full-body motion tracking of activities of daily living predicts disease trajectory in Duchenne muscular dystrophy. *Nat Med*. (2023) 29(1):95–103. doi: 10.1038/s41591-022-02045-1

18. Johnson SA, Karas M, Burke KM, Straczkiewicz M, Scheier ZA, Clark AP, et al. Wearable device and smartphone data quantify ALS progression and may provide novel outcome measures. *NPJ Digit Med*. (2023) 6(1):34. doi: 10.1038/s41746-023-00778-y

19. Giboin LS, Simillion C, Rennig J, Bamdadian A, Kinsella F, McColgan P, et al. A digital motor score for sensitive detection of disease progression in early manifest Huntington's disease (2023). Available online at: https://medically.roche.com/global/en/medical-material.a86d594d-0db2-42d9-9bef-6622e0b574d2.qr.html?cid=slprxx2304nehdchdi2023 (Accessed December 19, 2024).

20. Mori H, Wiklund SJ, Zhang JY. Quantifying the benefits of digital biomarkers and technology-based study endpoints in clinical trials: project moneyball. *Digit Biomark*. (2022) 6(2):36–46. doi: 10.1159/000525255

21. Adams JL, Kangarloo T, Gong Y, Khachadourian V, Tracey B, Volfson D, et al. Using a smartwatch and smartphone to assess early Parkinson's disease in the WATCH-PD study over 12 months. *NPJ Park.'s Dis*. (2024) 10:112. doi: 10.1038/s41531-024-00721-2

22. Czech MD, Badley D, Yang L, Shen J, Crouthamel M, Kangarloo T, et al. Improved measurement of disease progression in people living with early Parkinson's disease using digital health technologies. *Commun. Med*. (2024) 4(1):49. doi: 10.1038/s43856-024-00481-3

23. Perumal SV, Sankar R. Gait and tremor assessment for patients with Parkinson's disease using wearable sensors. *Ict Express*. (2016) 2(4):168–74. doi: 10.1016/j.icte.2016.10.005

24. Dinov ID, Heavner B, Tang M, Glusman G, Chard K, Darcy M, et al. Predictive big data analytics: a study of Parkinson's disease using large, complex, heterogeneous, incongruent, multi-source and incomplete observations. *PLoS One*. (2016) 11(8): e0157077. doi: 10.1371/journal.pone.0157077

25. Alaskar H, Hussain A. Prediction of Parkinson disease using gait signals. *2018 11th International Conference on Developments in ESystems Engineering (DeSE)* (2018). p. 23–6

26. Gao C, Sun H, Wang T, Tang M, Bohnen NI, Müller ML, et al. Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease. *Sci Rep*. (2018) 8(1):7129. doi: 10.1038/s41598-018-24783-4

27. Tsoulos IG, Mitsi G, Stavrakoudis A, Papapetropoulos S. Application of machine learning in a Parkinson's disease digital biomarker dataset using neural network construction (NNC) methodology discriminates patient motor Status. *Frontiers Ict*. (2019) 6:10. doi: 10.3389/fict.2019.00010

28. Dadu A, Satone V, Kaur R, Hashemi SH, Leonard H, Iwaki H, et al. Identification and prediction of Parkinson's disease subtypes and progression using machine learning in two cohorts. *NPJ Park.'s Dis*. (2022) 8(1):172. doi: 10.1038/s41531-022-00439-z

29. Taylor K, Lipsmeier F, Volkova-Volkmar E, Rukina D, Anzures-Cabrera J, Essioux L, et al. Prasinezumab reduced progression of Parkinson's disease motor features measured by roche PD Mobile application v2 sensor features: PASADENA phase II part 1 [abstract]. *Mov Disord*. (2021) 36(suppl 1). https://www.mdsabstracts.org/abstract/prasinezumab-reduced-progression-of-parkinsons-disease-motor-features-measured-by-roche-pd-mobile-application-v2-sensor-features-pasadena-phase-ii-part-1/

30. Sotirakis C, Su Z, Brzezicki MA, Conway N, Tarassenko L, FitzGerald JJ, et al. Identification of motor progression in Parkinson's disease using wearable sensors and machine learning. *NPJ Park.'s Dis*. (2023) 9(1):142. doi: 10.1038/s41531-023-00581-2

31. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. (1986) 73(1):13–22. doi: 10.1093/biomet/73.1.13

32. Wang L, Zhou J, Qu A. Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics*. (2012) 68(2):353–60. doi: 10.1111/j.1541-0420.2011.01678.x

33. Pereira MF, Buchanan T, Höglinger GU, Bogdanovic M, Tofaris G, Prangnell S, et al. Longitudinal changes of early motor and cognitive symptoms in progressive supranuclear palsy: the OxQUIP study. *BMJ Neurology Open*. (2022) 4:e000214. doi: 10.1136/bmjno-2021-000214

34. Svetnik V, Liaw A, Tong C, Wang T. Application of breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. In: Roli F, Kittler J, Windeatt T, editors. *Multiple Classifier Systems. MCS 2004. Lecture Notes in Computer Science, vol 3077*. Berlin, Heidelberg: Springer (2004). doi: 10.1007/978-3-540-25966-4_33

35. Zampieri C, Salarian A, Carlson-Kuhta P, Aminian K, Nutt JG, Horak FB. The instrumented timed up and go test: potential outcome measure for disease modifying therapies in Parkinson's disease. *J Neurology Neurosurg Psychiatry*. (2010) 81(2):171. doi: 10.1136/jnnp.2009.173740

36. Koop MM, Ozinga SJ, Rosenfeldt AB, Alberts JL. Quantifying turning behavior and gait in Parkinson's disease using mobile technology. *Ibro Reports*. (2018) 5:10–6. doi: 10.1016/j.ibror.2018.06.002

37. Lowry K, Woods T, Malone A, Krajek A, Smiley A, Van Swearingen J. The figure-of-8 walk test used to detect the loss of motor skill in walking among persons with Parkinson's disease. *Physiother Theory Pract*. (2022) 38(4):552–60. doi: 10.1080/09593985.2020.1774948

38. Lipsmeier F, Taylor KI, Postuma RB, Volkova-Volkmar E, Kilchenmann T, Mollenhauer B, et al. Reliability and validity of the Roche PD Mobile application for remote monitoring of early Parkinson's disease. *Sci Rep-uk*. (2022) 12(1):12081. doi: 10.1038/s41598-022-15874-4

39. Stack E, Ashburn A. Dysfunctional turning in Parkinson's disease. *Disabil Rehabil*. (2008) 30(16):1222–9. doi: 10.1080/09638280701829938

40. Hulbert S, Ashburn A, Robert L, Verheyden G. Narrative review of turning deficits in people with Parkinson's disease. *Disabil Rehabil*. (2014) 37(15):1382–9. doi: 10.3109/09638288.2014.961661

41. Mak MK, Patla A, Hui-Chan C. Sudden turn during walking is impaired in people with Parkinson's disease. *Exp Brain Res*. (2008) 190(1):43–51. doi: 10.1007/s00221-008-1446-1

42. Ragothaman A, Miranda-Dominguez O, Brumbach BH, Giritharan A, Fair DA, Nutt JG, et al. Relationship between brain volumes and objective balance and gait measures in Parkinson's disease. *J Park.'s Dis.* (2021) 12(1):283–94. doi: 10.3233/JPD-202403

43. Bloem BR, Marks WJ, de Lima ALS, Kuijf ML, van Laar T, Jacobs BPF, et al. The personalized Parkinson project: examining disease progression through broad biomarkers in early Parkinson's disease. *BMC Neurol.* (2019) 19(1):160. doi: 10.1186/s12883-019-1394-3

44. Atri R, Urban K, Marebwa B, Simuni T, Tanner C, Siderowf A, et al. Deep learning for daily monitoring of Parkinson's disease outside the clinic using wearable sensors. *Sensors.* (2022) 22(18):6831. doi: 10.3390/s22186831

45. Khan P, Fazlul Kader MD, Riazul Islam SM, Rahman AB, Sshahriar Kamal MD, Uddin Toha M, et al. Machine learning and deep learning approaches for brain disease diagnosis: principles and recent advances. *IEEE Access.* (2021) 9:37622–55. doi: 10.1109/ACCESS.2021.3062484

46. Rudin C, Chen C, Chen Z, Huang H, Semenova L, Zhong C. Interpretable machine learning: fundamental principles and 10 grand challenges. *Stat Surv.* (2022) 16:1–85. doi: 10.1214/21-SS133

47. US Food and Drug Administration. Good machine learning practice for medical device development: guiding principles (2021). Available online at: https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles (Accessed December 19, 2024).