



Motor Progression in Early-Stage Parkinson's Disease: A Clinical Prediction Model and the Role of Cerebrospinal Fluid Biomarkers

Ling-Yan Ma^{1,2†}, Yu Tian^{3†}, Chang-Rong Pan³, Zhong-Lue Chen⁴, Yun Ling⁴, Kang Ren⁴, Jing-Song Li^{3*} and Tao Feng^{1,2,5*}

¹ Department of Neurology, Center for Movement Disorders, Beijing Tiantan Hospital, Capital Medical University, Beijing, China, ² China National Clinical Research Center for Neurological Diseases, Beijing, China, ³ Engineering Research Center of Electronic Medical Record (EMR) and Intelligent Expert System, College of Biomedical Engineering and Instrument Science, Zhejiang University, Ministry of Education, Hangzhou, China, ⁴ Gyenno Science Co. Ltd., Shenzhen, China, ⁵ Parkinson's Disease Center, Beijing Institute for Brain Disorders, Beijing, China

OPEN ACCESS

Edited by:

Dennis Qing Wang,
Southern Medical University, China

Reviewed by:

Angelo Antonini,
University of Padua, Italy
Xie Fen,
Southern Medical University, China

*Correspondence:

Tao Feng
bxbkjys@sina.com
Jing-Song Li
ljs@zju.edu.cn

[†]These authors have contributed
equally to this work

Received: 08 November 2020

Accepted: 28 December 2020

Published: 25 January 2021

Citation:

Ma L-Y, Tian Y, Pan C-R, Chen Z-L,
Ling Y, Ren K, Li J-S and Feng T
(2021) Motor Progression in
Early-Stage Parkinson's Disease: A
Clinical Prediction Model and the Role
of Cerebrospinal Fluid Biomarkers.
Front. Aging Neurosci. 12:627199.
doi: 10.3389/fnagi.2020.627199

Background: The substantial heterogeneity of clinical symptoms and lack of reliable progression markers in Parkinson's disease (PD) present a major challenge in predicting accurate progression and prognoses. Increasing evidence indicates that each component of the neurovascular unit (NVU) and blood-brain barrier (BBB) disruption may take part in many neurodegenerative diseases. Since some portions of CSF are eliminated along the neurovascular unit and across the BBB, disturbing the pathways may result in changes of these substances.

Methods: Four hundred seventy-four participants from the Parkinson's Progression Markers Initiative (PPMI) study (NCT01141023) were included in the study. Thirty-six initial features, including general information, brief clinical characteristics and the current year's classical scale scores, were used to build five regression models to predict PD motor progression represented by the coming year's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Part III score after redundancy removal and recursive feature elimination (RFE)-based feature selection. Then, a threshold range was added to the predicted value for more convenient model application. Finally, we evaluated the CSF and blood biomarkers' influence on the disease progression model.

Results: Eight hundred forty-nine cases were included in the study. The adjusted R^2 values of three different categories of regression model, linear, Bayesian and ensemble, all reached 0.75. Models of the same category shared similar feature combinations. The common features selected among the categories were the MDS-UPDRS Part III score, Montreal Cognitive Assessment (MOCA) and Rapid Eye Movement Sleep Behavior Disorder Questionnaire (RBDSQ) score. It can be seen more intuitively that the model can achieve certain prediction effect through threshold range. Biomarkers had no significant impact on the progression model within the data in the study.

Conclusions: By using machine learning and routinely gathered assessments from the current year, we developed multiple dynamic models to predict the following year's

motor progression in the early stage of PD. These methods will allow clinicians to tailor medical management to the individual and identify at-risk patients for future clinical trials examining disease-modifying therapies.

Keywords: Parkinson's disease, motor progression, predictive model, Parkinson's progression markers initiative, machine learning

INTRODUCTION

Parkinson's disease (PD) is a chronic progressive neurodegenerative disorder characterized by a broad spectrum of gradual motor and non-motor impairments (Selikhova et al., 2009). In the clinical course of PD, both linear (Gottipati et al., 2017; Holden et al., 2018) and non-linear (Vu et al., 2012; Reinoso et al., 2015) progression have been reported in the advancement of motor and non-motor symptoms. In an era of increasing focus on individualized management and disease-modifying therapies, there is a need to develop tools to predict motor progression at the individual level. However, the substantial heterogeneity (Foltynie et al., 2002; Selikhova et al., 2009; Ma et al., 2015; Qian and Huang, 2019) of clinical symptoms and lack of reliable progression markers present a major challenge in predicting accurate progression and prognoses.

The current literature on PD progression consists largely of associative analyses focusing on predictors such as gender, age, clinical subtype (Aleksovski et al., 2018), genes (Deng et al., 2019), cognitive status and baseline motor score (Reinoso et al., 2015). Greater progression of motor scores has been associated with several factors, such as male gender, older age at diagnosis, akinetic-rigid subtype, cognitive impairment (Reinoso et al., 2015), right-side onset (Baumann et al., 2014), orthostatic hypotension, and rapid eye movement sleep behavior disorder (Fereshtehnejad et al., 2015). Apart from associative analyses, a few prognostic models have been developed that focus on predicting different aspects of PD at the individual level, including logistic regression and Bayesian classification models to predict cognitive impairment (Schrag et al., 2017; Hogue et al., 2018; Gramotnev et al., 2019), machine-learning, random survival forests to predict time to initiation of symptomatic treatment (Simuni et al., 2016) and Bayesian machine-learning methods to predict motor progression (Latourelle et al., 2017) and data mining and classification techniques to predicting faster symptoms worsening at baseline patients evaluation (Tsiouris et al., 2017). Based on the Parkinson's Progression Markers Initiative (PPMI) database, Latourelle et al. developed comprehensive multivariable prognostic models to predict the annual rate of change in PD (Latourelle et al., 2017). The authors included relatively comprehensive indicators, including baseline molecular and clinical variables, to construct an ensemble of models to better clarify and analyze the predictors. However, due to the large heterogeneity of motor symptoms and the complexity of related factors, it was difficult to predict motor progression with high accuracy. The models yielded a cross-validated R^2 value of 41% for the PPMI cohort and 9% for the LABS-PD cohort.

Increasing evidence indicates that each component of the neurovascular unit and blood-brain barrier (BBB) disruption

may take part in many neurodegenerative diseases (Yamazaki and Kanekiyo, 2017). Since some portions of Cerebrospinal Fluid (CSF) are eliminated along the neurovascular unit and across the BBB, disturbing the pathways may result in changes of these substances. CSF biomarkers in PD, such as α -synuclein species, lysosomal enzymes, markers of amyloid and tau pathology, and neurofilament light chain, have been suggested to possess potential diagnostic and prognostic value of PD (Parnetti et al., 2019). In large areas of undeveloped countries, costly tests, such as genetic and CSF testing, and image detection for PD (PET and MIBG), could exert a great burden on patients and medical security systems. Furthermore, the corresponding variables are not data routinely used for common clinical activities and are difficult to obtain. Considering clinical needs and utility, by using machine-learning methods with PD patient data from the PPMI database, we aim to develop multiple dynamic models to predict motor progression based on general information and classical clinical scales, displayed in the form of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Part III score. We also explore the influence of CSF and blood biomarkers on the prediction model. The use of baseline assessments (e.g., age, gender, disease duration, motor and non-motor examination, validated self-report questionnaires), which are either already routinely performed or could be reasonably implemented during a typical neurologist's office visit, can facilitate widespread implementation of this cost-efficient predictive model in real world applications.

METHODS

Participants

Data used in the preparation of the article were obtained from the PPMI database. The PPMI is an international, multicenter, prospective study designed to discover, and validate biomarkers of disease progression in newly diagnosed PD participants (National Clinical Trials identifier NCT01141023). Each PPMI recruitment site received approval from an institutional review board or ethics committee on human experimentation before study initiation. Written informed consent for research was obtained from all individuals participating in the study. The PPMI database was accessed on December 9, 2020, to obtain data from baseline visits ($n = 474$) and follow-up visits (the numbers of follow-up visits for five separate years are 380, 323, 257, 259, 230). For up-to-date information on the study, please visit www.ppmi-info.org.

Model Variables

The main model outcomes of interest were motor function in the coming year. The primary outcome was the MDS-UPDRS Part

III score. Predictive variables included demographics, disease duration, and measures of motor (off state) and non-motor function. Motor assessments included the MDS-UPDRS Parts II-IV, Hoehn and Yahr Stage (H&Y), PD subtype evaluations (tremor dominant (TD), postural instability and gait difficulty (PIGD) and TD/PIGD scores). For non-motor evaluation, the Epworth Sleepiness Scale Score (ESS) and REM Sleep Behavior Disorder Screening Questionnaire Score (RBDSQ) were applied to measure a patient's sleep quality and disturbances. The Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease (QUIP) score was used to evaluate PD impulse control disorders. The Geriatric Depression Scale (GDS) score was designed specifically to screen for depression in geriatric patients. The total score of the assessment of autonomic dysfunction in Parkinson's disease (SCOPA-AUT) was used to evaluate autonomic nervous function in the prior month. The MDS-UPDRS Part I score and component scores were used to evaluate mood and mentation.

In addition, we assessed whether biomarkers including CSF amyloid, CSF α -synuclein (and its ratio to CSF amyloid), serum uric acid had an impact on the progression model in the study.

Data Preprocessing

Each patient had a maximum of five samplings performed at different intervals. During data processing, we excluded negative differences in the MDS-UPDRS Part III score between the coming year and the prior year. According to the PPMI protocol, the "off state" was evaluated as more than 6 h after the last dose of dopaminergic therapy. Considering that the disease was always in a progressive state for all patients, a decrease in the MDS-UPDRS Part III score with increasing disease duration was thought to be unrealistic; therefore, patient samples that displayed this trend were discarded. Furthermore, as eliminating redundant information can improve the quality of the prediction model, we performed Spearman rank correlation analysis to ensure that no variables were correlated with a correlation coefficient of more than 0.5. Prior to model construction, all variables were transformed to a z-score by subtracting the mean value from each of the observed values and dividing by the standard deviation because the clinical phenotypes were measured on different scales and with different score ranges.

Regression Model Construction

First, the preprocessed data were divided into training and test sets. There were two dividing principles: (1) After ordering by diagnosis date, the first 75% of patients were included in the training data set; the remaining patients were included in the test set. (2) To avoid data overlapping, samples from the same patient were placed into either the training or test set.

We used five regression algorithms, including linear, ridge, Bayesian, random forest, and gradient boosting decision tree regression, to build the model with the training set. Features were selected using the recursive feature elimination (RFE) method. By evaluating model performance, the most irrelevant variables were eliminated in each iteration of the regression, and then the contributions of the remaining variables to the model were ranked. To select the best combination of features, RFE was

performed with cross-validation to calculate the verification error of all subsets of features. The subset with the smallest error was regarded as the final feature combination. Next, regression models were built with parameter adjustment to achieve the best performance. Finally, the models were validated with the test set.

Assessment of Biomarkers' Influence on Parkinson's Disease Progression Model

Due to the great difference between the absence of biomarkers and other clinical variables, the initial inclusion would result in smaller amounts of data. So, biomarkers were added after model construction based on the clinical and scale scores data to evaluate their influence on the basic model. First, correlation analysis was performed between biomarkers and disease progression. Then the variables with significant correlation were selected to improve the models and we compared the model difference between before and after adding the biomarkers variables.

Statistical Analysis and Performance Measures

Descriptive statistics of the patients' demographic and clinical characteristics are summarized in **Table 1**. The Kolmogorov-Smirnov test was used to test the normality of continuous data. Continuous variables are described by the mean, standard deviation, maximum, and minimum. Categorical variables are expressed as percentages. The *t*-test was used to compare regression models before and after adding biomarkers variables. Two-tailed $p < 0.05$ were considered to indicate statistical significance.

The performance of the models was assessed by the root mean square error (RMSE) and the adjusted coefficient of determination (adjusted R-squared coefficient or adjusted R^2). R^2 is a statistical measure in the regression model, equal to the ratio of the regression sum of squares to the total sum of squares and which reflects the degree of agreement between the data and the model. The influence of the number of variables on the goodness is excluded in the adjusted R^2 . Because doctors may score the same patient with minor differences in terms of the clinical significance of the symptom, we added a threshold range to each predicted value, which acted as the midpoint of the range, observed whether the true value fell within the range and calculated the accuracy for more convenient application of the model.

RESULTS

Patient Demographic and Clinical Characteristics

In this study, 474 (312 males and 162 females; M:F ratio 1.926:1; mean age 61.473 ± 9.753 years) PD patients were enrolled. The mean age of disease onset was 59.460 years, and the mean disease duration was 6.710 months at baseline. None of the continuous data were normally distributed according to the Kolmogorov-Smirnov test. Descriptive statistics of the

TABLE 1 | Descriptive statistics of the demographic and clinical variables of the Parkinson's progression markers initiative study participants.

Variables	Mean (SD; Min; Max)	Percentage (%)
Age	61.473 (9.753; 33.498; 84.884)	
Gender		Male: 65.823 Female: 34.177
Age at symptom onset	59.460 (10.054; 25.370; 83.008)	
Side most affected at PD onset		Left: 39.873 Right: 56.962 Symmetric: 3.165
Family History of PD		First degree family w/PD: 14.768 Non-1st degree family w/PD: 11.392 No family w/PD: 73.840
Duration (Months)	6.710 (6.664; 0.400; 37.000)	
Total rigidity score	3.508 (2.703; 0.000; 13.000)	
Hoehn & Yahr stage		Stage 1:45.992 Stage 2:53.376 Stage 3: 0.633
TD/PIGD classification (OFF)		TD: 71.730 PIGD: 17.300 Indeterminate: 10.970
Tremor score (OFF)	4.340 (3.105; 0.000; 18.000)	
MDS-UPDRS part III score (OFF)	19.937 (9.141; 2.000; 51.000)	
ESS score	6.162 (3.758; 0.000; 20.000)	
RBDSQ score	4.162 (2.686; 0.000; 12.000)	
MOCA score	27.133 (2.325; 17.000; 30.000)	
QUIP score	0.325 (0.697; 0.000; 4.000)	
GDS score	2.464 (2.672; 0.000; 14.000)	
SCOPA-AUT total score	10.103 (6.760; 0.000; 44.000)	
MDS-UPDRS part I score	5.956 (4.571; 0.000; 27.000)	
Initial symptom (at diagnosis)—resting tremor		Symptom present at diagnosis: 78.692
Initial symptom (at diagnosis)—rigidity		Symptom present at diagnosis: 73.840
Initial symptom (at diagnosis)—bradykinesia		Symptom present at diagnosis: 82.068
Initial symptom (at diagnosis)—postural instability		Symptom present at diagnosis: 7.595

SD, Standard deviation; Min, Minimum; Max, Maximum; TD, Tremor Dominant; PIGD, Postural Instability/Gait Difficulty; MDS-UPDRS, Movement Disorder Society-Unified Parkinson's Disease Rating Scale; ESS, Epworth Sleepiness Scale; RBDSQ, REM Sleep Behavior Disorder Screening Questionnaire; MOCA, Montreal Cognitive Assessment; QUIP, Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease; GDS, Geriatric Depression Scale; SCOPA-AUT, Assessment of autonomic dysfunction in Parkinson's disease.

general and total score variables are shown in **Table 1** and other detailed score variables are shown in **Supplementary Table 1**. A heatmap of the Spearman correlation coefficients between the variables are depicted in **Figure 1**. To reduce redundant information, we eliminated variables with strong correlations. So, most of the individual items were removed and the total scores were retained. In total, 24 features remained and are shown in **Table 2**. Apart from these, we also calculated the correlations between the 24 individual variables and the outcome event. The variables with the top five highest correlations were MDS-UPDRS Part III Score (0.882), Duration of Disease since Diagnosis (0.269), MDS-UPDRS Part I Apathy (0.197), SCOPA-AUT Total Score (0.196), and MDS-UPDRS Part I Fatigue (0.154). The correlation coefficients and probability values of all 24 variables are shown in **Supplementary Table 2**.

Prediction Model Construction and Performance Measures

After eliminating the missing data, a total of 849 cases were left, among which 633 cases were placed in the training set and 216 in the test set according to the aforementioned dividing principles. The disease progression as expressed by the MDS-UPDRS Part III score of the coming year was predicted based on the information from the previous year. We built regression models based on five algorithms, and their performances are shown in **Table 3**. The models can be divided into three categories: linear, Bayesian and ensemble methods. The adjusted R^2 of all the models reached above 0.75. The feature combinations obtained by the RFE method for models in the same category were similar and the feature importance or coefficient of each model is shown in **Table 4**. The models from each of the categories jointly selected the same three features: MDS-UPDRS Part III score, MOCA

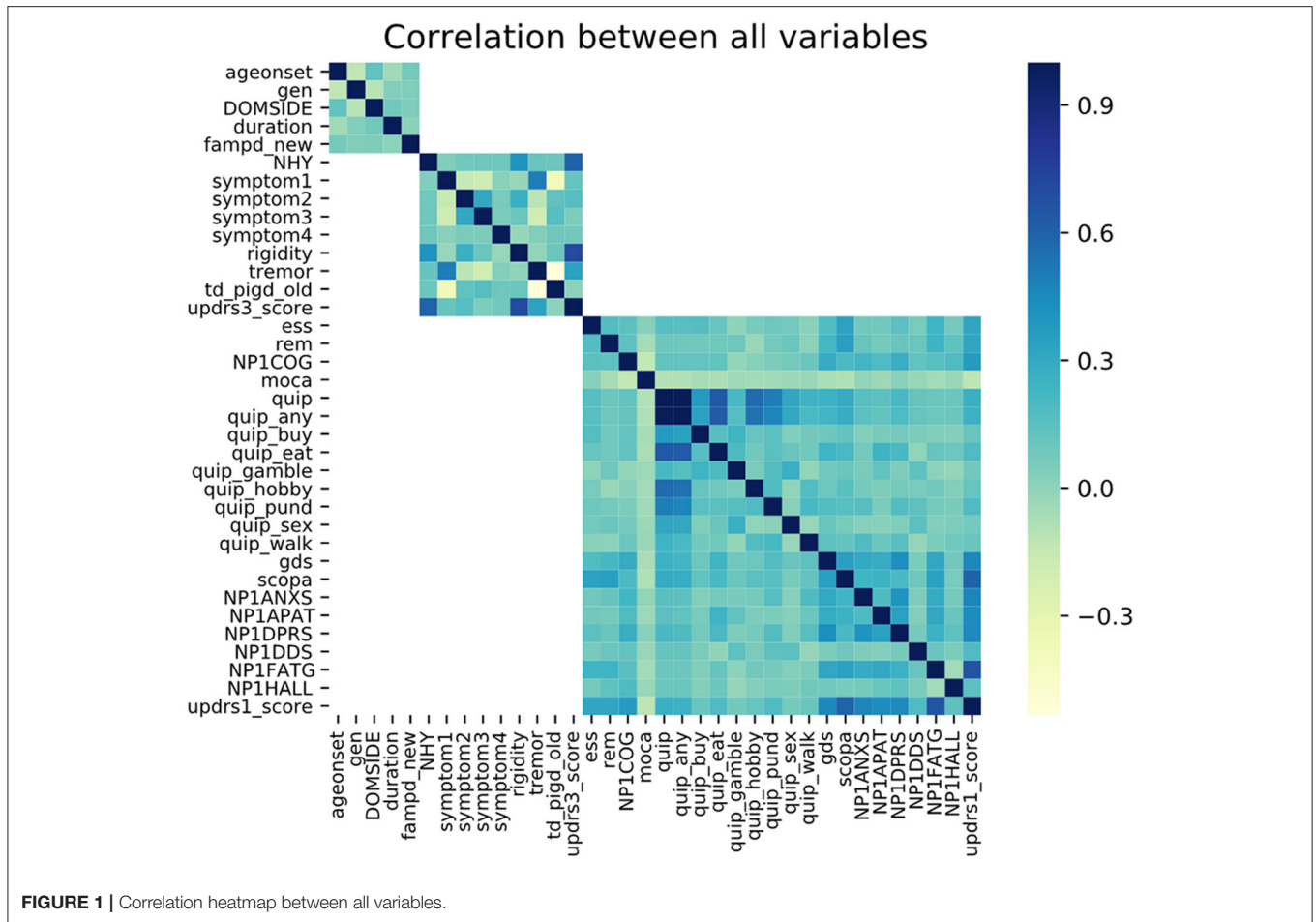


TABLE 2 | The remaining 24 features after eliminating features with strong correlation.

Category	Features
General	Age at symptom onset gender side most affected at PD onset Duration (Months) family history of PD
Motor	Four initial symptoms at diagnosis TD/PIGD classification MDS-UPDRS Part III Score
Non-motor	ESS score RBDSQ score QUIP score GDS score SCOPA-AUT total score MOCA score MDS-UPDRS part I analytic score

Score and RBDSQ score. All coefficients of regression models are shown in **Table 5**.

Taking as an example, the random forest (RF) regression had an adjusted R^2 of 0.770 and an RMSE of 5.678. The importance of each feature in the model is computed as the total reduction in

TABLE 3 | The performance of five regression models.

Model category	Model	Adjusted R^2	RMSE
Linear methods	Linear regression	0.779	5.585
	Ridge regression	0.779	5.584
Bayesian method	Bayesian regression	0.778	5.579
Ensemble methods	Random forest	0.770	5.678
	Gradient boosting	0.753	5.918

the criterion introduced by that feature and is shown in **Table 4**. **Figure 2** presents the evaluation of the predictions on the test set, with the predicted value as the midpoint of threshold ranges with half-range widths of 5, 6, and 7 for adjustment. Finally, we calculated the proportion of the true value that fell into the range as the accuracy.

Biomarkers Influence on Parkinson's Disease Progression Model

At the end of the study, the influence of biomarkers on Parkinson's disease progression model was analyzed. First, we performed spearman correlation analysis between biomarkers

TABLE 4 | The feature importance or coefficient of features selected through RFE method in five regression models.

Features selected through RFE method	Feature importance or coefficient				
	Linear regression	Ridge regression	Bayesian regression	Random forest	Gradient boosting
MDS-UPDRS part III score (OFF)	0.970	0.970	0.976	0.929	0.958
Age at symptom onset	/	/	/	0.009	0.002
Duration (months)	/	/	/	0.028	0.030
Family history of PD	0.523	0.521	0.326	/	/
MDS-UPDRS part I anxious mood	0.484	0.479	0.209	/	/
RBDSQ score	0.151	0.151	0.103	0.010	0.007
MDS-UPDRS part I features of dopamine dysregulation syndrome	-1.780	-1.738	-1.302	/	/
Initial symptom (at diagnosis)—postural instability	/	/	0.427	/	/
ESS score	/	/	/	0.010	/
SCOPA-AUT total score	/	/	/	0.007	/
MOCA score	-0.031	-0.031	-0.001	0.007	0.003

TABLE 5 | All coefficients of regression models.

Model	Coefficients
Linear regression	/
Ridge regression	Regularization strength = 1.32
Bayesian regression	Maximum number of iterations = 300
Random forest	The number of trees in the forest = 260 The maximum depth of the tree = 4 The function to measure the quality of a split = “the mean absolute error”
Gradient boosting	The number of boosting stages to perform = 30 Learning rate = 0.1 The function to measure the quality of a split = “the mean squared error with improvement score by Friedman”

and disease progression and found that the CSF amyloid and CSF α -synuclein were significantly associated with disease progression, of which the correlation coefficient were, respectively, -0.132 , -0.160 . The correlation analysis results of other variables are shown in **Supplementary Table 1**. After incorporating the two variables into feature combinations selected by feature selection based on clinical characters and scales scores and eliminating the missing data, a total of 441 cases were left. The models were reconstructed based on the feature combinations before and after the inclusion of biomarkers variables and the performance of two sets of models were compared by *t*-test. The *P*-value were 0.408 and 0.883 respectively, showing no significant difference and indicating that these biomarkers had no significant impact on the progression model within the data in the study. The

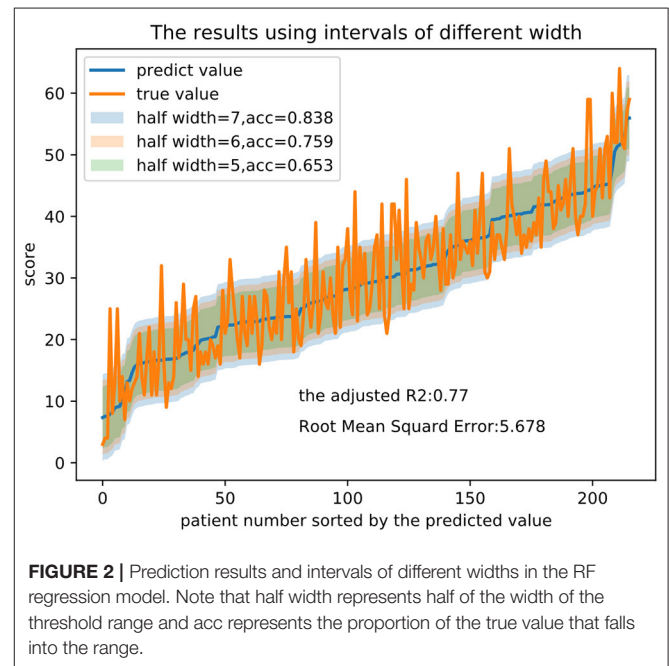


FIGURE 2 | Prediction results and intervals of different widths in the RF regression model. Note that half width represents half of the width of the threshold range and acc represents the proportion of the true value that falls into the range.

R^2 and RMSE of different regression models are shown in **Supplementary Table 3**.

DISCUSSION

In this study, we analyzed longitudinal data from the PPMI database to develop a predictive model for motor progression in patients with early PD. Five algorithms representing three model

categories (linear, ensemble and Bayesian) were developed; and the adjusted R^2 values of the models all reached 0.75. Our findings indicate that the models can practically predict the MDS-UPDRS Part III score of the coming year based on the clinically available characteristics obtained in the current year.

Our results suggest that all five algorithms (linear, ridge, Bayesian, RF and gradient boosting decision tree), categorized into linear, Bayesian and ensemble methods, have similar accuracies and features. Three common predictors were selected among the three categories: MDS-UPDRS Part III score, MOCA Score, and RBDSQ. In particular, functional status in the current year as measured by the MDS-UPDRS Part III (correlation coefficient 0.882; for more details, see the **Supplementary Table 2**) was consistent with and a relatively strong predictor of motor performance for the coming year. This is understandable, given that motor symptoms progress based on prior clinical features. Essentially, the results indicate that the worse one's baseline motor dysfunction is, the higher the MDS-UPDRS Part III score will be the following year.

Sleep disturbances such as disrupted circadian rhythm, insomnia, excessive daytime sleepiness (EDS), and rapid eye movement sleep behavior disorder (RBD) correlate with faster progression of motor symptoms and lower quality of life (Arnulf, 2005; Dulovic and Vos, 2018; Pagano et al., 2018). In our study, the scores of the scales used to evaluate RBD and EDS were important predictors for the progression of motor symptoms in the RF regression model, with RBD evaluation being a common predictor among the five algorithms. Subjective EDS in PD has been associated with advanced motor impairment and disease progression, male gender and the use of anti-parkinsonian medications (Arnulf, 2005; Dulovic and Vos, 2018). RBD is an aggravating factor of motor symptoms (Pagano et al., 2018), autonomic dysfunction, and dementia (Chahine et al., 2016). In a longitudinal analysis of early PD, the presence of RBD was found to predispose a patient toward a more aggressive phenotype characterized by a rapid progression of motor symptoms (Pagano et al., 2018). The promotion of neurodegeneration caused by sleep dysfunction has been proposed to drive further sleep alterations, creating a detrimental self-perpetuating cycle (Musiek and Holtzman, 2016).

Cognitive impairment at baseline is also significantly associated with faster disease progression and greater motor impairment, which has been identified in other studies (Velseboer et al., 2013; Fereshtehnejad et al., 2015; Reinoso et al., 2015). Fereshtehnejad et al. (2015) found that, besides UPDRS values, signs of cognitive impairment, orthostatic hypotension and rapid eye movement sleep behavior disorder at baseline evaluation, could suggest that patients will express a much faster decline in motor symptoms. This is mainly due to an increase in L-dopa non-responsive symptoms, which suggest a diffuse destruction of extra-nigrostriatal pathways in parallel with the nigrostriatal pathway (Velseboer et al., 2013).

There are some differences in the predictors selected among the models, which may be related to their different operating principles. The core idea underlying ensemble learning is to learn a series of basic classifiers from training data and then combine these relatively weak classifiers into a strong classifier. Then, the

predictive ability of the categorical variables can be given full play, allowing the ensemble learning model to achieve a predictive effect with fewer variables. We observed this behavior among our ensemble learning models as well.

Parkinson's disease is a clinically heterogeneous disease with varied progression patterns (Qian and Huang, 2019; Beheshti et al., 2020; Haumesser et al., 2020; LeWitt et al., 2020; Shen et al., 2020). In addition to genetic factors (Gao et al., 2020), the availability of objective fluid biomarkers specifically associated with motor or cognitive trajectories of PD subtypes could allow reliable prediction of clinical outcomes (Qian and Huang, 2019; Xie et al., 2019). As for blood biomarkers, studies on blood oligomeric α -synuclein showed increased quantities in patients with PD both in serum (Williams et al., 2016) and in red blood cells (RBCs) (Wang et al., 2015; Zhao et al., 2016; Daniele et al., 2018). Similarly, plasma phosphorylated α -synuclein is higher in patients with Parkinson's disease compared with controls (Foulds et al., 2013). In addition to α -synuclein, there are several substances playing essential roles in PD. The lower plasma levels of serum superoxide dismutase (SOD), total cholesterol, high-density lipoprotein cholesterol (HDL-C), and low-density lipoprotein cholesterol (LDL-C) and increased level of high-sensitivity C-reactive protein (hsCRP) were found in PD, which might be important markers to assess the PD severity (Yang et al., 2020). One study showed a significant decrease in the ubiquitous mitochondrial creatine kinase (uMtCK) activity in the PD group and a correlation between serum uMtCK activities and the disease progression rate, duration, and age at onset in PD patients (Xu et al., 2019). Compared with healthy subjects, the serum levels of Trefoil factor 3 (TFF3) and cholinesterase activity were lower, while homocysteine (Hcy) was higher in patients with Parkinson's disease dementia (PDD) and vascular parkinsonism with dementia (VPD). Significant correlations between TFF3/ChE activity/Hcy levels and PDD/VPD severities were found, including motor dysfunction, declining cognition, and mood/gastrointestinal symptoms (Zou et al., 2018). To explore the vascular, inflammatory, metabolic risk factors of dementia in PD with type 2 diabetes mellitus (DM) (PD-DM), lower LDL, and higher fibrinogen were the most significant risk factors in PD-DM with dementia (Wang et al., 2020).

Evidence suggests that measures of CSF A β 1-42, T-tau, P-tau181, and α -synuclein have prognostic and diagnostic potential in early-stage PD (Kang et al., 2013). Levels of α -synuclein in CSF are decreased in PD and other synucleinopathies and may serve as a marker to assist in diagnosis and prognosticating progression (Hong et al., 2010; Mollenhauer et al., 2011). Research has revealed that the NLR family pyrin domain containing 3 protein (NLRP3) inflammasome may facilitate the secretion of extracellular vesicles, as well as exosomal transmission of proteins like aggregated α -synuclein (Si et al., 2020). However, in prodromal and early PD, CSF α -synuclein does not correlate with PD's progression and does not reflect ongoing dopaminergic neurodegeneration (Mollenhauer et al., 2019). In our study, CSF biomarkers and serum uric acid had no significant impact on the progression model, indicating that easily accessible clinical assessments have sufficient capacity to predict disease progression.

Previously, based on the PPMI database, Latourelle et al. developed comprehensive multivariable prognostic models to predict the annual rate of change in PD (Latourelle et al., 2017). There are several differences between the two studies. The first one concerns to be the methodology and objectives. The previous study mainly focused on causal analysis, which differs from our research on prediction analysis. Although the value of R^2 was small in the previous study, causal analysis can also be conducted to determine the effect of independent variables on dependent variables (Latourelle et al., 2017). In Latourelle's study, the R^2 value of the model was 41% for the PPMI database and 9% for the Lab-PD cohort. The shortcomings of a small R^2 can be offset by a large sample size. In our work, maximization of the R^2 value was critical for prediction. The adjusted R^2 value for the models in each of the three categories reached 0.75, indicating that the selected features had a good predictive ability for motor scores. In general, predicting the patient's future condition can assist doctors in making decisions on when to intervene. Second, the selected features varied between the two studies. The model variables in Latourelle et al.'s study included many clinical, genetic and laboratory examinations. However, the variables in our study were all general data and clinical evaluation indicators, which could be obtained by a single doctor in the outpatient department without the need for relatively complex laboratory and imaging examinations. Thus, the use of these variables can greatly enhance the clinical practicality of our model.

Our research possesses a number of strengths. First, our prediction model was developed based on data from the previous year in order to predict annual motor symptoms, a dynamic process that has a high degree of practical clinical application value. Additionally, this study included information that was easy to obtain in the patient interview process as model features, making the model feasible for practical application. Doctors can embed the model creation process in an electronic medical record system to predict the next year's motor function automatically based on the patient's current clinical data. Second, the PPMI database contains data from different hospitals in different regions, which helps improve the accuracy of prediction.

However, there are also several limitations in our study. First, the subtypes of PD were not considered, and only uniform predictions across subtypes were made. Second, only the MDS-UPDRS Part III total score was predicted as the model result, and no subdivision prediction was made for a single item or symptom category score (such as limb rigidity, central axis slowing, tremor, gait, etc.). It is possible that the relevant trends in variability would not be reflected in the total score. Third, our study focused on the early stage of PD. Thus, the model does not apply to patients with advanced PD.

CONCLUSIONS

In the present study, by using machine learning and routinely gathered assessments, we developed convenient predictive models synthesizing multiple clinical characteristics to provide 75% accuracy in predicting motor progression. CSF biomarkers and serum uric acid had no significant impact on the progression

model, indicating that easily accessible clinical assessments have sufficient capacity to predict disease progression. The use of these models can predict motor evaluations at the individual level, allowing clinicians to tailor medical management for each patient and identify at-risk patients for future clinical trials investigating delaying motor progression. Future predictive models based on other large cohorts, assigned into training, and validation sets are needed to verify the accuracy of our results.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: www.ppmi-info.org.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Parkinson's Research and the National Institute of Neurological Disorders and Stroke (1P20NS092529-01). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

L-YM and YT participated in the design of the study, drafted the manuscript, and carried out the conceptualization of the study. C-RP, Z-LC, YL, and KR performed data analysis and drafted the manuscript. TF and J-SL carried out the conceptualization of the study, reviewing, and critiquing the article at the same time. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by Natural Science Foundation of China (Nos. 82071422 and 81771367), National Key R&D Program of China (No. 2016YFC1306501), Capital Characteristic Clinic Project (No. Z171100001017041), Beijing Municipal Science and Technology Commission (Nos. Z151100003915117 and Z151100003915150), and Beijing Natural Science Foundation (No. 7164254). This work was also supported by grants from the Michael J. Fox Foundation for Parkinson's Research and the National Institute of Neurological Disorders and Stroke (Grant No. 1P20NS092529-01). PPMI—a public-private partnership—is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Avid, Biogen, BioLegend, Bristol-Myers Squibb, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, Sanofi, Servier, Takeda, Teva, UCB, and Golub Capital.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnagi.2020.627199/full#supplementary-material>

REFERENCES

- Aleksovski, D., Miljkovic, D., Bravi, D., and Antonini, A. (2018). Disease progression in Parkinson subtypes: the PPMI dataset. *Neurol. Sci.* 39, 1971–1976. doi: 10.1007/s10072-018-3522-z
- Arnulf, I. (2005). Excessive daytime sleepiness in parkinsonism. *Sleep Med. Rev.* 9, 185–200. doi: 10.1016/j.smrv.2005.01.001
- Baumann, C. R., Held, U., Valko, P. O., Wienecke, M., and Waldvogel, D. (2014). Body side and predominant motor features at the onset of Parkinson's disease are linked to motor and nonmotor progression. *Mov. Disord.* 29, 207–213. doi: 10.1002/mds.25650
- Beheshti, I., Mishra, S., Sone, D., Khanna, P., and Matsuda, H. (2020). T1-weighted MRI-driven brain age estimation in Alzheimer's disease and Parkinson's disease. *Aging Dis.* 11, 618–628. doi: 10.14336/AD.2019.0617
- Chahine, L. M., Xie, S. X., Simuni, T., Tran, B., Postuma, R., Amara, A., et al. (2016). Longitudinal changes in cognition in early Parkinson's disease patients with REM sleep behavior disorder. *Parkinsonism Relat. Disord.* 27, 102–106. doi: 10.1016/j.parkreldis.2016.03.006
- Daniele, S., Frosini, D., Pietrobbono, D., Petrozzi, L., Lo Gerfo, A., Baldacci, F., et al. (2018). α -Synuclein heterocomplexes with β -amyloid are increased in red blood cells of Parkinson's disease patients and correlate with disease severity. *Front. Mol. Neurosci.* 11:53. doi: 10.3389/fnmol.2018.00053
- Deng, X., Xiao, B., Allen, J. C., Ng, E., Foo, J. N., Lo, Y. L., et al. (2019). Parkinson's disease GWAS-linked park16 carriers show greater motor progression. *J. Med. Genet.* 56, 765–768. doi: 10.1136/jmedgenet-2018-105661
- Dulovic, M., and Vos, M. (2018). Sleep dysfunction in Parkinson's disease: novel molecular mechanism and implications for therapy. *Mov. Disord.* 33, 1558–1559. doi: 10.1002/mds.27495
- Fereshtehnejad, S. M., Romenets, S. R., Anang, J. B., Latreille, V., Gagnon, J. F., and Postuma, R. B. (2015). New clinical subtypes of parkinson disease and their longitudinal progression: a prospective cohort comparison with other phenotypes. *JAMA Neurol.* 72, 863–873. doi: 10.1001/jamaneurol.2015.0703
- Foltniey, T., Brayne, C., and Barker, R. A. (2002). The heterogeneity of idiopathic Parkinson's disease. *J. Neurol.* 249:138–145. doi: 10.1007/PL00007856
- Foulds, P. G., Diggle, P., Mitchell, J. D., Parker, A., Hasegawa, M., Masuda-Suzukake, M., et al. (2013). A longitudinal study on α -synuclein in blood plasma as a biomarker for Parkinson's disease. *Sci. Rep.* 3:2540. doi: 10.1038/srep02540
- Gao, X., Huang, Z., Feng, C., Guan, C., Li, R., Xie, H., et al. (2020). Multimodal analysis of gene expression from postmortem brains and blood identifies synaptic vesicle trafficking genes to be associated with Parkinson's disease. *Brief Bioinform.* doi: 10.1093/bib/bbaa244. [Epub ahead of print].
- Gottipati, G., Karlsson, M. O., and Plan, E. L. (2017). Modeling a composite score in Parkinson's disease using item response theory. *AAPS J.* 19, 837–845. doi: 10.1208/s12248-017-0058-8
- Gramotnev, G., Gramotnev, D. K., and Gramotnev, A. (2019). Parkinson's disease prognostic scores for progression of cognitive decline. *Sci. Rep.* 9:17485. doi: 10.1038/s41598-019-54029-w
- Haumesser, J. K., Beck, M. H., Pellegrini, F., Kühn, J., Neumann, W. J., Altschüler, J., et al. (2020). Subthalamic beta oscillations correlate with dopaminergic degeneration in experimental parkinsonism. *Exp. Neurol.* 335:113513. doi: 10.1016/j.expneurol.2020.113513
- Hogue, O., Fernandez, H. H., and Floden, D. P. (2018). Predicting early cognitive decline in newly-diagnosed Parkinson's patients: a practical model. *Parkinsonism Relat. Disord.* 56, 70–75. doi: 10.1016/j.parkreldis.2018.06.031
- Holden, S. K., Finseth, T., Sillau, S. H., and Berman, B. D. (2018). Progression of MDS-UPDRS scores over five years in *de novo* Parkinson disease from the parkinson's progression markers initiative cohort. *Mov. Disord. Clin. Pract.* 5, 47–53. doi: 10.1002/mdc3.12553
- Hong, Z., Shi, M., Chung, K. A., Quinn, J. F., Peskind, E. R., Galasko, D., et al. (2010). DJ-1 and alpha-synuclein in human cerebrospinal fluid as biomarkers of Parkinson's disease. *Brain* 133, 713–726. doi: 10.1093/brain/awq008
- Kang, J. H., Irwin, D. J., Chen-Plotkin, A. S., Siderowf, A., Caspell, C., Coffey, C. S., et al. (2013). Association of cerebrospinal fluid β -amyloid 1-42, T-tau, P-tau181, and α -synuclein levels with clinical features of drug-naïve patients with early Parkinson disease. *JAMA Neurol.* 70, 1277–1287. doi: 10.1001/jamaneurol.2013.3861
- Latourelle, J. C., Beste, M. T., Hadzi, T. C., Miller, R. E., Oppenheim, J. N., Valko, M. P., et al. (2017). Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson's disease: a longitudinal cohort study and validation. *Lancet Neurol.* 16, 908–916. doi: 10.1016/S1474-4422(17)30328-9
- LeWitt, P. A., Kymes, S., and Hauser, R. A. (2020). Parkinson disease and orthostatic hypotension in the elderly: recognition and management of risk factors for falls. *Aging Dis.* 11, 679–691. doi: 10.14336/AD.2019.0805
- Ma, L. Y., Chan, P., Gu, Z. Q., Li, F. F., and Feng, T. (2015). Heterogeneity among patients with Parkinson's disease: cluster analysis and genetic association. *J. Neurol. Sci.* 351, 41–45. doi: 10.1016/j.jns.2015.02.029
- Mollenhauer, B., Caspell-Garcia, C. J., Coffey, C. S., Taylor, P., Singleton, A., Shaw, L. M., et al. (2019). Longitudinal analyses of cerebrospinal fluid α -synuclein in prodromal and early Parkinson's disease. *Mov. Disord.* 34, 1354–1364. doi: 10.1002/mds.27806
- Mollenhauer, B., Locascio, J. J., Schulz-Schaeffer, W., Sixel-Döring, F., Trenkwalder, C., and Schlossmacher, M. G. (2011). α -Synuclein and tau concentrations in cerebrospinal fluid of patients presenting with parkinsonism: a cohort study. *Lancet Neurol.* 10, 230–240. doi: 10.1016/S1474-4422(11)70014-X
- Musiek, E. S., and Holtzman, D. M. (2016). Mechanisms linking circadian clocks, sleep, and neurodegeneration. *Science* 354, 1004–1008. doi: 10.1126/science.aah4968
- Pagano, G., De Micco, R., Yousaf, T., Wilson, H., Chandra, A., and Politis, M. (2018). REM behavior disorder predicts motor progression and cognitive decline in Parkinson disease. *Neurology* 91, e894-894.e905. doi: 10.1212/WNL.00000000000006134
- Parnetti, L., Gaetani, L., Eusebi, P., Paciotti, S., Hansson, O., El-Agnaf, O., et al. (2019). CSF and blood biomarkers for Parkinson's disease. *Lancet Neurol.* 18, 573–586. doi: 10.1016/S1474-4422(19)30024-9
- Qian, E., and Huang, Y. (2019). Subtyping of Parkinson's Disease - Where Are We Up To. *Aging Dis.* 10, 1130–1139. doi: 10.14336/AD.2019.0112
- Reinoso, G., Allen, J. C. Jr, Au, W. L., Seah, S. H., Tay, K. Y., and Tan, L. C. (2015). Clinical evolution of Parkinson's disease and prognostic factors affecting motor progression: 9-year follow-up study. *Eur. J. Neurol.* 22, 457–463. doi: 10.1111/ene.12476
- Schrag, A., Siddiqui, U. F., Anastasiou, Z., Weintraub, D., and Schott, J. M. (2017). Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson's disease: a cohort study. *Lancet Neurol.* 16, 66–75. doi: 10.1016/S1474-4422(16)30328-3
- Selikhova, M., Williams, D. R., Kempster, P. A., Holton, J. L., Revesz, T., and Lees, A. J. (2009). A clinico-pathological study of subtypes in Parkinson's disease. *Brain* 132, 2947–2957. doi: 10.1093/brain/awp234
- Shen, Y. T., Yuan, Y. S., Wang, M., Zhi, Y., Wang, J. W., Wang, L. N., et al. (2020). Dysfunction in superior frontal gyrus associated with diphasic dyskinesia in Parkinson's disease. *NPJ Parkinsons. Dis.* 6:30. doi: 10.1038/s41531-020-00133-y
- Si, X. L., Fang, Y. J., Li, L. F., Gu, L. Y., Yin, X. Z., Yan YP, et al. (2020). From inflammasome to Parkinson's disease: does the NLRP3 inflammasome facilitate exosome secretion and exosomal alpha-synuclein transmission in Parkinson's disease. *Exp Neurol.* doi: 10.1016/j.expneurol.2020.113525. [Epub ahead of print].
- Simuni, T., Long, J. D., Caspell-Garcia, C., Coffey, C. S., Lasch, S., Tanner, C. M., et al. (2016). Predictors of time to initiation of symptomatic therapy in early Parkinson's disease. *Ann. Clin. Transl. Neurol.* 3, 482–494. doi: 10.1002/acn3.317
- Tsiouris, K. M., Rigas, G., Gatsios, D., Antonini, A., Konitsiotis, S., Koutsouris, D. D., et al. (2017). Predicting rapid progression of Parkinson's disease at baseline patients evaluation. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2017, 3898–3901. doi: 10.1109/EMBC.2017.8037708
- Velseboer, D. C., Broeders, M., Post, B., van Geloven, N., Speelman, J. D., Schmand, B., et al. (2013). Prognostic factors of motor impairment, disability, and quality of life in newly diagnosed PD. *Neurology* 80, 627–633. doi: 10.1212/WNL.0b013e318281cc99
- Vu, T. C., Nutt, J. G., and Holford, N. H. (2012). Progression of motor and nonmotor features of Parkinson's disease and their response to treatment. *Br. J. Clin. Pharmacol.* 74, 267–283. doi: 10.1111/j.1365-2125.2012.04192.x
- Wang, T., Yuan, F., Chen, Z., Zhu, S., Chang, Z., Yang, W., et al. (2020). Vascular, inflammatory and metabolic risk factors in relation to dementia

- in Parkinson's disease patients with type 2 diabetes mellitus. *Aging* 12, 15682–15704. doi: 10.18632/aging.103776
- Wang, X., Yu, S., Li, F., and Feng, T. (2015). Detection of α -synuclein oligomers in red blood cells as a potential biomarker of Parkinson's disease. *Neurosci. Lett.* 599, 115–119. doi: 10.1016/j.neulet.2015.05.030
- Williams, S. M., Schulz, P., and Sierks, M. R. (2016). Oligomeric α -synuclein and β -amyloid variants as potential biomarkers for Parkinson's and Alzheimer's diseases. *Eur. J. Neurosci.* 43, 3–16. doi: 10.1111/ejn.13056
- Xie, F., Gao, X., Yang, W., Chang, Z., Yang, X., Wei, X., et al. (2019). Advances in the research of risk factors and prodromal biomarkers of Parkinson's disease. *ACS Chem. Neurosci.* 10, 973–990. doi: 10.1021/acscchemneuro.8b00520
- Xu, J., Fu, X., Pan, M., Zhou, X., Chen, Z., Wang, D., et al. (2019). Mitochondrial creatine kinase is decreased in the serum of idiopathic Parkinson's disease patients. *Aging Dis.* 10, 601–610. doi: 10.14336/AD.2018.0615
- Yamazaki, Y., and Kanekiyo, T. (2017). Blood-brain barrier dysfunction and the pathogenesis of Alzheimer's disease. *Int. J. Mol. Sci.* 18:1965. doi: 10.3390/ijms18091965
- Yang, W., Chang, Z., Que, R., Weng, G., Deng, B., Wang, T., et al. (2020). Contra-directional expression of plasma superoxide dismutase with lipoprotein cholesterol and high-sensitivity C-reactive protein as important markers of Parkinson's disease severity. *Front. Aging Neurosci.* 12:53. doi: 10.3389/fnagi.2020.00053
- Zhao, H. Q., Li, F. F., Wang, Z., Wang, X. M., and Feng, T. (2016). A comparative study of the amount of α -synuclein in ischemic stroke and Parkinson's disease. *Neurol. Sci.* 37, 749–754. doi: 10.1007/s10072-016-2485-1
- Zou, J., Chen, Z., Liang, C., Fu, Y., Wei, X., Lu, J., et al. (2018). Trefoil factor 3, cholinesterase and homocysteine: potential predictors for Parkinson's disease dementia and vascular parkinsonism dementia in advanced stage. *Aging Dis.* 9, 51–65. doi: 10.14336/AD.2017.0416

Conflict of Interest: Z-LC, YL, and KR were employed by company Gyenno Science Co. Ltd., Shenzhen, China.

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

Copyright © 2021 Ma, Tian, Pan, Chen, Ling, Ren, Li and Feng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.