



## Article

# Machine Learning for Early Detection of Cognitive Decline in Parkinson's Disease Using Multimodal Biomarker and Clinical Data

Raziye Mohammadi <sup>1</sup>, Samuel Y. E. Ng <sup>2</sup>, Jayne Y. Tan <sup>3</sup>, Adeline S. L. Ng <sup>1,3</sup>, Xiao Deng <sup>3</sup>, Xinyi Choi <sup>2</sup>, Dede L. Heng <sup>2</sup>, Shermyn Neo <sup>3</sup>, Zheyu Xu <sup>3</sup>, Kay-Yaw Tay <sup>1,3</sup>, Wing-Lok Au <sup>1,3</sup>, Eng-King Tan <sup>2,3</sup>, Louis C. S. Tan <sup>1,2,3</sup>, Ewout W. Steyerberg <sup>4</sup>, William Greene <sup>5</sup> and Seyed Ehsan Saffari <sup>1,3,\*</sup>

<sup>1</sup> Duke-NUS Medical School, National University of Singapore, Singapore 169857, Singapore; raziye.mohammadi@duke-nus.edu.sg (R.M.); adeline.ng.s.l@singhealth.com.sg (A.S.L.N.); tay.kay.yaw@singhealth.com.sg (K.-Y.T.); au.wing.lok@singhealth.com.sg (W.-L.A.); louis.tan.c.s@singhealth.com.sg (L.C.S.T.)

<sup>2</sup> Department of Research, National Neuroscience Institute, Singapore 308433, Singapore; samuel.ng.y.e@singhealth.com.sg (S.Y.E.N.); choi.xinyi@singhealth.com.sg (X.C.); dede.liana.heng@singhealth.com.sg (D.L.H.); tan.eng.king@singhealth.com.sg (E.-K.T.)

<sup>3</sup> Department of Neurology, National Neuroscience Institute, Singapore 308433, Singapore; jayne.tan.yi@singhealth.com.sg (J.Y.T.); deng.xiao@sgh.com.sg (X.D.); shermyn.neo.x.m@singhealth.com.sg (S.N.); xu.zhe.yu@singhealth.com.sg (Z.X.)

<sup>4</sup> Department of Biomedical Data Sciences, Leiden University Medical Center, 2333 ZD Leiden, The Netherlands; e.w.steyerberg@lumc.nl

<sup>5</sup> Department of Econometrics, Stern School of Business, New York University, New York, NY 10012, USA; wgreene@stern.nyu.edu

\* Correspondence: ehsan.saffari@duke-nus.edu.sg



**Citation:** Mohammadi, R.; Ng, S.Y.E.; Tan, J.Y.; Ng, A.S.L.; Deng, X.; Choi, X.; Heng, D.L.; Neo, S.; Xu, Z.; Tay, K.-Y.; et al. Machine Learning for Early Detection of Cognitive Decline in Parkinson's Disease Using Multimodal Biomarker and Clinical Data. *Biomedicines* **2024**, *12*, 2758. <https://doi.org/10.3390/biomedicines12122758>

Academic Editors: Simone Battaglia and Masaru Tanaka

Received: 3 October 2024

Revised: 25 November 2024

Accepted: 29 November 2024

Published: 3 December 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract: Background:** Parkinson's disease (PD) is the second most common neurodegenerative disease, primarily affecting the middle-aged to elderly population. Among its nonmotor symptoms, cognitive decline (CD) is a precursor to dementia and represents a critical target for early risk assessment and diagnosis. Accurate CD prediction is crucial for timely intervention and tailored management of at-risk patients. This study used machine learning (ML) techniques to predict the CD risk over five-year in early-stage PD. **Methods:** Data from the Early Parkinson's Disease Longitudinal Singapore (2014 to 2018) was used to predict CD defined as a one-unit annual decrease or a one-unit decline in Montreal Cognitive Assessment over two consecutive years. Four ML methods—**AutoScore, Random Forest, K-Nearest Neighbors and Neural Network**—were applied using baseline demographics, clinical assessments and blood biomarkers. **Results:** Variable selection identified key predictors of CD, including education year, diastolic lying blood pressure, diastolic standing blood pressure, systolic lying blood pressure, Hoehn and Yahr scale, body mass index, phosphorylated tau at threonine 181, total tau, Neurofilament light chain and suppression of tumorigenicity 2. Random Forest was the most effective, achieving an AUC of 0.93 (95% CI: 0.89, 0.97), using 10-fold cross-validation. **Conclusions:** Here, we demonstrate that ML-based models can identify early-stage PD patients at high risk for CD, supporting targeted interventions and improved PD management.

**Keywords:** neurodegenerative disease; cognitive dysfunction; Parkinson's disease; early diagnosis; risk assessment; machine learning; blood biomarkers; Montreal cognitive assessment; nonmotor symptoms; predictive models

## 1. Introduction

Parkinson's disease (PD) is a prevalent neurodegenerative disorder impacting millions worldwide, characterized by the gradual deterioration of both motor and non-motor functions [1–3]. While motor symptoms such as bradykinesia, resting tremor, rigidity,

and postural instability are typically the earliest manifestations, PD is increasingly recognized for its extensive non-motor symptoms, including cognitive dysfunction, autonomic disturbances, and psychiatric issues [4]. Among these, cognitive decline is particularly concerning, as it is associated with a significant risk of progression to dementia, with nearly half of PD patients developing dementia within a decade of diagnosis [5–8]. Cognitive decline often begins early in the disease course and is accompanied by structural brain changes, including gray matter alterations in the temporal regions, hippocampus, frontal, and parietal lobes, as well as white matter changes in the corpus callosum and cingulate gyrus [9,10]. Recent studies have highlighted the prevalence of mild cognitive impairment in PD (PD-MCI), which affects 20–33% of patients at diagnosis and serves as a precursor to Parkinson’s disease dementia (PDD) [11,12]. Notably, 60–80% of those with PD-MCI progress to PDD within 12 years [13,14]. Furthermore, approximately 30–35% of individuals with early-stage PD experience cognitive decline, emphasizing the need for early detection and intervention [15,16].

Given the profound impact of cognitive decline on PD patients and the significant economic burden it imposes, developing an accurate and cost-effective predictive model for cognitive decline in early PD stages is crucial [17]. Early identification of cognitive decline offers a valuable opportunity for timely interventions to enhance cognitive reserve, preserve cognitive function, and potentially prevent further deterioration. However, the considerable heterogeneity in cognitive trajectories among PD patients complicates prognostication and poses challenges for clinical trials aimed at addressing this critical aspect of the disease [18]. Therefore, predictive biomarkers for cognitive decline in PD are urgently needed to improve early detection and provide insights into the mechanisms driving cognitive decline in certain PD patients while sparing others [19]. Despite this urgency, practical methodologies integrating baseline demographics, clinical assessments, and blood biomarkers for early detection of PD-related cognitive decline remain limited. Addressing this critical gap, this study aims to provide a straightforward, reliable, and easy-to-use model for predicting cognitive decline in early PD using machine learning (ML) techniques and accessible baseline features.

Biomarkers play a pivotal role in PD, offering promise for early diagnosis, disease monitoring, and clinical trial design. Substantial evidence supports the idea that converting  $\alpha$ -synuclein from soluble monomers to aggregated, insoluble forms in the brain is a hallmark of PD pathology [20]. This pathological process may also be found in human bodily fluids such as cerebrospinal fluid (CSF) and blood plasma [21,22]. Extensive research has delved into various biomarker types, encompassing clinical, genetic, CSF, and imaging biomarkers. These are increasingly pivotal in predicting cognitive decline during early diagnosis and disease prognostication [23–27]. Specifically, in CSF, higher levels of phosphorylated tau (p-tau) and lower levels of amyloid  $\beta$ 42 have been linked to an elevated risk of dementia in PD patients [28,29]. Additionally, elevated neurofilament light chain (NfL) levels in CSF have been shown to predict cognitive decline, further underscoring the importance of these biomarkers in understanding and managing the disease [30].

While CSF biomarkers have shown significant promise, blood biomarkers stand out as particularly advantageous in PD due to their accessibility and cost-effectiveness compared to CSF and imaging biomarkers [31,32]. Prior research has explored the relationship between blood biomarkers and cognitive decline in early PD. For instance, increased physical activity has been shown to attenuate the vulnerability associated with the apolipoprotein E  $\epsilon$ 4 (APOE  $\epsilon$ 4) allele to early cognitive decline in patients with PD [33]. Another study highlighted the potential of elevated  $\alpha$ -synuclein and total tau (t-tau), along with reduced amyloid-beta-40 (A $\beta$ -40) levels, as biomarkers for the early detection of cognitive impairment in PD patients [34]. Additionally, a pilot study suggested that lower serum uric acid levels in the early stages of the disease may be associated with the later development of MCI [35]. Recent findings by Sekiya, et al. (2022) [36] further highlight the widespread presence of  $\alpha$ -synuclein oligomers in various brain regions of PD patients, especially in the neocortex, and their association with cognitive impairment, suggesting their potential

significance in early PD pathology. Furthermore, elevated plasma NfL levels and reduced epidermal growth factor levels have been linked to cognitive decline in PD patients [37–39]. Despite these advancements, studies investigating plasma biomarkers and correlations to cognitive decline in PD remain limited.

To the best of our knowledge, no existing risk prediction models have utilized ML methods incorporating baseline clinical, demographic, and blood biomarkers to predict the risk of cognitive decline in early PD. This study aims to fill this gap by developing a risk prediction model using ML algorithms capable of detecting complex patterns and interactions not discernible through traditional analysis methods.

Previous studies have explored various approaches to predicting cognitive decline in PD using different data sources and methodologies. For instance, a study used data from the Parkinson's Progression Markers Initiative (PPMI) to accurately predict cognitive impairment at a 2-year follow-up [40]. Combining age, non-motor assessments, dopamine transporter (DAT) imaging, and CSF biomarkers effectively predicted Montreal Cognitive Assessment (MoCA) scores at the 2-year follow-up in newly diagnosed PD patients. Another study, also using PPMI data, developed a multimodal ML model to predict cognitive decline in early PD patients by utilizing the change in MoCA scores as the outcome, calculated from the difference between the baseline and 4-year follow-up data [41]. Additionally, a cross-sectional study using data from the Early Parkinson's disease Longitudinal Singapore (PALS) study examined cognitive impairment by comparing PD-MCI patients and those with normal cognition (PD-NC). This study highlighted the significant associations between PD-MCI and several factors, including triglycerides (TG), apolipoprotein A1 (ApoA1), and the SNCA rs6826785 genetic marker, suggesting their potential role in early cognitive decline in PD patients [42].

Building on these insights, the current study utilized data from the PALS cohort to develop a risk prediction model for predicting cognitive decline over a five-year period in individuals with early PD by incorporating baseline characteristics and employing various ML algorithms.

## 2. Materials and Methods

### 2.1. Study Design and Population

This study utilized data from the PALS prospective cohort study to predict cognitive decline using ML techniques. Data were used from 214 PD patients, collected over five years between 2014 and 2018, with all participants meeting the National Institute of Neurological Disorders and Stroke (NINDS) clinical criteria for PD. Participants had to have more than 6 years of education and were able to read and write English or Mandarin to enroll in the study. Exclusion criteria included significant medical conditions hindering regular follow-up and orthopedic issues potentially affecting study outcomes. Functional status was measured using the Hoehn and Yahr (HY) rating scale, while motor symptom severity was assessed via the Movement Disorder Society-Unified Parkinson's disease Rating Scale (MDS-UPDRS) Part III. Cognitive function was assessed using the MoCA. Where dopaminergic therapy had begun, the dosage was calculated and reported as cumulative levodopa equivalent daily dose (LEDD) [43,44].

Patients were defined as having 'early PD' based on the following inclusion criteria: (i) motor symptoms within two years, and (ii) diagnosis of PD within one year according to the NINDS criteria as determined by a specialist in movement disorders. Ethics approval was obtained from the Singapore Health Services Centralized Institutional Review Board (CIRB) for the use of human participants in this study, and all participants provided informed written consent. After excluding patients with missing MoCA scores for the first three years, 193 PD patients were included in the final analysis.

### 2.2. Outcome Definition

In the context of early PD, the MoCA is used as a key measure to assess cognitive decline. For the purposes of this study, cognitive decline was defined as either a one-unit

annual decrease in MoCA score or a one-unit decline observed over two consecutive years during the five-year follow-up period. This threshold of a one-unit decline is clinically relevant, as even a small change can indicate an early sign of cognitive deterioration in early PD.

### 2.3. Input Variables

Input variables included baseline demographics, clinical assessments, and blood biomarkers. The baseline demographics included age, gender, years of education, body mass index (BMI), smoking status, alcohol consumption, coffee consumption, and tea consumption. Clinical assessments encompassed standing and lying systolic blood pressure (SBP), standing and lying diastolic blood pressure (DBP), HY, total MoCA score, total motor score, diabetes mellitus, hypertension, and hyperlipidemia. The blood biomarkers analyzed were suppression of tumorigenicity 2 (ST2), NFL, t-tau, phosphorylated tau at threonine 181 (p-tau181), apolipoprotein E (APOE), and alpha-synuclein gene promoter (REP1).

### 2.4. Data Imputation and Transformation

Missing values were imputed using a random forest-based imputation method, which estimates missing values by leveraging relationships observed in existing data [45]. This approach imputes missing data using mean/mode, and then iteratively fits a random forest (RF) to predict the missing values until a stopping criterion or maximum iterations are reached. Continuous input features were transformed into binary variables for easier clinical interpretation. For input variables without well-established cut-off points, including blood biomarkers and total motor score, the Youden index was used. This method, which incorporates both sensitivity and specificity, is a commonly used measure of overall diagnostic performance. It identifies the cut-off point that optimizes the biomarker's differentiating ability when equal weight is assigned to both sensitivity and specificity [46].

### 2.5. Feature Selection

Initially, 24 variables were included in the study. The RF importance was calculated for each feature. The mean RF importance across all features was used as a threshold, and features with an RF importance below mean were excluded from the dataset.

### 2.6. Statistical Analysis and ML Methods

Descriptive statistics, including mean and standard deviation (SD) or median and first and third quartile, were reported for numeric variables, depending on the normality assumption, while categorical variables were presented as frequency and percentages. Univariate logistic regression analysis was performed to investigate the association of baseline patients' characteristics with progression outcome, and odds ratios (OR) along with 95% confidence interval (CI) were calculated. Four ML methods including AutoScore, RF, k-nearest neighbors (KNN), and neural network (NN), along with logistic regression as a baseline statistical approach, were employed. To ensure optimal model performance, hyperparameter tuning was performed for all methods using 10-fold cross-validation. For RF, a grid search was employed to tune the number of trees (ntree), the number of variables selected at each split (mtry), the minimum node size (nodesize), and the maximum number of terminal nodes (maxnodes). The following values were explored: ntree = (100, 200, 500), mtry = (3, 4, 5), nodesize = (5, 10, 15), and maxnodes = (5, 10, 20). The optimal parameters were chosen based on cross-validated performance metrics, specifically the area under the curve (AUC). For NN, one hidden layer was considered, and hyperparameters including the number of hidden units (size = (1, 2, 3, 4, 5)) and weight decay (decay = (0, 0.01, 0.1)) were optimized, with accuracy as the performance metric. Similarly, for KNN, the number of neighbors was optimized within the range of 1 to 10 using grid search, also with accuracy as the performance metric. For AutoScore, as all variables were initially binary, a score table was generated from the model outputs to create interpretable clinical scores. Following hyperparameter tuning, models with the optimal parameters were compared

using performance metrics including AUC, sensitivity, and specificity, along with their corresponding 95% CI. Sensitivity and specificity were determined by identifying the optimal threshold on the receiver operating characteristic (ROC) curve, defined as the point closest to the top-left corner, representing a balance between high sensitivity and specificity.

Each ML method was evaluated using the following two modeling strategies: (i) including all variables, and (ii) selecting the ten most important variables based on feature importance scores derived from the RF importance approach. AUC as an overall accuracy metric was used to identify the best-performing model. Model calibration was evaluated using a binned plot, which is recommended for smaller datasets. In this approach, predicted probabilities are grouped into 10 equal-sized bins, and for each bin, the midpoint of the predicted probability is plotted against the true fraction of positive cases. If the model is well calibrated, the points will fall near the diagonal line [47]. A risk score table for the model utilizing selected variables was generated using the AutoScore method, an easy-to-use ML algorithm designed to facilitate risk assessment [48]. Statistical significance was set at  $p$ -value < 0.05. All data analyses were conducted using R software 4.4.2. (R Core Team (2024); R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>, accessed on 30 November 2024).

### 3. Results

#### 3.1. Baseline Characteristics and Descriptive Statistics by Outcome

A total of 193 early PD patients completed baseline assessments and were included in our study, with 58% of the participants being male. At baseline, the mean age was 63.6 years (SD = 8.94 years), and the mean years of education was 10.7 years (SD = 4.37 years). The descriptive statistics of baseline variables are detailed in Table 1. Cognitive decline as the primary outcome was observed in 44 (23%) subjects. Significant findings include that patients with 10 or more years of education had a lower risk of cognitive decline compared to those with fewer years (OR = 0.36,  $p$ -value = 0.006). Elevated lying SBP, lying DBP, and standing SBP were associated with a higher risk of cognitive decline (OR = 2.13,  $p$ -value = 0.045; OR = 2.60,  $p$ -value = 0.009; OR = 2.13,  $p$ -value = 0.045, respectively).

**Table 1.** Summary statistics of demographic, clinical assessments, blood biomarkers, and their associations with progression outcomes using univariate logistic regression.

	Total	No Progression	Progression	OR (95% CIs)	$p$ -Value
	N = 193	N = 149	N = 44		
<b>Demographic Characteristics</b>					
Male Gender	112 (58.0%)	85 (57.0%)	27 (61.4%)	1.2 (0.6, 2.4)	0.737
Smoker	56 (29.0%)	43 (28.9%)	13 (29.5%)	1.0 (0.5, 2.1)	1.000
Years of education ( $\geq 10$ years)	135 (69.9%)	112 (75.2%)	23 (52.3%)	0.4 (0.2, 0.7)	0.006
Tea drinking	180 (93.3%)	138 (92.6%)	42 (95.5%)	1.6 (0.4, 11.5)	0.736
Coffee drinking	175 (90.7%)	135 (90.6%)	40 (90.9%)	1.0 (0.3, 3.8)	1.000
Alcohol drinking	125 (64.8%)	101 (67.8%)	24 (54.5%)	0.6 (0.3, 1.1)	0.151
BMI ( $>25$ kg/m <sup>2</sup> )	64 (33.2%)	48 (32.2%)	16 (36.4%)	1.2 (0.6, 2.4)	0.740
Age ( $>65$ years)	97 (50.3%)	72 (48.3%)	25 (56.8%)	1.4 (0.7, 2.8)	0.413
<b>Clinical Assessments</b>					
Lying SBP ( $\geq 140$ mmHg)	95 (49.2%)	67 (45.0%)	28 (63.6%)	2.1 (1.1, 4.3)	0.045
Lying DBP ( $\geq 80$ mmHg)	67 (34.7%)	44 (29.5%)	23 (52.3%)	2.6 (1.3, 5.2)	0.009
Standing SBP ( $\geq 140$ mmHg)	85 (44.0%)	63 (42.3%)	22 (50.0%)	1.4 (0.7, 2.7)	0.463
Standing DBP ( $\geq 80$ mmHg)	95 (49.2%)	67 (45.0%)	28 (63.6%)	2.1 (1.1, 4.3)	0.045
Diabetes mellitus	31 (16.1%)	25 (16.8%)	6 (13.6%)	0.8 (0.3, 2.0)	0.791
Hypertension	88 (45.6%)	68 (45.6%)	20 (45.5%)	1.0 (0.5, 2.0)	1.000
Hyperlipidemia	92 (47.7%)	73 (49.0%)	19 (43.2%)	0.8 (0.4, 1.6)	0.613



Table 1. Cont.

	Total	No Progression	Progression	OR (95% CIs)	p-Value
	N = 193	N = 149	N = 44		
<b>Demographic Characteristics</b>					
MoCA	26 [23.0, 28.0]	26 [23.0, 28.0]	26 [23.0, 28.0]	1.0 (0.9, 1.1)	0.771
Total motor score	20.0 [15.0; 26.0]	19.0 [15.0; 26.0]	22.0 [17.0; 29.0]	1.0 (1.0, 1.1)	0.062
HY	2.00 [1.0; 3.0]	2.00 [1.50; 2.0]	2.00 [2.00; 2.0]	2.0 (0.8, 4.8)	0.112
<b>Blood Biomarkers</b>					
APOE4 (Non-carriers)	153 (79.3%)	120 (80.5%)	33 (75.0%)	0.7 (0.3, 1.7)	0.559
REP1 (Short)	88 (45.6%)	66 (44.3%)	22 (50.0%)	1.3 (0.6, 2.5)	0.620
ST2	11,600 [8750; 14,800]	11,500 [8400; 14,900]	12,600 [9430; 14,800]	1.0 (1.0, 1.0)	0.375
NfL	13.7 [10.1; 18.9]	13.9 [10.2; 18.7]	13.3 [9.9; 21.7]	1.0 (1.0, 1.1)	0.702
t-tau	1.17 [0.9; 1.5]	1.1 [0.9; 1.6]	1.3 [0.9; 1.5]	1.3 (0.9, 1.8)	0.350
p-tau181	20.3 [15.7; 24.8]	20.50 [15.4; 24.3]	20.1 [15.8; 28.9]	1.0 (1.0, 1.1)	0.666

Data are expressed as frequency (%) or median (quartile); *p*-values are from univariate logistic regression models assessing the association of each variable with cognitive decline progression. Abbreviations: N: number, OR: odds ratio, CIs: confidence intervals, BMI: body mass index, SBP: systolic blood pressure, DBP: diastolic blood pressure, MoCA: Montreal Cognitive Assessment, HY: Hoehn and Yahr scale, APOE: apolipoprotein E, REP1: alpha-synuclein gene promoter, ST2: suppression of tumorigenicity 2, NfL: neurofilament light chain, t-tau: total tau, p-tau181: phosphorylated tau at threonine 181.

### 3.2. Feature Selection Analysis

RF importance scores were calculated for each feature (see Figure 1). The mean RF importance across all features was used as a threshold, and features with an RF importance score below the mean were excluded. This strategy selected 10 out of the initial 24 features: lying DBP, NfL, years of education, p-tau181, ST2, BMI, lying SBP, standing DBP, t-tau, and HY scale.

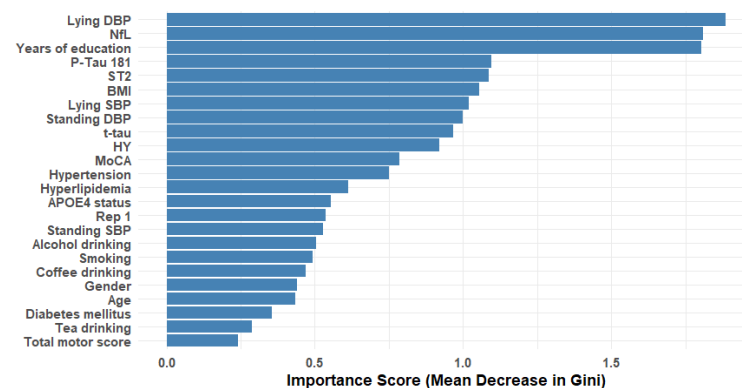


Figure 1. Feature importance ranked by mean decrease in Gini score.

### 3.3. Model Performance and ROC Analysis

The results of the various ML methods are summarized in Table 2. In Model 1, which included all variables, the RF algorithm achieved the highest AUC at 0.999, indicating near-perfect discrimination between patients with and without cognitive decline. The NN also performed exceptionally well, with an AUC of 0.996. AutoScore and logistic regression showed moderate performance, with AUCs of 0.797 and 0.806, respectively. KNN had the lowest AUC at 0.766. In terms of sensitivity, both RF and NN were outstanding, with sensitivities of 1.000 and 0.977, respectively, highlighting their excellent ability to identify patients with cognitive decline. AutoScore had the lowest sensitivity at 0.636. In Model 2, which included the ten most important variables, RF again had the highest AUC at 0.930, followed by NN and KNN with AUCs of 0.918 and 0.843, respectively. Logistic regression and AutoScore exhibited similar moderate performance, with AUCs of 0.770 and 0.771,

respectively. NN achieved the highest sensitivity at 0.841, indicating strong detection capability with fewer variables. AutoScore, RF, and KNN all performed well, each with a sensitivity of 0.818. Overall, RF consistently showed the highest AUC across both models, demonstrating superior performance in distinguishing between patients with and without cognitive decline. NN also performed strongly, particularly in Model 2, where it exhibited the highest sensitivity.

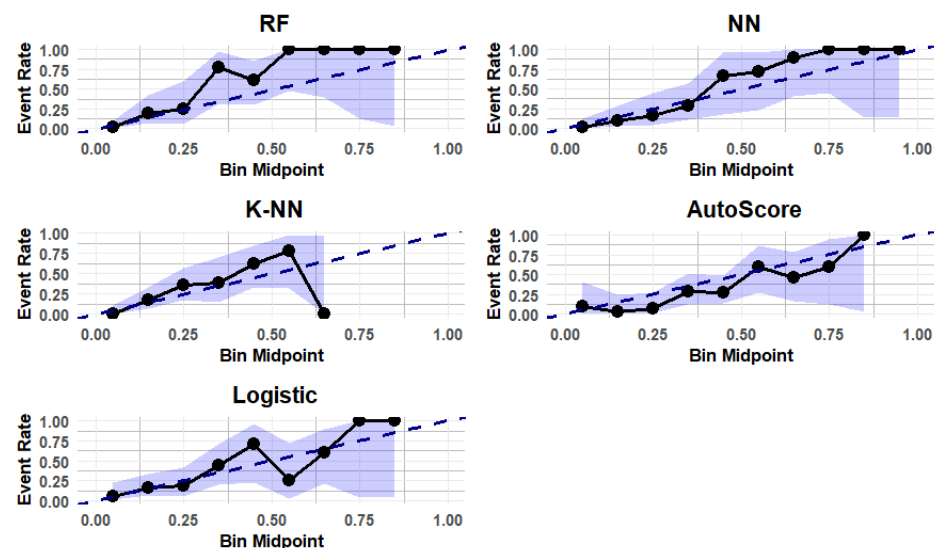
**Table 2.** The performance of four ML methods \* under two modeling strategies.

Algorithm	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
<b>Model 1: All Variables</b>			
AutoScore	0.797 (0.720, 0.8736)	0.636 (0.500, 0.773)	0.825 (0.765, 0.879)
RF	0.999 (0.997, 1.000)	1.000 (0.920, 1.000)	0.987 (0.952, 0.998)
KNN	0.766 (0.690, 0.842)	0.750 (0.597, 0.868)	0.678 (0.596, 0.752)
NN	0.996 (0.989, 1.000)	0.977 (0.880, 0.999)	0.987 (0.952, 0.998)
Logistic	0.806 (0.731, 0.881)	0.682 (0.524, 0.814)	0.819 (0.747, 0.877)
<b>Model 2: Top Ten Variables</b>			
AutoScore	0.771 (0.691, 0.851)	0.818 (0.705, 0.909)	0.631 (0.557, 0.705)
RF	0.930 (0.889, 0.971)	0.818 (0.673, 0.918)	0.872 (0.808, 0.921)
KNN	0.843 (0.788, 0.899)	0.818 (0.673, 0.918)	0.711 (0.632, 0.783)
NN	0.918 (0.872, 0.965)	0.841 (0.699, 0.934)	0.832 (0.762, 0.888)
Logistic	0.770 (0.690, 0.849)	0.795 (0.647, 0.902)	0.631 (0.548, 0.708)

\* Results obtained through ROC analysis on the training dataset using 10-fold cross-validation. Top ten variables: lying DBP, NFL, years of education, p-tau 181, ST2, BMI, lying SBP, standing DBP, t-tau, and HY.

### 3.4. Calibration of the Predictive Models

Calibration was assessed using a binned plot for both models. Figure 2 presents the calibration results specifically for Model 2, featuring a binned plot with a 99% CI based on the internal data. The plot shows that both RF and NN tend to overestimate the predicted probabilities, while KNN underestimates them. In contrast, logistic regression and AutoScore demonstrate better calibration performance.



**Figure 2.** Calibration plot with 99% CI for Model 2 across different methods. The black points indicate the observed event rates for each bin, while the black line connects these points to show the trend. The dashed dark blue diagonal line represents the ideal calibration line, indicating perfect agreement between predicted probabilities and observed event rates.

### 3.5. Score of Risk Factors Based on AutoScore Algorithm

The risk score generated for model 2 by AutoScore algorithm is summarized in Table 3. The high sensitivity demonstrated by AutoScore in this model highlights the importance of the identified risk factors. Results indicate that lower education ( $<10$  years), higher NfL ( $\geq 21.5$  pg/mL), higher ST2 ( $\geq 14,185.8$  pg/mL), higher t-tau ( $\geq 2.1$  pg/mL), higher p-tau181 ( $\geq 27.1$  pg/mL), higher standing and lying DBP ( $\geq 80$  mmHg), higher lying SBP ( $\geq 140$  mmHg), higher BMI ( $\geq 25$  kg), and higher HY scale ( $\geq 2$ ) are significant risk factors that increase the likelihood of cognitive decline in early PD patients.

**Table 3.** Risk scores generated by AutoScore algorithm \* for Model 2.

Variable	Interval	Partial Score
Lying DBP	Normal	0
	High	16
NfL	Normal	0
	High	18
Years of education	$\geq 10$	0
	$<10$	16
p-tau 181	Normal	0
	High	11
ST2	Normal	0
	High	9
BMI	$<25$	0
	$\geq 25$	3
Lying SBP	Normal	0
	High	1
Standing DBP	Normal	0
	High	11
t-tau	Normal	0
	High	11
HY	$<2$	0
	$\geq 2$	4

\* AutoScore, an interpretable ML-based tool for generating automatic clinical score, provides risk factor scores for each feature. This capability translates complex model predictions into a more understandable format for clinical decision making. Abbreviations: DBP: diastolic blood pressure, NfL: neurofilament light chain, p-tau 181: phosphorylated tau at threonine 181, ST2: suppression of tumorigenicity 2, BMI: Body mass index, SBP: systolic blood pressure, t-tau: total tau, HY: Hoehn and Yahr scale.

### 3.6. Summary of Key Findings

Overall, feature selection using RF importance scores narrowed down the variables to ten, which were used in various ML methods. The RF algorithm demonstrated the highest AUC (0.999) in Model 1, and also achieved the highest AUC (0.930) in Model 2, which incorporated the most important variables. Both RF and NN exhibited high sensitivity, particularly in identifying patients with cognitive decline. The risk factors identified through AutoScore emphasize the clinical relevance of these variables in predicting cognitive decline in early PD patients.

## 4. Discussion

Cognitive impairment is one of the most common non-motor symptoms in PD and can be more devastating for both patients and caregivers than motor symptoms [49]. Cognitive decline is increasingly recognized as a prevalent issue even in newly diagnosed PD patients [8,10]. The search for objective biomarkers is driven by their potential to enhance early and accurate diagnosis, monitor disease progression, and optimize clinical trial design and interpretation. While alpha-synuclein remains a promising biomarker candidate, the complex and heterogeneous nature of PD underscores the necessity for a comprehensive biomarker panel [31,32]. In light of these considerations, this study aimed to develop accurate predictive models for cognitive decline in early PD patients. Leveraging



a combination of blood biomarkers, clinical data, and demographic characteristics, ML techniques were employed to achieve this objective.

This study compared two different modeling strategies to predict cognitive decline in early PD patients, emphasizing the importance of practicality in clinical settings. Model 1, which included all available variables, demonstrated the highest performance metrics, particularly with the RF and NN algorithms. However, the complexity and cost associated with obtaining comprehensive datasets limit its utility in routine clinical practice. In contrast, Model 2, incorporating only the ten most significant variables identified through feature selection, not only maintained strong performance metrics but also enhanced practicality for clinicians. Notably, this model showcased remarkable sensitivity with the NN algorithm, suggesting its potential to effectively detect cognitive decline using a streamlined approach. By focusing on easily obtainable variables, such as years of education and blood pressure, Model 2 could facilitate timely interventions, allowing healthcare providers to identify patients at risk for rapid cognitive decline. Implementing such practical models in clinical workflows could significantly improve early detection and management strategies, ultimately enhancing patient outcomes in early-stage PD.

The ten primary variables derived from feature selection, comprising a blend of demographics, clinical parameters, and blood biomarkers, included years of education, BMI, NfL, t-tau, p-tau181, ST2, standing DBP, lying DBP, lying SBP, and HY scale. Notably, several of these features have been identified as significant risk factors in prior research, with further details provided below.

In this study, BMI emerged as a significant demographic variable with potential implications for managing early PD. Although research specifically examining the effect of BMI on early PD is lacking, several studies have explored its association with cognitive decline in PD. For instance, an analysis of data from PPMI identified that higher baseline BMI, along with modifiable comorbidities such as depression and sleep disorders, contributed to an accelerated rate of cognitive decline in PD patients [50]. Similarly, another study using PPMI data found that PD patients with a metabolically unhealthy normal weight (MUNW) phenotype experienced more rapid cognitive decline, particularly in global cognition and visuospatial perception, over a 48-month period compared to those in other BMI-metabolic status categories [51]. Conversely, Yoo et al. (2019) reported that PD patients with a higher-than-normal BMI at diagnosis exhibited a slower cognitive decline and a reduced risk of developing dementia over a six-year period compared to those with under/normal weight, suggesting that a higher BMI may have a protective effect against cognitive deterioration in PD [52]. Additionally, Kim et al. (2012) observed that a decrease in BMI during the initial six months of follow-up in PD patients could serve as an early indicator of future dementia risk, enabling clinicians to predict a faster rate of cognitive decline [53]. These findings underscore the importance of monitoring BMI in PD patients, as it may inform clinical decisions regarding interventions aimed at preserving cognitive function and improving overall patient outcomes.

In addition to BMI, years of education also emerged as a significant demographic predictor in this study. This finding is consistent with a recent cross-sectional study using PALS data, which demonstrated that fewer years of education are associated with higher MDS-UPDRS Part III and an elevated risk of MCI in early PD [42]. Lower educational attainment may therefore be a marker for greater vulnerability to motor and cognitive declines in PD, underscoring the potential role of education in influencing disease progression and patient outcomes.

In the present study, standing DBP, lying DBP, and lying SBP emerged as significant clinical predictors of cognitive decline in early PD. This aligns with previous research underscoring the role of hypertension in cognitive decline among early PD patients. Previous research has shown that PD-MCI patients exhibited significantly higher diastolic blood pressure variability (BPV) during follow-up compared to those with non-MCI PD, suggesting BPV as a potential predictive marker of cognitive decline [54]. Additionally, an analysis using PPMI data indicated that elevated visit-to-visit variability in systolic blood

pressure (systolic VIM) was associated with a faster decline in global cognitive function, assessed by the MoCA score, in PD-MCI patients [55]. Further emphasizing the importance of blood pressure management, another study found that, on average, every 10 mmHg increase in pulse pressure was associated with a 0.08 reduction in cognitive Z-scores in early PD [56]. These findings collectively highlight the critical need for effective blood pressure management in early PD to mitigate the risk of cognitive decline.

Through this study, the HY scale emerged as a significant predictor of cognitive decline in early PD, underscoring its clinical relevance beyond motor symptom assessment. This finding aligns with previous research demonstrating a strong association between motor impairment severity, as measured by the HY scale, and cognitive deficits in PD patients. For example, a study by Siciliano et al. (2017) compared cognitive performance in de novo PD patients and found that those at HY stage II scored significantly lower on neuropsychological tests compared to those at HY stage I, indicating that greater motor impairment correlates with increased cognitive dysfunction [57]. Additionally, the predictive power of the HY scale for disease progression is further supported by studies such as the PASADENA trial, a Phase II randomized, double-blind, placebo-controlled study investigating the efficacy and safety of prasinezumab in early PD [58], and analyses using PPMI data [59]. These studies identified the HY stage, along with other biomarkers like dopamine transporter SPECT imaging, as the key predictors of clinical progression in early PD. These findings highlight the critical role of the HY scale in the early detection and management of PD, aiding clinicians in predicting and potentially mitigating cognitive decline.

Regarding blood biomarkers, four biomarkers, NfL, p-tau181, t-tau, and ST2, were identified as significant predictors of cognitive decline in early PD. Our findings for NfL align closely with previous research, reinforcing its role as a valuable prognostic biomarker. For instance, one study demonstrated that elevated serum NfL levels are positively associated with an increased risk of early PD-related symptoms, suggesting that serum NfL could serve as a promising biomarker for early PD [60]. Additionally, another study discovered, through a study using PALS data, that higher plasma NfL levels were linked to a frontal pattern of neurodegeneration, which also correlated with cognitive performance in early PD [61]. This supports the potential future role of plasma NfL as an accessible biomarker for neurodegeneration and cognitive dysfunction in PD. Ng et al. (2020) further highlighted that higher plasma NfL levels were associated with worse cognition and motor function in the postural instability gait disorder (PIGD) subtype of PD, predicting motor and cognitive decline over two years [62]. Similarly, Aamodt et al. (2021) reported that PD participants with high plasma NfL levels were significantly more likely to develop incident cognitive impairment (HR = 5.34,  $p$ -value = 0.005). Although their ROC analysis demonstrated only modest performance for plasma NfL alone in predicting the conversion from normal cognition to MCI or dementia, they noted that incorporating plasma NfL into a multi-marker panel could enhance predictive accuracy [37]. In line with these findings, Batzu et al. (2022) reported that higher plasma NfL levels in PD patients were associated with lower Mini-Mental State Examination (MMSE) scores at baseline, even after adjusting for age, gender, and education [63].

To our knowledge, there have been no extensive studies specifically exploring the roles of blood biomarkers including t-tau, p-tau181, and ST2 in predicting cognitive decline in early PD. In this context, Batzu et al. (2022) conducted a cross-sectional study that found significantly higher plasma p-tau181 concentrations in PD subjects compared to healthy controls at baseline [63]. However, their follow-up over two years did not reveal a significant association between plasma p-tau181 levels and either baseline or longitudinal cognitive performance. Another study highlighted the potential of elevated  $\alpha$ -synuclein and t-tau, along with reduced A $\beta$ -40 levels, as biomarkers for the early detection of cognitive impairment in PD patients [34].

Most research in this area has focused on CSF biomarkers. For instance, Almgren et al. (2023) used PPMI data to develop a ML model for predicting cognitive decline in

de novo PD, incorporating CSF biomarkers, clinical test scores, basic demographics, and baseline cognition [41]. Their findings showed that higher levels of CSF beta-amyloid were significantly associated with less cognitive decline, while higher baseline MoCA scores, elevated CSF t-tau, anxiety, and autonomic dysfunction were linked to greater cognitive decline. Similarly, Tao et al. (2022) investigated the associations between non-motor symptoms and CSF biomarkers in early PD using PPMI data [64]. They found that PD patients with cognitive impairment had significantly lower levels of CSF  $\alpha$ -synuclein,  $A\beta_{1-42}$ , and t-tau compared to PD patients without cognitive impairment. Additionally, Terrelonge et al. (2016) explored the role of CSF biomarkers in predicting cognitive impairment in early PD, revealing that lower baseline levels of CSF  $A\beta_{1-42}$  were significantly associated with a higher risk of cognitive impairment over a two-year period, while no significant associations were found for t-tau or p-tau181 [65]. These findings emphasize the importance of CSF biomarkers as early indicators of cognitive decline risk in PD, underscoring their potential clinical utility for early diagnosis and targeted intervention in PD-related cognitive impairment.

In terms of ST2, a study measuring the plasma soluble decoy receptor form of ST2 (sST2) levels in controls and patients with Alzheimer's disease (AD), frontotemporal dementia (FTD), and PD found that sST2 levels were elevated across all disease groups compared to controls, with the highest levels observed in FTD, followed by AD and PD [66,67]. However, to our knowledge, no studies have specifically investigated plasma ST2 levels in the context of early-stage PD. This highlights the novelty of our study in exploring the association of ST2 with cognitive decline in early PD.

In other words, this study is among the first to explore the association of plasma biomarkers, including t-tau, p-tau181, and ST2, with cognitive decline in early PD. This novel approach provides new insights into how these plasma biomarkers might predict cognitive deterioration in early-stage PD.

The performance of different ML methods in predicting cognitive decline in early PD was evaluated, with RF and NN consistently showing superior results compared to AutoScore and KNN. Model 1, which included all available variables, demonstrated the highest performance, while Model 2, focusing on the top ten variables, provided a more practical approach with notable performance in predicting cognitive decline in early PD.

Previous studies have leveraged ML algorithms to enhance the prediction of cognitive decline and other outcomes in PD. For instance, Zhang et al. (2023) used demographic variables, hospital admission data, and clinical assessments, while grouping predictors based on their cost and accessibility, to build models predicting PD risk. Penalized logistic regression and XGBoost emerged as the most accurate algorithms, with penalized logistic regression achieving an AUC of 0.94 [3]. Deng et al. (2023) conducted a cross-sectional study on PALS data, identifying eight key variables associated with MCI in early PD using ShapleyVIC-assisted and backward selection methods [42]. Their final model included fewer years of education, a shorter history of hypertension, higher MDS-UPDRS motor scores, elevated levels of TG and ApoA1, and noncarrier status of the SNCA rs6826785 genetic marker. These findings align with the present study, which also identified fewer years of education and a history of hypertension as significant predictors of cognitive decline. The combined insights from these studies underscore the importance of a multifaceted approach in using ML to predict cognitive outcomes as a longitudinal outcome in early PD, integrating demographic, clinical, biochemical, and genetic factors for more accurate and practical predictive models.

#### *Limitations and Future Avenues*

This study employed 10-fold cross-validation to assess model performance; however, several limitations should be noted. The high AUC observed could be influenced by the small sample size, which may lead to overly optimistic performance estimates, even with 10-fold cross-validation. Although one-tenth of the data are set aside for validation in each iteration, small datasets can result in higher variance in performance metrics, and the

results may not generalize well to larger, independent datasets. This raises concerns about producing biased performance estimates. Therefore, a more conservative approach, such as nested cross-validation, may provide more reliable performance estimates. Additionally, the calibration results for Model 2 showed that while RF and NN tend to overestimate predicted probabilities, KNN underestimates them. In contrast, logistic regression and AutoScore exhibit better calibration performance. However, these results should be interpreted with caution due to the small sample size and the use of the same dataset for both training and evaluation. These limitations highlight the necessity for further validation with independent, external datasets to ensure the robustness and generalizability of the findings.

## 5. Conclusions

This study demonstrates the potential of ML methods in accurately predicting cognitive decline in individuals with early-stage PD. By integrating baseline demographic, clinical, and blood biomarker data, these models offer valuable insights for the early identification of patients at a high risk of cognitive deterioration, providing opportunities for timely interventions and improved patient outcomes. While a comprehensive model incorporating all available variables achieved the highest predictive performance, the practicality of utilizing certain biomarkers in clinical settings may be limited due to their cost and accessibility. A more streamlined model focusing on key biomarkers, however, maintained strong predictive capabilities, offering a more practical and feasible approach for real-world clinical implementation. These findings underscore the theoretical implications of integrating data-driven approaches in neurodegenerative disease management and highlight opportunities for translating ML models into clinical practice. Future research should explore strategies to enhance model interpretability, validate findings across diverse populations, and assess long-term impacts on patient care. Additionally, further methodological developments are needed to optimize biomarker selection and address practical implementation challenges, paving the way for broader adoption of predictive analytics in personalized medicine.

**Author Contributions:** Conceptualization, R.M., A.S.L.N., E.-K.T., L.C.S.T. and S.E.S.; Methodology, R.M., E.W.S., W.G. and S.E.S.; Software, R.M. and S.E.S.; Validation, R.M. and S.E.S.; Formal Analysis, R.M. and S.E.S.; Investigation, R.M. and S.E.S.; Resources, S.N., J.Y.T., A.S.L.N., X.D., X.C., D.L.H., S.N., Z.X., K.-Y.T., W.-L.A., E.-K.T., L.C.S.T. and S.E.S.; Data Curation, S.N., J.Y.T., A.S.L.N., X.D., X.C. and S.E.S.; Writing—Original Draft Preparation, R.M. and S.E.S.; Writing—Review and Editing, R.M., S.E.S., J.Y.T., A.S.L.N., X.D., X.C., D.L.H., S.N., Z.X., K.-Y.T., W.-L.A., E.-K.T., L.C.S.T., E.W.S., W.G. and S.E.S.; Visualization, R.M. and S.E.S.; Supervision, A.S.L.N., E.-K.T., L.C.S.T., W.G. and S.E.S.; Project Administration, R.M., S.Y.E.N. and J.Y.T.; Funding Acquisition, E.-K.T., L.C.S.T. and S.E.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Singapore Ministry of Health's National Medical Research Council (MOH-OFLCG18May-0002, MOH-CSAINV21-0005, CNIG22jul-0004).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of SingHealth (IRB reference number: CIRB 2019-2433) and National University of Singapore (IRB reference number: NUS-IRB-2022-899).

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

**Data Availability Statement:** The study data will be made available upon reasonable request to the corresponding author. The data are not publicly available due to privacy and ethical concerns.

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Abbreviations

PD	Parkinson's disease
CD	Cognitive decline
MoCA	Montreal Cognitive Assessment
AUC	Area under the curve
MCI	Mild cognitive impairment
PDD	Parkinson's disease dementia
ML	Machine learning
RF	Random forest
CSF	Cerebrospinal fluid
PPMI	Parkinson's Progression Markers Initiative
ApoA1	Apolipoprotein A1
TG	Triglycerides
HY	Hoehn and Yahr
MDS-UPDRS	Movement Disorder Society-Unified Parkinson's disease Rating Scale
BMI	Body mass index
SBP	Systolic blood pressure
DBP	Diastolic blood pressure
ST2	Suppression of tumorigenicity 2
NfL	Neurofilament light chain
t-tau	Total tau
p-tau 181	Phosphorylated tau at threonine 181
APOE	Apolipoprotein E
REP1	Alpha-synuclein gene promoter
OR	Odds ratio
SD	Standard deviation
CI	Confidence interval
KNN	K-nearest neighbors
NN	Neural network
ROC	Receiver operating characteristic
AD	Alzheimer's disease
FTD	Frontotemporal dementia

## References

1. Aarsland, D.; Creese, B.; Politis, M.; Chaudhuri, K.R.; Ffytche, D.H.; Weintraub, D.; Ballard, C. Cognitive decline in Parkinson disease. *Nat. Rev. Neurol.* **2017**, *13*, 217–231. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Ciucci, M.R.; Grant, L.M.; Rajamanickam, E.S.P.; Hilby, B.L.; Blue, K.V.; Jones, C.A.; Kelm-Nelson, C.A. *Early Identification and Treatment of Communication and Swallowing Deficits in Parkinson Disease*; Semin. Speech Lang; Thieme Medical Publishers: New York, NY, USA, 2013; pp. 185–202.
3. Zhang, J.; Zhou, W.; Yu, H.; Wang, T.; Wang, X.; Liu, L.; Wen, Y. Prediction of Parkinson's Disease Using Machine Learning Methods. *Biomolecules* **2023**, *13*, 1761. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Simuni, T.; Sethi, K. Nonmotor manifestations of Parkinson's disease. *Ann. Neurol. Off. J. Am. Neurol. Assoc. Child Neurol. Soc.* **2008**, *64*, S65–S80. [\[CrossRef\]](#)
5. Williams-Gray, C.H.; Mason, S.L.; Evans, J.R.; Foltynie, T.; Brayne, C.; Robbins, T.W.; Barker, R.A. The CamPaIGN study of Parkinson's disease: 10-year outlook in an incident population-based cohort. *J. Neurol. Neurosurg. Psychiatry* **2013**, *84*, 1258–1264. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Riedel, O.; Klotsche, J.; Spottke, A.; Deuschl, G.; Förstl, H.; Henn, F.; Heuser, I.; Oertel, W.; Reichmann, H.; Riederer, P. Frequency of dementia, depression, and other neuropsychiatric symptoms in 1,449 outpatients with Parkinson's disease. *J. Neurol.* **2010**, *257*, 1073–1082. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Deng, X.; Saffari, S.E.; Xiao, B.; Ng, S.Y.E.; Chia, N.; Choi, X.; Heng, D.L.; Ng, E.; Xu, Z.; Tay, K.-Y. Disease Progression of Data-Driven Subtypes of Parkinson's Disease: 5-Year Longitudinal Study from the Early Parkinson's Disease Longitudinal Singapore (PALS) Cohort. *JPD* **2024**, *14*, 1051–1059. [\[CrossRef\]](#)
8. Deng, X.; Saffari, S.E.; Xiao, B.; Ng, S.Y.E.; Chia, N.; Choi, X.; Heng, D.L.; Xu, Z.; Tay, K.-Y.; Au, W.-L. Disease progression in Parkinson's disease patients with mild cognitive impairment: 5-year longitudinal study from the early Parkinson's disease longitudinal Singapore (PALS) cohort. *Aging* **2024**, *16*, 11491. [\[CrossRef\]](#)
9. Battaglia, S.; Avenanti, A.; Vecsei, L.; Tanaka, M. Neurodegeneration in cognitive impairment and mood disorders for experimental, clinical and translational neuropsychiatry. *Biomedicines* **2024**, *12*, 574. [\[CrossRef\]](#)



10. Fang, C.; Lv, L.; Mao, S.; Dong, H.; Liu, B. Cognition deficits in Parkinson's disease: Mechanisms and treatment. *Parkinsons Dis.* **2020**, *2020*, 2076942. [\[CrossRef\]](#)
11. Kandiah, N.; Mak, E.; Ng, A.; Huang, S.; Au, W.L.; Sitoh, Y.Y.; Tan, L.C.S. Cerebral white matter hyperintensity in Parkinson's disease: A major risk factor for mild cognitive impairment. *Park. Relat. Disord.* **2013**, *19*, 680–683. [\[CrossRef\]](#)
12. Pigott, K.; Rick, J.; Xie, S.X.; Hurtig, H.; Chen-Plotkin, A.; Duda, J.E.; Morley, J.F.; Chahine, L.M.; Dahodwala, N.; Akhtar, R.S. Longitudinal study of normal cognition in Parkinson disease. *Neurology* **2015**, *85*, 1276–1282. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Hely, M.A.; Reid, W.G.; Adena, M.A.; Halliday, G.M.; Morris, J.G. The Sydney multicenter study of Parkinson's disease: The inevitability of dementia at 20 years. *Mov. Disord.* **2008**, *23*, 837–844. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Lawson, R.A.; Yarnall, A.J.; Duncan, G.W.; Breen, D.P.; Khoo, T.K.; Williams-Gray, C.H.; Barker, R.A.; Burn, D.J. Stability of mild cognitive impairment in newly diagnosed Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **2017**, *88*, 648–652. [\[CrossRef\]](#)
15. Poletti, M.; Emre, M.; Bonuccelli, U. Mild cognitive impairment and cognitive reserve in Parkinson's disease. *Park. Relat. Disord.* **2011**, *17*, 579–586. [\[CrossRef\]](#)
16. Chua, C.Y.; Koh, M.R.E.; Chia, N.S.-Y.; Ng, S.Y.-E.; Saffari, S.E.; Wen, M.-C.; Chen, R.Y.-Y.; Choi, X.; Heng, D.L.; Neo, S.X. Subjective cognitive complaints in early Parkinson's disease patients with normal cognition are associated with affective symptoms. *Park. Relat. Disord.* **2021**, *82*, 24–28. [\[CrossRef\]](#)
17. Pressley, J.C.; Louis, E.D.; Tang, M.X.; Cote, L.; Cohen, P.D.; Glied, S.; Mayeux, R. The impact of comorbid disease and injuries on resource use and expenditures in parkinsonism. *Neurology* **2003**, *60*, 87–93. [\[CrossRef\]](#)
18. Greenland, J.C.; Williams-Gray, C.H.; Barker, R.A. The clinical heterogeneity of Parkinson's disease and its therapeutic implications. *Eur. J. Neurosci.* **2019**, *49*, 328–338. [\[CrossRef\]](#)
19. Shen, J.; Amari, N.; Zack, R.; Skrinak, R.T.; Unger, T.L.; Posavi, M.; Tropea, T.F.; Xie, S.X.; Van Deerlin, V.M.; Dewey, R.B., Jr.; et al. Plasma MIA, CRP, and albumin predict cognitive decline in Parkinson's disease. *Ann. Neurol.* **2022**, *92*, 255–269. [\[CrossRef\]](#)
20. Martin, F.L.; Williamson, S.J.M.; Paleologou, K.E.; Allsop, D.; El-Agnaf, O.M.A.  $\alpha$ -Synuclein and the pathogenesis of Parkinson's disease. *Protein Pept. Lett.* **2004**, *11*, 229–237. [\[CrossRef\]](#)
21. Alves, G.; Brønne, K.; Aarsland, D.; Blennow, K.; Zetterberg, H.; Ballard, C.; Kurz, M.W.; Andreasson, U.; Tysnes, O.-B.; Larsen, J.P. CSF amyloid- $\beta$  and tau proteins, and cognitive performance, in early and untreated Parkinson's Disease: The Norwegian ParkWest study. *J. Neurol. Neurosurg. Psychiatry* **2010**, *81*, 1080–1086. [\[CrossRef\]](#)
22. Deng, X.; Saffari, S.E.; Liu, N.; Xiao, B.; Allen, J.C.; Ng, S.Y.E.; Chia, N.; Tan, Y.J.; Choi, X.; Heng, D.L. Biomarker characterization of clinical subtypes of Parkinson Disease. *NPJ Park. Dis.* **2022**, *8*, 109. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Deng, X.; Saffari, S.E.; Ng, S.Y.E.; Chia, N.; Tan, J.Y.; Choi, X.; Heng, D.L.; Xu, Z.; Tay, K.-Y.; Au, W.-L. Blood lipid biomarkers in early Parkinson's disease and Parkinson's disease with mild cognitive impairment. *J. Park. Dis.* **2022**, *12*, 1937–1943. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Hoogland, J.; De Bie, R.M.A.; Williams-Gray, C.H.; Muslimović, D.; Schmand, B.; Post, B. Catechol-O-methyltransferase val158met and cognitive function in Parkinson's disease. *Mov. Disord.* **2010**, *25*, 2550–2554. [\[CrossRef\]](#)
25. Kim, R.; Kim, H.J.; Shin, J.H.; Lee, C.Y.; Jeon, S.H.; Jeon, B. Serum inflammatory markers and progression of nonmotor symptoms in early Parkinson's disease. *Mov. Disord.* **2022**, *37*, 1535–1541. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Michael, J.; Fox Foundation. FDA Issues Letter of Support Encouraging Use of Synuclein-Based Biomarker (Asyn-SAA) in Clinical Trials 2024. Available online: <https://web.archive.org/web/20240930072011/https://www.michaeljfox.org/publication/fda-issues-letter-support-encouraging-use-synuclein-based-biomarker-asyn-saa-clinical> (accessed on 30 September 2024).
27. Hu, X.; Yang, Y.; Gong, D. Changes of cerebrospinal fluid A $\beta$  42, t-tau, and p-tau in Parkinson's disease patients with cognitive impairment relative to those with normal cognition: A meta-analysis. *Neurol. Sci.* **2017**, *38*, 1953–1961. [\[CrossRef\]](#)
28. Mollenhauer, B.; Bibl, M.; Wiltfang, J.; Steinacker, P.; Ciesielczyk, B.; Neubert, K.; Trenkwalder, C.; Otto, M. *Total tau Protein, Phosphorylated tau (181p) Protein,  $\beta$ -Amyloid1–42, and  $\beta$ -Amyloid1–40 in Cerebrospinal Fluid of Patients with Dementia with Lewy Bodies*; De Gruyter: Berlin, Germany, 2006.
29. Siderowf, A.; Xie, S.X.; Hurtig, H.; Weintraub, D.; Duda, J.; Chen-Plotkin, A.; Shaw, L.M.; Van Deerlin, V.; Trojanowski, J.Q.; Clark, C. CSF amyloid  $\beta$  1–42 predicts cognitive decline in Parkinson disease. *Neurology* **2010**, *75*, 1055–1061. [\[CrossRef\]](#)
30. Lerche, S.; Wurster, I.; Rösen, B.; Zimmermann, M.; Machetanz, G.; Wiethoff, S.; Dehnert, M.; Rietschel, L.; Riebenbauer, B.; Deuschle, C. CSF NFL in a longitudinally assessed PD cohort: Age effects and cognitive trajectories. *Mov. Disord.* **2020**, *35*, 1138–1144. [\[CrossRef\]](#)
31. Parnetti, L.; Gaetani, L.; Eusebi, P.; Paciotti, S.; Hansson, O.; El-Agnaf, O.; Mollenhauer, B.; Blennow, K.; Calabresi, P. CSF and blood biomarkers for Parkinson's disease. *Lancet Neurol.* **2019**, *18*, 573–586. [\[CrossRef\]](#)
32. Youssef, P.; Hughes, L.; Kim, W.S.; Halliday, G.M.; Lewis, S.J.; Cooper, A.; Dzamko, N. Evaluation of plasma levels of NFL, GFAP, UCHL1 and tau as Parkinson's disease biomarkers using multiplexed single molecule counting. *Sci. Rep.* **2023**, *13*, 5217. [\[CrossRef\]](#)
33. Kim, R.; Park, S.; Yoo, D.; Jun, J.-S.; Jeon, B. Association of physical activity and APOE genotype with longitudinal cognitive change in early Parkinson disease. *Neurology* **2021**, *96*, e2429–e2437. [\[CrossRef\]](#)
34. Chen, N.-C.; Chen, H.-L.; Li, S.-H.; Chang, Y.-H.; Chen, M.-H.; Tsai, N.-W.; Yu, C.-C.; Yang, S.-Y.; Lu, C.-H.; Lin, W.-C. Plasma levels of  $\alpha$ -synuclein, A $\beta$ -40 and T-tau as biomarkers to predict cognitive impairment in Parkinson's disease. *Front. Aging Neurosci.* **2020**, *12*, 112. [\[CrossRef\]](#) [\[PubMed\]](#)

35. Pellecchia, M.T.; Savastano, R.; Moccia, M.; Picillo, M.; Siano, P.; Erro, R.; Valletlunga, A.; Amboni, M.; Vitale, C.; Santangelo, G. Lower serum uric acid is associated with mild cognitive impairment in early Parkinson's disease: A 4-year follow-up study. *J. Neural Transm.* **2016**, *123*, 1399–1402. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Sekiya, H.; Tsuji, A.; Hashimoto, Y.; Takata, M.; Koga, S.; Nishida, K.; Futamura, N.; Kawamoto, M.; Kohara, N.; Dickson, D.W. Discrepancy between distribution of alpha-synuclein oligomers and Lewy-related pathology in Parkinson's disease. *Acta Neuropathol. Commun* **2022**, *10*, 133. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Aamodt, W.W.; Waligorska, T.; Shen, J.; Tropea, T.F.; Siderowf, A.; Weintraub, D.; Grossman, M.; Irwin, D.; Wolk, D.A.; Xie, S.X. Neurofilament light chain as a biomarker for cognitive decline in Parkinson disease. *Mov. Disord.* **2021**, *36*, 2945–2950. [\[CrossRef\]](#)
38. Chen-Plotkin, A.S.; Hu, W.T.; Siderowf, A.; Weintraub, D.; Goldmann Gross, R.; Hurtig, H.I.; Xie, S.X.; Arnold, S.E.; Grossman, M.; Clark, C.M. Plasma epidermal growth factor levels predict cognitive decline in Parkinson disease. *Ann. Neurol.* **2011**, *69*, 655–663. [\[CrossRef\]](#)
39. Ma, L.-Z.; Zhang, C.; Wang, H.; Ma, Y.-H.; Shen, X.-N.; Wang, J.; Tan, L.; Dong, Q.; Yu, J.-T. Serum neurofilament dynamics predicts cognitive progression in de novo Parkinson's disease. *J. Park. Dis.* **2021**, *11*, 1117–1127. [\[CrossRef\]](#)
40. Schrag, A.; Siddiqui, U.F.; Anastasiou, Z.; Weintraub, D.; Schott, J.M. Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson's disease: A cohort study. *Lancet Neurol.* **2017**, *16*, 66–75. [\[CrossRef\]](#)
41. Almgren, H.; Camacho, M.; Hanganu, A.; Kibreb, M.; Camicioli, R.; Ismail, Z.; Forkert, N.D.; Monchi, O. Machine learning-based prediction of longitudinal cognitive decline in early Parkinson's disease using multimodal features. *Sci. Rep.* **2023**, *13*, 13193. [\[CrossRef\]](#)
42. Deng, X.; Ning, Y.; Saffari, S.E.; Xiao, B.; Niu, C.; Ng, S.Y.E.; Chia, N.; Choi, X.; Heng, D.L.; Tan, Y.J. Identifying clinical features and blood biomarkers associated with mild cognitive impairment in Parkinson disease using machine learning. *Eur. J. Neurol.* **2023**, *30*, 1658–1666. [\[CrossRef\]](#)
43. Ng, S.Y.-E.; Chia, N.S.-Y.; Abbas, M.M.; Saffari, E.S.; Choi, X.; Heng, D.L.; Xu, Z.; Tay, K.-Y.; Au, W.-L.; Tan, E.-K. Physical activity improves anxiety and apathy in early Parkinson's disease: A longitudinal follow-up study. *Front. Neurol.* **2021**, *11*, 625897. [\[CrossRef\]](#)
44. Yong, A.C.W.; Tan, Y.J.; Zhao, Y.; Lu, Z.; Ng, E.Y.L.; Ng, S.Y.E.; Chia, N.S.Y.; Choi, X.; Heng, D.; Neo, S. SNCA Rep1 microsatellite length influences non-motor symptoms in early Parkinson's disease. *Aging* **2020**, *12*, 20880. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Stekhoven, D.J. Using the missForest package. *R Package* **2011**, 1–11. [\[CrossRef\]](#)
46. Youden, W.J. Index for rating diagnostic tests. *Cancer* **1950**, *3*, 32–35. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Niculescu-Mizil, A.; Caruana, R. Predicting good probabilities with supervised learning. In Proceedings of the Twenty-Second International Conference (ICML 2005), Bonn, Germany, 7–11 August 2005; pp. 625–632.
48. Saffari, S.E.; Ning, Y.; Xie, F.; Chakraborty, B.; Volovici, V.; Vaughan, R.; Ong, M.E.H.; Liu, N. AutoScore-Ordinal: An interpretable machine learning framework for generating scoring models for ordinal outcomes. *BMC Med. Res. Methodol.* **2022**, *22*, 286. [\[CrossRef\]](#)
49. Roheger, M.; Kalbe, E.; Liepelt-Scarfone, I. Progression of cognitive decline in Parkinson's disease. *J. Park. Dis.* **2018**, *8*, 183–193. [\[CrossRef\]](#)
50. Forbes, E.; Tropea, T.F.; Mantri, S.; Xie, S.X.; Morley, J.F. Modifiable comorbidities associated with cognitive decline in Parkinson's disease. *Mov. Disord. Clin. Pract.* **2021**, *8*, 254–263. [\[CrossRef\]](#)
51. Zhang, L.; Gu, L.-Y.; Dai, S.-B.; Zheng, R.; Jin, C.-Y.; Fang, Y.; Yang, W.-Y.; Tian, J.; Yin, X.-Z.; Zhao, G.-H. Associations of body mass index-metabolic phenotypes with cognitive decline in Parkinson's disease. *Eur. Neurol.* **2022**, *85*, 24–30. [\[CrossRef\]](#)
52. Yoo, H.S.; Chung, S.J.; Lee, P.H.; Sohn, Y.H.; Kang, S.Y. The influence of body mass index at diagnosis on cognitive decline in Parkinson's disease. *J. Clin. Neurol.* **2019**, *15*, 517–526. [\[CrossRef\]](#)
53. Kim, H.J.; Oh, E.S.; Lee, J.H.; Moon, J.S.; Oh, J.E.; Shin, J.W.; Lee, K.J.; Baek, I.C.; Jeong, S.-H.; Song, H.-J. Relationship between changes of body mass index (BMI) and cognitive decline in Parkinson's disease (PD). *Arch. Gerontol. Geriatr.* **2012**, *55*, 70–72. [\[CrossRef\]](#)
54. Kwon, K.-Y.; Pyo, S.J.; Lee, H.M.; Seo, W.-K.; Koh, S.-B. Cognition and visit-to-visit variability of blood pressure and heart rate in de novo patients with Parkinson's disease. *J. Mov. Disord* **2016**, *9*, 144. [\[CrossRef\]](#)
55. Xiao, Y.; Yang, T.; Zhang, L.; Wei, Q.; Ou, R.; Hou, Y.; Liu, K.; Lin, J.; Jiang, Q.; Shang, H. Association between the blood pressure variability and cognitive decline in Parkinson's disease. *Brain Behav.* **2023**, *13*, e3319. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Doiron, M.; Langlois, M.; Dupré, N.; Simard, M. The influence of vascular risk factors on cognitive function in early Parkinson's disease. *Int. J. Geriatr. Psychiatry* **2018**, *33*, 288–297. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Siciliano, M.; De Micco, R.; Trojano, L.; De Stefano, M.; Baiano, C.; Passaniti, C.; De Mase, A.; Russo, A.; Tedeschi, G.; Tessitore, A. Cognitive impairment is associated with Hoehn and Yahr stages in early, de novo Parkinson disease patients. *Park. Relat. Disord.* **2017**, *41*, 86–91. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Pagano, G.; Boess, F.G.; Taylor, K.I.; Ricci, B.; Mollenhauer, B.; Poewe, W.; Boulay, A.; Anzures-Cabrera, J.; Vogt, A.; Marchesi, M. 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. [\[CrossRef\]](#)
59. Jackson, H.; Anzures-Cabrera, J.; Taylor, K.I.; Pagano, G.; Investigators, P.; Prasinezumab Study, G. Hoehn and Yahr stage and striatal Dat-SPECT uptake are predictors of Parkinson's disease motor progression. *Front. Neurosci.* **2021**, *15*, 765765. [\[CrossRef\]](#)

60. Wang, X.; Yang, X.; He, W.; Song, X.; Zhang, G.; Niu, P.; Chen, T. The association of serum neurofilament light chains with early symptoms related to Parkinson's disease: A cross-sectional study. *J. Affect. Disord.* **2023**, *343*, 144–152. [[CrossRef](#)]
61. Welton, T.; Tan, Y.J.; Saffari, S.E.; Ng, S.Y.; Chia, N.S.; Yong, A.C.; Choi, X.; Heng, D.L.; Shih, Y.-C.; Hartono, S. Plasma neurofilament light concentration is associated with diffusion-tensor MRI-based measures of neurodegeneration in early Parkinson's disease. *J. Park. Dis.* **2022**, *12*, 2135–2146. [[CrossRef](#)]
62. Ng, A.S.L.; Tan, Y.J.; Yong, A.C.W.; Saffari, S.E.; Lu, Z.; Ng, E.Y.; Ng, S.Y.E.; Chia, N.S.Y.; Choi, X.; Heng, D. Utility of plasma Neurofilament light as a diagnostic and prognostic biomarker of the postural instability gait disorder motor subtype in early Parkinson's disease. *Mol. Neurodegener.* **2020**, *15*, 1–8. [[CrossRef](#)]
63. Batzu, L.; Rota, S.; Hye, A.; Heslegrave, A.; Trivedi, D.; Gibson, L.L.; Farrell, C.; Zinzalias, P.; Rizos, A.; Zetterberg, H. Plasma p-tau181, neurofilament light chain and association with cognition in Parkinson's disease. *npj Park. Dis.* **2022**, *8*, 154. [[CrossRef](#)]
64. Tao, M.; Dou, K.; Xie, Y.; Hou, B.; Xie, A. The associations of cerebrospinal fluid biomarkers with cognition, and rapid eye movement sleep behavior disorder in early Parkinson's disease. *Front. Neurosci.* **2022**, *16*, 1049118. [[CrossRef](#)]
65. Terrelonge, M.; Marder, K.S.; Weintraub, D.; Alcalay, R.N. CSF  $\beta$ -amyloid 1–42 predicts progression to cognitive impairment in newly diagnosed Parkinson disease. *J. Mol. Neurosci.* **2016**, *58*, 88–92. [[CrossRef](#)] [[PubMed](#)]
66. Tan, Y.J.; Saffari, S.E.; Zhao, Y.; Ng, E.Y.; Yong, A.C.; Ng, S.Y.; Chia, N.S.; Choi, X.; Heng, D.; Neo, S. Longitudinal Study of SNCA Rep1 Polymorphism on Executive Function in Early Parkinson's Disease. *J. Park. Dis.* **2022**, *12*, 865–870. [[CrossRef](#)] [[PubMed](#)]
67. Tan, Y.J.; Siow, I.; Saffari, S.E.; Ting, S.K.S.; Li, Z.; Kandiah, N.; Tan, L.C.S.; Tan, E.K.; Ng, A.S.L. Plasma soluble ST2 levels are higher in neurodegenerative disorders and associated with poorer cognition. *J. Alzheimer's Dis.* **2023**, *92*, 573–580. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.