



ORIGINAL ARTICLE

# Accuracy of Machine Learning Using the Montreal Cognitive Assessment for the Diagnosis of Cognitive Impairment in Parkinson's Disease

Junbeom Jeon,<sup>1</sup> Kiyong Kim,<sup>2</sup> Kyeongmin Baek,<sup>1</sup> Seok Jong Chung,<sup>3,4</sup> Jeehee Yoon,<sup>1</sup> Yun Joong Kim<sup>3,4</sup>

<sup>1</sup>Department of Computer Engineering, Hallym University, Chuncheon, Korea

<sup>2</sup>Department of Electronic Engineering, Kyonggi University, Suwon, Korea

<sup>3</sup>Department of Neurology, Yonsei University College of Medicine, Seoul, Korea

<sup>4</sup>Department of Neurology, Yongin Severance Hospital, Yonsei University Health System, Yongin, Korea

## ABSTRACT

**Objective** The Montreal Cognitive Assessment (MoCA) is recommended for assessing general cognition in Parkinson's disease (PD). Several cutoffs of MoCA scores for diagnosing PD with cognitive impairment (PD-CI) have been proposed, with varying sensitivity and specificity. This study investigated the utility of machine learning algorithms using MoCA cognitive domain scores for improving diagnostic performance for PD-CI.

**Methods** In total, 2,069 MoCA results were obtained from 397 patients with PD enrolled in the Parkinson's Progression Markers Initiative database with a diagnosis of cognitive status based on comprehensive neuropsychological assessments. Using the same number of MoCA results randomly sampled from patients with PD with normal cognition or PD-CI, discriminant validity was compared between machine learning (logistic regression, support vector machine, or random forest) with domain scores and a cutoff method.

**Results** Based on cognitive status classification using a dataset that permitted sampling of MoCA results from the same individual ( $n = 221$  per group), no difference was observed in accuracy between the cutoff value method ( $0.74 \pm 0.03$ ) and machine learning ( $0.78 \pm 0.03$ ). Using a more stringent dataset that excluded MoCA results ( $n = 101$  per group) from the same patients, the accuracy of the cutoff method ( $0.66 \pm 0.05$ ), but not that of machine learning ( $0.74 \pm 0.07$ ), was significantly reduced. Inclusion of cognitive complaints as an additional variable improved the accuracy of classification using the machine learning method ( $0.87$ – $0.89$ ).

**Conclusion** Machine learning analysis using MoCA domain scores is a valid method for screening cognitive impairment in PD.

**Keywords** Depression; Machine learning; Mild cognitive impairment; Montreal Cognitive Assessment; Parkinson's disease; Regression analysis.

Cognitive impairment is common in Parkinson's disease (PD). Mild cognitive impairment (MCI) affects 20%–50% of patients with PD.<sup>1</sup> More than 25% of cases of newly diagnosed PD with MCI (PD-MCI) progress to PD dementia (PDD) within 3 years.<sup>2</sup> The cumulative prevalence of dementia in PD over an 8- to 12-

year follow-up has been reported to be 60%–83%.<sup>3-5</sup> Crucially, the occurrence of dementia in PD has a major impact on functional independence, nursing home placement, mortality, psychiatric morbidities, and caregiver burden. The Movement Disorders Society (MDS) Task Force has proposed diagnostic criteria

Received: January 22, 2022 Revised: March 26, 2022 Accepted: April 13, 2022

Corresponding author: Yun Joong Kim, MD, PhD

Department of Neurology, Yongin Severance Hospital, 363 Dongbaekjukjeon-daero, Giheung-gu, Yongin 16995, Korea / Tel: +82-31-5189-8140 / Fax: +82-31-5189-8565 / E-mail: yunjkim@yuhs.ac

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

guidelines for PD-MCI and PDD and recommends cognitive assessments using abbreviated (level I) or comprehensive assessments (level II) comprising neuropsychological tests with at least two tests in each of the five cognitive domains.<sup>6</sup> The Montreal Cognitive Assessment (MoCA) is the most widely recommended level I test. The MoCA is more sensitive than the Mini-Mental State Examination owing to the inclusion of tools for frontal executive function and no ceiling effect<sup>7,8</sup> and is recommended for global cognitive assessment; however, it also permits the evaluation of specific cognitive domains.<sup>1,9,10</sup> The MoCA has been employed to differentiate PD with cognitive impairment (PD-CI) from PD with normal cognition (PD-NC), and several different cutoffs of MoCA total scores have been proposed, with varying sensitivity and specificity.<sup>7-9,11-13</sup> Using a cutoff score in the MoCA has the following limitations:<sup>14</sup> MoCA scores are influenced by not only age but also education, occupational experience, or translated versions of the MoCA. Therefore, application of the MoCA in a less educated population is not accurate. Cognitive change in PD is affected by the duration of PD, the presence of depression and the severity of parkinsonian motor symptoms, which are not incorporated in the determination of cognitive diagnosis using a cutoff score. In PD-MCI or PDD, executive dysfunction or visuospatial dysfunction is more frequently reported.<sup>15-18</sup> Although the evaluation of individual cognitive domains is possible using the MoCA, no study has used multiple cognitive domains of the MoCA in screening PD-CI.

The growing use of big data has catalyzed the adoption of machine learning and deep learning technologies in clinical studies. Machine learning is a statistical methodology for determining patterns based on a training dataset using a specific algorithm to make learning-based predictions for other datasets.<sup>19</sup> Although the cutoff method based on MoCA total scores has been widely used in clinical domains, machine learning methods based on individual domain scores of the MoCA may exhibit superior performance compared to the previous cutoff method. However, no study to date has systematically evaluated the comparative effectiveness of machine learning algorithms for the diagnosis of cognitive impairment in PD using the MoCA. In this study, we investigated the extent to which machine learning algorithms based on MoCA cognitive domain scores could be used to improve the accuracy of PD-CI diagnosis using comprehensive neuropsychological tests compared to a conventional cutoff method.

## MATERIALS & METHODS

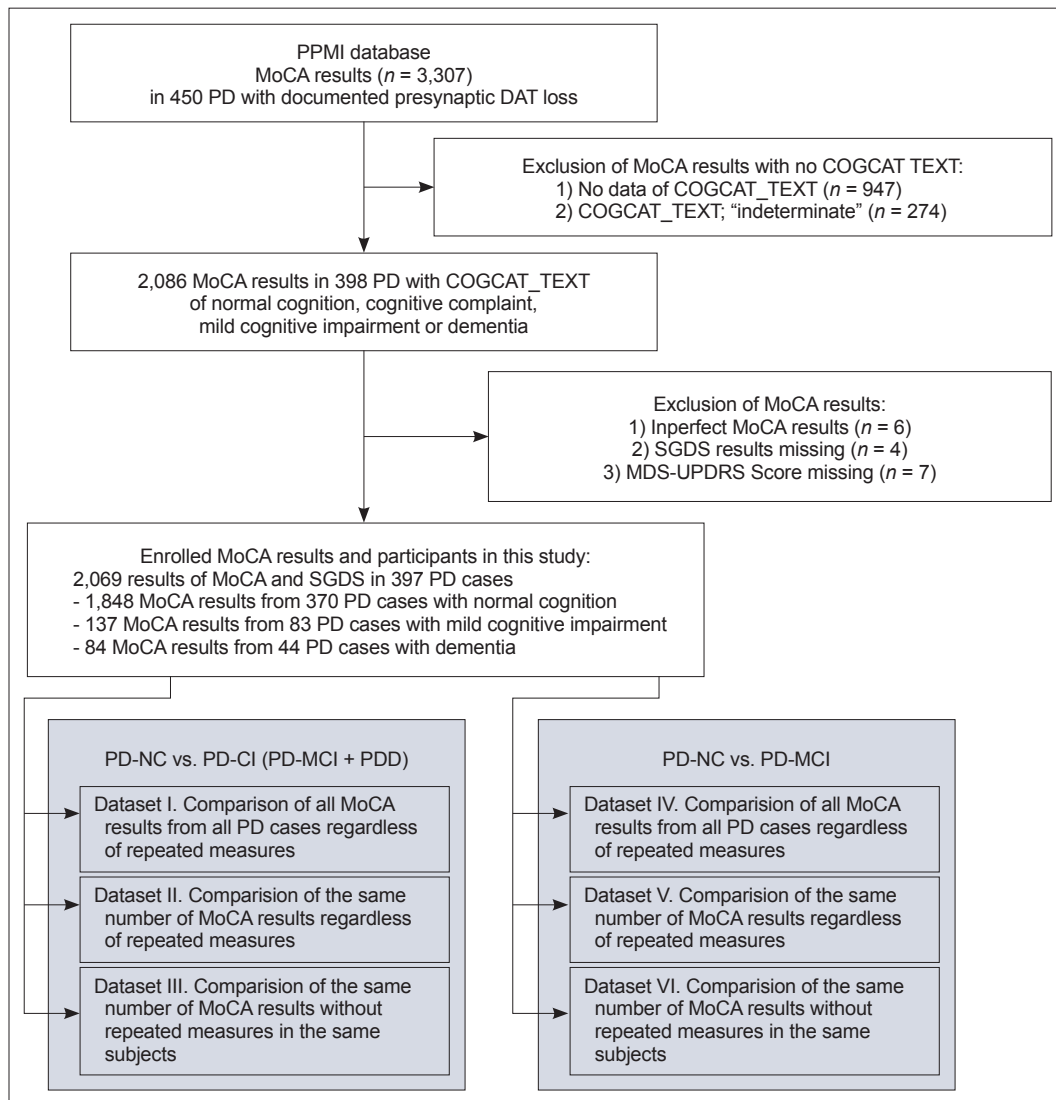
### Participants

This study used the database from the Parkinson's Progression Markers Initiative (PPMI) cohort, which is an observational, in-

ternational study cohort designed to identify clinical, imaging, genetic, and biospecimen markers for PD progression to accelerate disease-modifying therapeutic trials. Clinical and neuroimaging data of patients with PD in the PPMI database were downloaded in April 2021. A flowchart of participant enrollment and performance of the MoCA is presented in Figure 1. A total of 3,307 MoCA results were assessed longitudinally in 450 patients with PD in whom presynaptic dopamine loss was documented by dopamine transporter imaging using <sup>123</sup>I-Ioflupane single-photon emission computed tomography. MoCA results from patients with PD with cognitive categorization of normal cognition, cognitive complaints, MCI, or dementia based on comprehensive neuropsychological tests (level II) were included. In total, 947 MoCA results from patients with no cognitive categorization and 274 MoCA results from patients with "indeterminate" cognitive categorization were excluded. We further excluded 17 MoCA results from patients with incomplete MoCA, short version of the Geriatric Depression Scale (SGDS), or Unified Parkinson's Disease Rating Scale by the MDS data. Finally, a total of 2,069 MoCA and SGDS results from 397 patients with PD were included, comprising 221 MoCA results from 101 patients with PD-CI and 1,848 MoCA results from 370 patients with PD-NC. Cognitive categorization of the 74 patients with PD varied from normal to cognitive impairment (either MCI or dementia) during the follow-up. To avoid potential bias due to skewness of sample size or repeated measures, we generated six datasets by random sampling as follows (Figure 1): Datasets I, II, and III were used to compare PD-NC and PD-CI (MCI or dementia). Datasets IV, V, and VI were used to compare PD-NC and PD-MCI. In datasets I and IV, all MoCA results from cases and controls were included, which has a flaw of skewness (unbalanced data). To overcome the skewness, datasets II and V were generated by randomly sampling the same number of MoCA results. However, the flaws in these datasets are that repeated measures in the same subjects are included. Last, to avoid the potential bias of repeated measures, we randomly sampled the MoCA from different subjects, not allowing repeated measurements. This study was approved by the Institutional Review Boards of each participating PPMI site. Written informed consent was obtained from all participants. The study is registered at <http://www.clinicaltrials.gov> (identifier: NCT01141023).

### Statistical analysis

The effectiveness of machine learning classification methods was compared using extensive analyses. Three methods were used to build prediction models: support vector machine (SVM), random forest (RF), and logistic regression (LR). SVM is a widely used machine learning method with excellent prediction accuracy.<sup>20</sup> RF is an ensemble model of decision trees that exhib-



**Figure 1.** Flowchart of participants and the enrollment process. PPMI, Parkinson's Progression Markers Initiative; MoCA, Montreal Cognitive Assessment; PD, Parkinson's disease; DAT, dopamine transporter; COGCAT\_TEXT, cognitive categorization; SGDS, short version of the Geriatric Depression Scale; MDS-UPDRS, Unified Parkinson's Disease Rating Scale by the Movement Disorders Society; PD-NC, PD with normal cognition; PD-CI, PD with cognitive impairment; PD-MCI, PD with mild cognitive impairment; PDD, PD dementia.

its robustness to noise and irrelevant factors and requires almost no fine-tuning of parameters to produce good predictions.<sup>21,22</sup> LR is a simple model that involves few parameters and is easily interpretable. LR was used as the basis to determine the relative importance of individual screening factors for the identification of cognitive status. Model development was initiated by stepwise LR using the training dataset. Three optimized prediction models (SVM, RF, and LR) based on informative factors were built. Model performance was evaluated and compared with that of a simple classification method using the MoCA total score of 26 as a cutoff.

In the PPMI datasets, there was a discrepancy between the number of MoCA results in the normal cognition and cognitive impairment groups, resulting in an unbalanced dataset. Predic-

tions derived from an unbalanced dataset tend to be inclined to the majority group.<sup>23</sup> We therefore generated two datasets from the raw data of the PPMI datasets, with equal numbers of tests between groups. One dataset included all 221 MoCA results from 101 patients with PD-CI and the same number of results randomly sampled from 370 patients with PD-NC. In the other dataset, 101 MoCA results in each group were independently randomly sampled from 101 patients with PD-CI and 370 patients with PD-NC.

Each of the three original datasets was randomly split into training and testing datasets at an 80:20 ratio, which is the most common split ratio used for machine learning models. The ratio of MoCA results with PD-CI in the training and testing datasets was kept the same as the original dataset using stratified

sampling.<sup>24</sup> We repeated the random data sampling and train-test splitting to generate 1,000 different training and testing datasets to evaluate the performance of the above classification models. As the name suggests, the training dataset is used for training the model, and the testing dataset is used for testing the accuracy of the model. The process of model training and testing was performed 1,000 times with randomly formed training and testing datasets, and the model accuracy was averaged across all the trained models. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were used to evaluate the classification models.

Cognitive categorization (variable name in PPMI: "COGCAT"), which was determined based on level II neuropsychological tests according to the MDS Task Force recommendations,<sup>6</sup> was used as a dependent variable. PD cases with cognitive complaints were categorized as normal cognition, and a new independent variable of cognitive complaints was added. Since the PPMI manual indicates that the diagnosis of MCI and dementia requires cognitive complaints, patients with MCI and dementia were considered to have cognitive complaints. The MoCA total score was used in the cutoff analysis. In the classification analyses using machine learning (LR, RF, and SVM), MoCA scores in executive, visuospatial, memory, language, and attention domains were used as indicated in previous literature.<sup>10</sup> The total number of words in the verbal fluency test (variable name in the PPMI dataset: "MCAVFNUM") was included. The depression score was used as a variable by including binary results of 15 items in the SGDS. Demographic data, including age, sex, years of education, handedness, and duration of PD diagnosis, were also incorporated as independent variables.

## RESULTS

The demographic and clinical characteristics of the PD groups are summarized in Table 1 and Supplementary Tables 1 and 2 (in the online-only Data Supplement). Patients in the PD-CI group comprised a higher proportion of female patients were older and less educated, and had longer PD duration and more severe parkinsonian motor symptoms. The depression score was higher and MoCA scores were lower in the PD-CI group than in the PD-NC group. Of patients in the PD-NC group, 20.3% presented with cognitive complaints. Patients with PD-NC with cognitive complaints comprised a higher proportion of female patients, were older, and had longer PD duration, more severe parkinsonian motor symptoms, higher depression scores, and lower MoCA scores. We first compared the performance of different classification methods between the PD-NC and PD-CI groups (Table 2). In the analysis of all MoCA results using a cutoff of

**Table 1.** Demographic data of the participants

	PD-CI	PD-NC	<i>p</i> value
No. of PD cases	101	370	
No. of MoCA results	221	1,848	
No. of repeated tests	2.19 ± 1.56	4.99 ± 2.14	
Sex, female	184 (83.3)	1,293 (70.0)	< 0.001
Age (yr)	70.45 ± 8.35	64.41 ± 9.71	< 0.001
Education (yr)	15.25 ± 3.45	15.80 ± 2.71	0.023
Duration of PD (mon)*	62.78 ± 26.19	52.29 ± 27.28	< 0.001
MDS-UPDRS III total score*	36.57 ± 15.11	27.64 ± 12.06	< 0.001
Hoehn & Yahr stage*			< 0.001
1	20 (9.0)	326 (17.6)	
2	151 (68.3)	1,405 (76.0)	
3	43 (19.5)	98 (5.3)	
4	6 (2.7)	17 (0.9)	
5	1 (0.5)	2 (0.1)	
SGDS total score*	6.46 ± 1.97	5.40 ± 1.47	< 0.001
MoCA score	21.33 ± 4.62	27.26 ± 2.38	< 0.001

Values are presented as mean ± standard deviation or *n* (%). \*data obtained from the time point of MoCA. PD, Parkinson's disease; PD-CI, PD with cognitive impairment; PD-NC, Parkinson's disease with normal cognition; MoCA, Montreal Cognitive Assessment; MDS-UPDRS, Unified Parkinson's Disease Rating Scale by the Movement Disorders Society; SGDS, short version of the Geriatric Depression Scale.

26, the accuracy was 0.78 with a sensitivity of 0.81 and specificity of 0.78 (PPV = 0.30, NPV = 0.97). Analyses with three different machine learning methods using the MoCA domain scores exhibited better performance (accuracy = 0.93) but lower sensitivity (0.39–0.46). Low PPV in the cutoff method and low sensitivity in the machine learning analysis using all MoCA results were biased because the number of MoCA results in the PD-NC group was 8.4-fold higher than that in the PD-CI group. To avoid the potential bias caused by unbalanced data in the classification analysis,<sup>23</sup> we next performed classification analyses using the same number of MoCA results randomly sampled from the PD-NC and PD-CI groups. There was no difference in accuracy between the cutoff method (0.74 ± 0.03) and machine learning (0.78 ± 0.03) for the classification of cognitive status using a liberal dataset that permitted sampling from the same individual (*n* = 221 MoCA results in each group). However, the use of a more stringent dataset that excluded MoCA results from the same patients (*n* = 101 in each group) resulted in a significant reduction in the accuracy of the cutoff method (0.66 ± 0.05) but not of machine learning (0.74 ± 0.07). When cognitive complaints were included as an additional variable, classification accuracy was improved in both datasets of 221-201 and 101-101 MoCA results (accuracy of 0.89 and 0.87, respectively). The addition of SGDS results as an independent variable did not improve the machine learning classification performance. Furthermore, analysis using binary results of 30 items in the MoCA instead of domain scores

**Table 2.** Diagnostic performance of domain scores of the MoCA for distinguishing PD with normal cognition and cognitive impairment (mild cognitive impairment or dementia) according to the classification model

Dataset	No. of results (patients) in PD-CI	No. of results (patients) in PD-NC	Classification method	Types of MoCA data	Inclusion of SGDS	Inclusion of cognitive complaints	ACC (mean ± SD)	Sensitivity	Specificity	PPV	NPV
I	221 (101)	1,848 (370)	Cutoff 26	Total score	No	No	0.78 ± 0.02	0.81	0.78	0.30	0.97
	221 (101)	1,848 (370)	Logistic regression	Domain scores	No	No	0.93 ± 0.01	0.46	0.98	0.77	0.94
	221 (101)	1,848 (370)	Random forest	Domain scores	No	No	0.93 ± 0.01	0.41	0.99	0.82	0.93
	221 (101)	1,848 (370)	SVM	Domain scores	No	No	0.93 ± 0.01	0.39	0.99	0.83	0.93
II	221 (101)	221 (370)	Cutoff 26	Total score	No	No	0.74 ± 0.03	0.71	0.78	0.77	0.72
	221 (101)	221 (370)	Random forest	Domain scores	No	No	0.78 ± 0.03	0.72	0.84	0.82	0.75
	221 (101)	221 (370)	Random forest	Domain scores	No	Yes	0.89 ± 0.03	0.91	0.86	0.87	0.91
	221 (101)	221 (370)	Random forest	Domain scores	Yes	No	0.79 ± 0.03	0.73	0.84	0.83	0.75
	221 (101)	221 (370)	Random forest	Domain scores	Yes	Yes	0.88 ± 0.03	0.89	0.87	0.88	0.88
III	101 (101)	101 (370)	Cutoff 26	Total score	No	No	0.66 ± 0.05	0.61	0.71	0.69	0.64
	101 (101)	101 (370)	Random forest	Domain scores	No	No	0.74 ± 0.07	0.70	0.78	0.77	0.71
	101 (101)	101 (370)	Random forest	Domain scores	No	Yes	0.87 ± 0.05	0.91	0.82	0.84	0.89
	101 (101)	101 (370)	Random forest	Domain scores	Yes	No	0.76 ± 0.06	0.74	0.78	0.78	0.74
	101 (101)	101 (370)	Random forest	Domain scores	Yes	Yes	0.87 ± 0.05	0.90	0.83	0.85	0.88

Values for diagnostic performance were the average of 1,000 iterations. In dataset I, all 221 MoCA results from 101 patients with PD-CI and 1,848 results from 370 patients with PD-NC were used. In dataset II, all 221 MoCA results from 101 patients with PD-CI and 221 results randomly sampled from 370 patients with PD-NC were used. In dataset III, 101 MoCA results were independently randomly selected from 101 patients with PD-CI and 370 patients with PD-NC. MoCA, Montreal Cognitive Assessment; PD, Parkinson's disease; PD-CI, PD with cognitive impairment; PD-NC, PD with normal cognition; SGDS, short version of the Geriatric Depression Scale; ACC, accuracy; SD, standard deviation; PPV, positive predictive value; NPV, negative predictive value; SVM, support vector machine.

did not improve the classification performance (Supplementary Table 3 in the online-only Data Supplement).

We next compared the classification performance between the PD-NC and PD-MCI groups after excluding 84 MoCA results from 18 patients with a cognitive categorization of dementia (Table 3). The overall classification performance for the PD-MCI and PD-NC groups was similar to that in previous analyses including PDD. Classification using machine learning with MoCA domain scores exhibited a similar performance to that using a cutoff of 26 for the MoCA total score. The addition of cognitive complaints as a variable improved the performance of the machine learning method, but the addition of depression scores did not affect the results.

## DISCUSSION

To our knowledge, this study is the first to investigate the performance of machine learning using the MoCA for PD-CI diagnosis. Previous studies investigating the validity of the MoCA for the diagnosis and screening of PD-CI used a cutoff score and reported an accuracy, sensitivity, and specificity of 56%–64%, 83%–90%, and 44%–75%, respectively.<sup>7,11,12</sup> Our study evaluated whether a machine learning analysis using MoCA domain scores with several covariates could distinguish PD-CI from PD-

NC more effectively than a cutoff method using the MoCA total score. The machine learning analysis was trained by MoCA results with cognitive categorization diagnosed using comprehensive neuropsychological assessments (level II criteria). We observed that the accuracy of the machine learning analysis was higher than that of the cutoff method in a stringent dataset. Notably, the inclusion of cognitive complaints, but not depression scores, as a variable improved the classification performance of machine learning.

As the PPMI datasets included longitudinally collected results of cognitive assessments using both level I and II tests, analyses of these data could be biased by multiple longitudinal results from the same patients. An imbalance in the number of MoCA results between the PD-NC and PD-CI groups could introduce another source of bias. Hence, we analyzed several iterations of datasets after sampling MoCA results from the PD-NC and PD-CI groups by applying different dataset stringencies. The accuracy of the cutoff method varied considerably (0.66–0.75) depending on the dataset stringency. In contrast, the accuracy of the machine learning analysis was not influenced by the dataset stringency, with the caveat that training was performed with sufficient data. Exclusion of PDD from the dataset did not influence the classification performance of machine learning. Moreover, machine learning analysis has several advantages as follows: Cognition and its measurement are strongly influenced by



**Table 3.** Diagnostic performance of domain scores of the MoCA for distinguishing PD with normal cognition and mild cognitive impairment according to the classification model

Dataset	No. of results (patients) in PD-MCI	No. of results (patients) in PD-NC	Classification method	Types of MoCA data	Inclusion of SGDS	Inclusion of cognitive complaints	ACC (mean $\pm$ SD)	Sensitivity	Specificity	PPV	NPV
IV	137 (83)	1,848 (370)	Cutoff 26	Total score	No	No	0.78 $\pm$ 0.02	0.76	0.78	0.20	0.98
	137 (83)	1,848 (370)	Logistic regression	Domain scores	No	No	0.94 $\pm$ 0.01	0.26	0.99	0.70	0.95
	137 (83)	1,848 (370)	Random forest	Domain scores	No	No	0.94 $\pm$ 0.01	0.23	0.99	0.73	0.95
	137 (83)	1,848 (370)	SVM	Domain scores	No	No	0.94 $\pm$ 0.00	0.11	1.00	0.88	0.94
V	137 (83)	137 (370)	Cutoff 26	Total score	No	No	0.75 $\pm$ 0.04	0.71	0.78	0.77	0.72
	137 (83)	137 (370)	Random forest	Domain scores	No	No	0.79 $\pm$ 0.04	0.78	0.80	0.80	0.78
	137 (83)	137 (370)	Random forest	Domain scores	No	Yes	0.91 $\pm$ 0.04	0.97	0.84	0.86	0.97
	137 (83)	137 (370)	Random forest	Domain scores	Yes	No	0.80 $\pm$ 0.04	0.80	0.80	0.80	0.80
	137 (83)	137 (370)	Random forest	Domain scores	Yes	Yes	0.92 $\pm$ 0.03	0.99	0.85	0.87	0.98
VI	83 (83)	83 (370)	Cutoff 26	Total score	No	No	0.74 $\pm$ 0.05	0.76	0.71	0.73	0.75
	83 (83)	83 (370)	Random forest	Domain scores	No	No	0.75 $\pm$ 0.07	0.79	0.72	0.74	0.77
	83 (83)	83 (370)	Random forest	Domain scores	No	Yes	0.88 $\pm$ 0.05	0.95	0.81	0.83	0.94
	83 (83)	83 (370)	Random forest	Domain scores	Yes	No	0.76 $\pm$ 0.07	0.79	0.73	0.75	0.78
	83 (83)	83 (370)	Random forest	Domain scores	Yes	Yes	0.87 $\pm$ 0.05	0.94	0.81	0.83	0.93

Values for diagnostic performance were the average of 1,000 iterations. In dataset IV, all 137 MoCA results from 83 patients with PD-MCI and 1,848 results from 370 patients with PD-NC were used. In dataset V, all 137 MoCA results from 83 patients with PD-MCI and 137 results randomly sampled from 370 patients with PD-NC were used. In dataset VI, 83 MoCA results were independently randomly selected from 83 patients with PD-MCI and 370 patients with PD-NC. MoCA, Montreal Cognitive Assessment; PD, Parkinson's disease; PD-MCI, PD with mild cognitive impairment; PD-NC, PD with normal cognition; SGDS, short version of the Geriatric Depression Scale; ACC, accuracy; SD, standard deviation; PPV, positive predictive value; NPV, negative predictive value; SVM, support vector machine.

individual patient variables and their interaction with disease-related factors, which include age, educational background, pre-morbid function, and cognitive reserve.<sup>25</sup> The use of a cutoff method is unable to reflect the effects of these complex interactions. Education level in the PPMI dataset was relatively high, with small differences. However, in a population of patients with PD with a low education level, classification using the cutoff method may be misleading, unless appropriate norms are provided according to age and education level.<sup>14</sup> Likewise, in non-English speaking populations, appropriate cutoffs for different versions of the MoCA translated to different languages should be provided. Whether machine learning analysis with the MoCA is applicable in non-English speaking populations should be explored in future studies. Although depression influences cognitive function, the inclusion of PD cases with depression did not affect the classification performance despite incorporating the depression score as a variable in the machine learning analysis.

The MDS-recommended diagnostic criteria of PD-MCI stipulate a gradual decline in cognition reported by either the patient or informant or observed by the clinician. According to the PPMI operations manual, the diagnosis of PD-MCI requires cognitive complaints by either the patient or informant (spouse, family member, or friend). Our study demonstrated that the inclusion of cognitive complaints as a variable markedly improved the accuracy of the machine learning analysis in all datasets. However,

the retrospective nature of the dataset should be noted because all patients with MCI and dementia were considered to have cognitive complaints based on the PPMI operations manual, whereas 20.3% of patients with PD-NC had cognitive complaints. Information regarding cognitive complaints in the PPMI dataset was not collected using a standardized protocol. Indeed, impaired self-awareness of cognitive deficits has been reported in 16% of patients with PD-MCI and 21.8% of patients with de novo PD.<sup>26,27</sup> The proportion of patients classified with PD-MCI increased from 33% to 41% by eliminating the need for cognitive complaints and performing the diagnosis based on neurobehavioral signs and symptoms derived from the patient or informant.<sup>12</sup> Therefore, our machine learning analysis results including cognitive complaints may have overestimated the classification performance. Future studies investigating the actual magnitude of improvement in PD-CI classification using machine learning analysis, including subjective cognitive complaints obtained from the patient or informant and based on standardized methods, are warranted.

The MoCA enables the detection of mild cognitive changes in PD as well as the evaluation of specific cognitive domains. Characteristic profiles of dysfunctional cognitive domains in patients with PD-MCI or PDD have been reported,<sup>28,29</sup> with executive function being the most frequently involved.<sup>15,18</sup> Visuospatial dysfunction is an early feature of PD-MCI and is severely affected in PDD.<sup>16,17</sup> However, our results did not reveal any differences

in the performance of machine learning analyses between the use of domain scores and the individual data of 30 items. This may be due to the limitation of the MoCA in the evaluation of cognitive domains. Another possible cause is the heterogeneity of PD-MCI, whereby the involvement of multiple cognitive domains has been reported in 43%–93% of cases.<sup>15,18,30</sup>

This study has a few limitations. Most participants did not have de novo PD. Furthermore, education level was high in the PPMI datasets, which may have affected the validity for populations with low education. The optimal cutoff scores for the MoCA vary by race and ethnicity.<sup>31</sup> Therefore, for generalizability of our finding that machine learning analysis is superior to cutoff methods, future studies that are performed in patients of diverse languages, cultures or education levels are needed. Since the results of machine learning may vary depending on the quality of training datasets, high-quality training datasets, including both level I and level II tests, are absolutely important. Types of antiparkinsonian drugs were not included as covariates. Given that the primary purpose of this study was to test the performance of machine learning analysis with the MoCA compared to cognitive assessment by comprehensive neuropsychological tests, we do not think that the effect of medications on cognition differentially affects level I and level II tests. Analyzing the discrimination ability of individual domains, which we think is beyond the purpose of this study, was not included.

In conclusion, our study demonstrates that machine learning analysis using MoCA domain scores is a valid method for screening cognitive impairment in PD. Future studies are warranted to validate the performance of machine learning analysis using the MoCA with the inclusion of a history of cognitive complaints in prospectively enrolled patients with de novo PD with diverse language, culture, or education levels.

### Supplementary Materials

The online-only Data Supplement is available with this article at <https://doi.org/10.14802/jmd.22012>.

### Conflicts of Interest

The authors have no financial conflicts of interest.

### Funding Statement

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. NRF-2020R1F1A1050128).

### Author Contributions

Conceptualization: Yun Joong Kim. Data curation: Junbeom Jeon, Kiyong Kim, Kyeongmin Baek, Jeehee Yoon. Formal analysis: all authors. Funding acquisition: Yun Joong Kim. Investigation: Yun Joong Kim, Jeehee Yoon. Methodology: Yun Joong Kim, Kiyong Kim, Jeehee Yoon. Project administration: Yun Joong Kim, Jeehee Yoon. Resources: Yun Joong Kim. Software: Junbeom Jeon, Kiyong Kim, Kyeongmin Baek, Jeehee Yoon. Supervision: Yun Joong Kim, Kiyong Kim, Jeehee Yoon. Validation: Junbeom Jeon, Ki-

yong Kim, Kyeongmin Baek, Seok Jong Chung, Jeehee Yoon. Visualization: Junbeom Jeon. Writing—original draft: Junbeom Jeon, Yun Joong Kim. Writing—review & editing: Yun Joong, Kim, Kiyong Kim, Jeehee Yoon.

### ORCID iDs

Junbeom Jeon	<a href="https://orcid.org/0000-0002-3901-3242">https://orcid.org/0000-0002-3901-3242</a>
Kiyong Kim	<a href="https://orcid.org/0000-0003-3175-8132">https://orcid.org/0000-0003-3175-8132</a>
Kyeongmin Baek	<a href="https://orcid.org/0000-0001-6869-6815">https://orcid.org/0000-0001-6869-6815</a>
Seok Jong Chung	<a href="https://orcid.org/0000-0001-6086-3199">https://orcid.org/0000-0001-6086-3199</a>
Jeehee Yoon	<a href="https://orcid.org/0000-0002-4327-3135">https://orcid.org/0000-0002-4327-3135</a>
Yun Joong Kim	<a href="https://orcid.org/0000-0002-2956-1552">https://orcid.org/0000-0002-2956-1552</a>

### REFERENCES

- Litvan I, Aarsland D, Adler CH, Goldman JG, Kulisevsky J, Mollenhauer B, et al. MDS Task Force on mild cognitive impairment in Parkinson's disease: critical review of PD-MCI. *Mov Disord* 2011;26:1814-1824.
- Pedersen KF, Larsen JP, Tysnes OB, Alves G. Prognosis of mild cognitive impairment in early Parkinson disease: the Norwegian ParkWest study. *JAMA Neurol* 2013;70:580-586.
- Hely MA, Reid WG, Adena MA, Halliday GM, Morris JG. The Sydney multicenter study of Parkinson's disease: the inevitability of dementia at 20 years. *Mov Disord* 2008;23:837-844.
- Aarsland D, Andersen K, Larsen JP, Lolk A, Kragh-Sørensen P. Prevalence and characteristics of dementia in Parkinson disease: an 8-year prospective study. *Arch Neurol* 2003;60:387-392.
- Buter TC, van den Hout A, Matthews FE, Larsen JP, Brayne C, Aarsland D. Dementia and survival in Parkinson disease: a 12-year population study. *Neurology* 2008;70:1017-1022.
- Litvan I, Goldman JG, Tröster AI, Schmand BA, Weintraub D, Petersen RC, et al. Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Mov Disord* 2012;27:349-356.
- Hoops S, Nazem S, Siderowf AD, Duda JE, Xie SX, Stern MB, et al. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology* 2009;73:1738-1745.
- Zadikoff C, Fox SH, Tang-Wai DF, Thomsen T, de Bie RM, Wadia P, et al. A comparison of the mini mental state exam to the Montreal cognitive assessment in identifying cognitive deficits in Parkinson's disease. *Mov Disord* 2008;23:297-299.
- Chou KL, Amick MM, Brandt J, Camicioli R, Frei K, Gitelman D, et al. A recommended scale for cognitive screening in clinical trials of Parkinson's disease. *Mov Disord* 2010;25:2501-2507.
- Hendershott TR, Zhu D, Llanes S, Poston KL. Domain-specific accuracy of the Montreal Cognitive Assessment subsections in Parkinson's disease. *Parkinsonism Relat Disord* 2017;38:31-34.
- Dalrymple-Alford JC, MacAskill MR, Nakas CT, Livingston L, Graham C, Crucian GP, et al. The MoCA: well-suited screen for cognitive impairment in Parkinson disease. *Neurology* 2010;75:1717-1725.
- Marras C, Armstrong MJ, Meaney CA, Fox S, Rothberg B, Reginold W, et al. Measuring mild cognitive impairment in patients with Parkinson's disease. *Mov Disord* 2013;28:626-633.
- Kandiah N, Zhang A, Cenina AR, Au WL, Nadkarni N, Tan LC. Montreal Cognitive Assessment for the screening and prediction of cognitive decline in early Parkinson's disease. *Parkinsonism Relat Disord* 2014;20:1145-1148.
- Kim JI, Sunwoo MK, Sohn YH, Lee PH, Hong JY. The MMSE and MoCA for screening cognitive impairment in less educated patients with Parkinson's disease. *J Mov Disord* 2016;9:152-159.
- Kalbe E, Rehberg SP, Heber I, Kronenbueger M, Schulz JB, Storch A, et al. Subtypes of mild cognitive impairment in patients with Parkinson's disease: evidence from the LANDSCAPE study. *J Neurol Neurosurg Psychiatry* 2016;87:1099-1105.
- Johnson DK, Langford Z, Garnier-Villarreal M, Morris JC, Galvin JE. Onset of mild cognitive impairment in Parkinson disease. *Alzheimer Dis*

- Assoc Disord 2016;30:127-133.
17. Johnson DK, Galvin JE. Longitudinal changes in cognition in Parkinson's disease with and without dementia. *Dement Geriatr Cogn Disord* 2011; 31:98-108.
  18. Goldman JG, Weis H, Stebbins G, Bernard B, Goetz CG. Clinical differences among mild cognitive impairment subtypes in Parkinson's disease. *Mov Disord* 2012;27:1129-1136.
  19. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* 2022;23:40-55.
  20. Noble WS. What is a support vector machine? *Nat Biotechnol* 2006;24: 1565-1567.
  21. Breiman L. Random forests. *Mach Learn* 2001;45:5-32.
  22. Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics* 2018;19:270.
  23. He HB, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng* 2009;21:1263-1284.
  24. Elfil M, Negida A. Sampling methods in clinical research; an educational review. *Emerg (Tehran)* 2017;5:e52.
  25. Goldman JG, Holden SK, Litvan I, McKeith I, Stebbins GT, Taylor JP. Evolution of diagnostic criteria and assessments for Parkinson's disease mild cognitive impairment. *Mov Disord* 2018;33:503-510.
  26. Orfei MD, Assogna F, Pellicano C, Pontieri FE, Caltagirone C, Pierantozzi M, et al. Anosognosia for cognitive and behavioral symptoms in Parkinson's disease with mild dementia and mild cognitive impairment: frequency and neuropsychological/neuropsychiatric correlates. *Parkinsonism Relat Disord* 2018;54:62-67.
  27. Yoo HS, Chung SJ, Lee YH, Ye BS, Sohn YH, Lee PH. Cognitive anosognosia is associated with frontal dysfunction and lower depression in Parkinson's disease. *Eur J Neurol* 2020;27:951-958.
  28. Williams-Gray CH, Foltynie T, Brayne CE, Robbins TW, Barker RA. Evolution of cognitive dysfunction in an incident Parkinson's disease cohort. *Brain* 2007;130(Pt 7):1787-1798.
  29. Williams-Gray CH, Evans JR, Goris A, Foltynie T, Ban M, Robbins TW, et al. The distinct cognitive syndromes of Parkinson's disease: 5 year follow-up of the CamPaIGN cohort. *Brain* 2009;132(Pt 11):2958-2969.
  30. Lawrence BJ, Gasson N, Loftus AM. Prevalence and subtypes of mild cognitive impairment in Parkinson's disease. *Sci Rep* 2016;6:33929.
  31. Milani SA, Marsiske M, Striley CW. Discriminative ability of Montreal Cognitive Assessment subtests and items in racial and ethnic minority groups. *Alzheimer Dis Assoc Disord* 2019;33:226-232.



**Supplementary Table 1.** Demographic data of participants with PD with normal cognition, mild cognitive impairment, and dementia

	PD-NC ( <i>n</i> = 1,848)	PD-MCI ( <i>n</i> = 137)	PDD ( <i>n</i> = 84)	<i>p</i> values			
				Comparison among three groups	PD-NC vs. PD-MCI	PD-NC vs. PDD	PD-MCI vs. PDD
Sex, female	1,293 (70.0)	114 (83.2)	70 (83.3)	< 0.001			
Age (yr)	64.41 ± 9.71	70.58 ± 8.76	70.24 ± 7.66	< 0.001	< 0.001	< 0.001	> 0.999
Education (yr)	15.80 ± 2.71	15.66 ± 2.85	14.58 ± 4.19	0.031	> 0.999	< 0.001	0.017
Duration of PD (mon)*	52.29 ± 27.28	60.44 ± 27.26	66.60 ± 24.02	< 0.001	0.002	< 0.001	0.306
MDS-UPDRS III total score*	27.64 ± 12.06	34.59 ± 13.63	39.79 ± 16.84	< 0.001	< 0.001	< 0.001	0.008
Hoehn & Yahr stage*				< 0.001			
1	326 (17.6)	13 (9.5)	7 (8.3)				
2	1,405 (76.0)	98 (71.5)	53 (63.1)				
3	98 (5.3)	25 (18.2)	18 (21.4)				
4	17 (0.9)	1 (0.7)	5 (6.0)				
5	2 (0.1)	0 (0.0)	1 (1.2)				
SGDS total score*	5.40 ± 1.47	6.11 ± 1.85	7.04 ± 2.02	< 0.001	< 0.001	< 0.001	< 0.001
MoCA score	27.26 ± 2.38	22.37 ± 3.99	19.63 ± 5.07	< 0.001	< 0.001	< 0.001	< 0.001

Values are presented as mean ± standard deviation of *n* (%). \*data obtained from the time point of MoCA. PD, Parkinson's disease; PD-NC, PD with normal cognition; PD-MCI, PD with mild cognitive impairment; PDD, PD dementia; MDS-UPDRS, Unified Parkinson's Disease Rating Scale by the Movement Disorders Society; SGDS, short version of the Geriatric Depression Scale; MoCA, Montreal Cognitive Assessment.

**Supplementary Table 2.** PD-NC with versus without cognitive complaints

	PD-NC without cognitive complaints	PD-NC with cognitive complaints	<i>p</i> value
No. of MoCA results	1,473	375	
Male/female ( <i>n</i> )	567/906	113/262	0.003
Age (yr)	63.47 ± 9.79	68.09 ± 8.45	< 0.001
Education (yr)	15.83 ± 2.72	15.69 ± 2.68	0.369
Duration of PD (yr)*	50.44 ± 27.16	59.58 ± 26.54	< 0.001
MDS-UPDRS III total score*	26.73 ± 11.97	31.23 ± 11.73	< 0.001
Hoehn & Yahr stage* (I:II:III:IV:V)	291:1,100:59:11:2	35:295:39:6:0	< 0.001
SGDS total score*	5.30 ± 1.43	5.77 ± 1.57	< 0.001
MoCA score	27.51 ± 2.24	26.27 ± 2.67	< 0.001

Values are presented as mean ± standard deviation unless otherwise indicated. \*data obtained from the time point of MoCA. PD-NC, Parkinson's disease with normal cognition; MoCA, Montreal Cognitive Assessment; PD, Parkinson's disease; MDS-UPDRS, Unified Parkinson's Disease Rating Scale by the Movement Disorders Society; SGDS, short version of the Geriatric Depression Scale.

**Supplementary Table 3.** Diagnostic performance of binary data of 30 items of the MoCA for distinguishing PD-NC and PD-CI according to classification model

Dataset	No. of results (patients) in PD-CI	No. of results (patients) in PD-NC	Classification method	Type of MoCA data	Inclusion of SGDS	Inclusion of cognitive complaints	ACC (mean $\pm$ SD)	Sensitivity	Specificity	PPV	NPV
I	221 (101)	1,848 (370)	Cutoff 26	Total score	No	No	0.78 $\pm$ 0.02	0.76	0.81	0.30	0.97
	221 (101)	1,848 (370)	Logistic regression	Binary data	No	No	0.93 $\pm$ 0.01	0.47	0.98	0.76	0.94
	221 (101)	1,848 (370)	Random forest	Binary data	No	No	0.93 $\pm$ 0.01	0.36	0.99	0.86	0.93
	221 (101)	1,848 (370)	SVM	Binary data	No	No	0.93 $\pm$ 0.00	0.43	0.99	0.81	0.94
II	221 (101)	221 (370)	Cutoff 26	Total score	No	No	0.74 $\pm$ 0.03	0.71	0.78	0.77	0.72
	221 (101)	221 (370)	Random forest	Binary data	No	No	0.76 $\pm$ 0.04	0.68	0.84	0.82	0.72
	221 (101)	221 (370)	Random forest	Binary data	No	Yes	0.88 $\pm$ 0.04	0.88	0.88	0.88	0.88
	221 (101)	221 (370)	Random forest	Binary data	Yes	No	0.76 $\pm$ 0.03	0.68	0.85	0.82	0.72
	221 (101)	221 (370)	Random forest	Binary data	Yes	Yes	0.87 $\pm$ 0.03	0.85	0.88	0.88	0.85
III	101 (101)	101 (370)	Cutoff 26	Total score	No	No	0.66 $\pm$ 0.05	0.61	0.71	0.69	0.64
	101 (101)	101 (370)	Random forest	Binary data	No	No	0.75 $\pm$ 0.07	0.72	0.78	0.78	0.72
	101 (101)	101 (370)	Random forest	Binary data	No	Yes	0.87 $\pm$ 0.05	0.90	0.84	0.85	0.89
	101 (101)	101 (370)	Random forest	Binary data	Yes	No	0.76 $\pm$ 0.06	0.73	0.78	0.78	0.73
	101 (101)	101 (370)	Random forest	Binary data	Yes	Yes	0.86 $\pm$ 0.05	0.88	0.84	0.85	0.87

Values for diagnostic performance were the average of 1,000 iterations. In dataset I, all 221 MoCA results from 101 patients with PD-CI and 1,848 results from 370 patients with PD-NC were used. In dataset II, all 221 MoCA results from 101 patients with PD-CI and 221 results randomly sampled from 1,848 results tested on 370 patients with PD-NC were used. In dataset III, 101 MoCA results were randomly selected independently from 101 patients with PD-CI and 370 patients with PD-NC, respectively. MoCA, Montreal Cognitive Assessment; PD-NC, Parkinson's disease with normal cognition; PD-CI, Parkinson's disease with cognitive impairment; SGDS, short version of the Geriatric Depression Scale; ACC, accuracy; SD, standard deviation; PPV, positive predictive value; NPV, negative predictive value; SVM, support vector machine.