

World Journal of *Psychiatry*

World J Psychiatr 2020 November 19; 10(11): 245-285



ORIGINAL ARTICLE**Case Control Study**

- 245 Best early-onset Parkinson dementia predictor using ensemble learning among Parkinson's symptoms, rapid eye movement sleep disorder, and neuropsychological profile

Byeon H

Observational Study

- 260 Testing the hypothesis of subtypes of nonadherence in schizophrenia and schizoaffective disorder: A prospective study

Vega D, Acosta FJ, Saavedra P

META-ANALYSIS

- 272 Effects of smartphone-based interventions and monitoring on bipolar disorder: A systematic review and meta-analysis

Liu JY, Xu KK, Zhu GL, Zhang QQ, Li XM

ABOUT COVER

Editorial board member of *World Journal of Psychiatry*, Dr. Lenin Pavón is an immunologist, having received his PhD at Instituto Politécnico Nacional (Mexico City, Mexico). Since 2000, he has served as Head of the Department of Psychoimmunology at Instituto Nacional de Psiquiatría "Ramón de la Fuente Muñiz" in Mexico City. His career-long work has focused on studying the neuroendocrine-immune interactions in stress, major depression, fibromyalgia, schizophrenia, multiple sclerosis, and Alzheimer's disease. Dr. Pavón's research efforts have yielded a variety of professional awards as well as two books, 88 articles in indexed journals, and 19 chapters in specialized books. At present, he is a member of the editorial board of several international journals and scientific societies. His ongoing research interests involve inflammatory mediators' participation in psychiatric disorders and immunomodulation by neurotransmitters and hormones. (L-Editor: Filipodia)

AIMS AND SCOPE

The primary aim of *World Journal of Psychiatry* (*WJP*, *World J Psychiatr*) is to provide scholars and readers from various fields of psychiatry with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

WJP mainly publishes articles reporting research results and findings obtained in the field of psychiatry and covering a wide range of topics including adolescent psychiatry, biological psychiatry, child psychiatry, community psychiatry, ethnopsychology, psychoanalysis, psychosomatic medicine, etc.

INDEXING/ABSTRACTING

The *WJP* is now abstracted and indexed in Science Citation Index Expanded (SCIE, also known as SciSearch®), Current Contents/Clinical Medicine, Journal Citation Reports/Science Edition, PubMed, and PubMed Central. The 2020 edition of Journal Citation Reports® cites the 2019 impact factor (IF) for *WJP* as 3.545; IF without journal self cites: 3.545; Ranking: 46 among 155 journals in psychiatry; and Quartile category: Q2.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: Jia-Hui Li; Production Department Director: Yun-Xiaojuan Wu; Editorial Office Director: Jia-Ping Yan.

NAME OF JOURNAL

World Journal of Psychiatry

ISSN

ISSN 2220-3206 (online)

LAUNCH DATE

December 31, 2011

FREQUENCY

Monthly

EDITORS-IN-CHIEF

Rajesh R Tampi

EDITORIAL BOARD MEMBERS

<https://www.wjgnet.com/2220-3206/editorialboard.htm>

PUBLICATION DATE

November 19, 2020

COPYRIGHT

© 2020 Baishideng Publishing Group Inc

INSTRUCTIONS TO AUTHORS

<https://www.wjgnet.com/bpg/gerinfo/204>

GUIDELINES FOR ETHICS DOCUMENTS

<https://www.wjgnet.com/bpg/GerInfo/287>

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH

<https://www.wjgnet.com/bpg/gerinfo/240>

PUBLICATION ETHICS

<https://www.wjgnet.com/bpg/GerInfo/288>

PUBLICATION MISCONDUCT

<https://www.wjgnet.com/bpg/gerinfo/208>

ARTICLE PROCESSING CHARGE

<https://www.wjgnet.com/bpg/gerinfo/242>

STEPS FOR SUBMITTING MANUSCRIPTS

<https://www.wjgnet.com/bpg/GerInfo/239>

ONLINE SUBMISSION

<https://www.f6publishing.com>

Case Control Study

Best early-onset Parkinson dementia predictor using ensemble learning among Parkinson's symptoms, rapid eye movement sleep disorder, and neuropsychological profile

Haewon Byeon

ORCID number: Haewon Byeon
0000-0002-3363-390X.

Author contributions: Byeon H designed the paper, was involved in study data interpretation, performed the statistical analysis, and assisted with writing the article.

Supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education, No. NRF-2018R1D1A1B07041091 and NRF-2019S1A5A8034211.

Institutional review board

statement: The study was reviewed and approved by the National Biobank of Korea Institutional Review Board, Approval No. KBN-2019-005.

Informed consent statement: All patients gave informed consent.

Conflict-of-interest statement: No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

Data sharing statement: Technical appendix, statistical code, and

Haewon Byeon, Department of Medical Big Data, College of AI Convergence, Inje University, Gimhae 50834, Gyeongsangnamdo, South Korea

Corresponding author: Haewon Byeon, DSc, PhD, Professor, Department of Medical Big Data, College of AI Convergence, Inje University, Gimhae 50834, Gyeongsangnamdo, South Korea.
bhwpuma@naver.com

Abstract

BACKGROUND

Despite the frequent progression from Parkinson's disease (PD) to Parkinson's disease dementia (PDD), the basis to diagnose early-onset Parkinson dementia (EOPD) in the early stage is still insufficient.

AIM

To explore the prediction accuracy of sociodemographic factors, Parkinson's motor symptoms, Parkinson's non-motor symptoms, and rapid eye movement sleep disorder for diagnosing EOPD using PD multicenter registry data.

METHODS

This study analyzed 342 Parkinson patients (66 EOPD patients and 276 PD patients with normal cognition), younger than 65 years. An EOPD prediction model was developed using a random forest algorithm and the accuracy of the developed model was compared with the naive Bayesian model and discriminant analysis.

RESULTS

The overall accuracy of the random forest was 89.5%, and was higher than that of discriminant analysis (78.3%) and that of the naive Bayesian model (85.8%). In the random forest model, the Korean Mini Mental State Examination (K-MMSE) score, Korean Montreal Cognitive Assessment (K-MoCA), sum of boxes in Clinical Dementia Rating (CDR), global score of CDR, motor score of Untitled Parkinson's Disease Rating (UPDRS), and Korean Instrumental Activities of Daily Living (K-IADL) score were confirmed as the major variables with high weight for EOPD prediction. Among them, the K-MMSE score was the most important factor in the final model.

dataset available from the corresponding author at bhwpuma@naver.com.

STROBE statement: The authors have read the STROBE Statement – checklist of items, and the manuscript was prepared and revised according to the STROBE Statement – checklist of items.

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

Manuscript source: Invited manuscript

Specialty type: Psychiatry

Country/Territory of origin: South Korea

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B
Grade C (Good): 0
Grade D (Fair): 0
Grade E (Poor): 0

Received: July 21, 2020

Peer-review started: July 21, 2020

First decision: September 17, 2020

Revised: September 27, 2020

Accepted: October 11, 2020

Article in press: October 11, 2020

Published online: November 19, 2020

P-Reviewer: Wang YP

S-Editor: Huang P

L-Editor: Webster JR

P-Editor: Li JH



CONCLUSION

It was found that Parkinson-related motor symptoms (e.g., motor score of UPDRS) and instrumental daily performance (e.g., K-IADL score) in addition to cognitive screening indicators (e.g., K-MMSE score and K-MoCA score) were predictors with high accuracy in EOPD prediction.

Key Words: Early-onset Parkinson dementia; Ensemble learning method; Neuropsychological test; Risk factor; Discriminant analysis; Naive Bayesian model

©The Author(s) 2020. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: It is believed that if the Korean Mini Mental State Examination (K-MMSE) is given priority over other cognitive screening tests in order to distinguish early-onset Parkinson dementia (EOPD) from Parkinson's disease, the accuracy of detecting EOPD will be higher than conducting other screening tests first. However, further epidemiological studies will be needed to fully comprehend the results of better accuracy of the K-MMSE than that of Korean Montreal Cognitive Assessment while detecting EOPD using the developed ensemble-based prediction model.

Citation: Byeon H. Best early-onset Parkinson dementia predictor using ensemble learning among Parkinson's symptoms, rapid eye movement sleep disorder, and neuropsychological profile. *World J Psychiatr* 2020; 10(11): 245-259

URL: <https://www.wjgnet.com/2220-3206/full/v10/i11/245.htm>

DOI: <https://dx.doi.org/10.5498/wjp.v10.i11.245>

INTRODUCTION

Dementia is a typical senile disease and indicates that a person with normal cognition experiences cognitive impairment due to various causes in the aging process. Dementia shows diverse symptoms, such as memory impairment; decreased cognitive functions, including language ability and frontal lobe executive function; and psychological symptoms of dementia (BPDS), depending on the type and progression of the disease^[1]. It burdens caregivers psychologically and economically as well as the dementia patients^[2]. In particular, care costs are extremely high as there is currently no cure for dementia, and it is necessary to take care of patients for a long time. As of 2019, South Korea spends KRW 14.6 trillion managing dementia, which is 0.8% of the GDP, and it is expected to increase more than seven times (KRW 106.5 trillion)^[3]. It was reported that South Korea had 700000 dementia patients out of 7 million elderly people in 2017, which is already over 10% of the total elderly population^[3]. It is approximately a 35% increase from 540000 dementia patients in 2012^[3]. It tends to increase steadily by more than 7% per year^[3]. Therefore, the reduction of dementia prevalence through the prevention, early diagnosis, and early management of dementia is the key to the mental health policy that the South Korean government must resolve as South Korea has an aging society.

Geriatricians evaluate the characteristics of dementia by classifying it into several types to diagnose dementia as soon as possible. Recently, many studies^[4,5] examined the characteristics of the disease after categorizing it into early-onset dementia (EOD: Occurring before 65 years old) and late-onset dementia (LOD: Occurring at 65 years old or later) based on the onset of dementia symptoms (age). These studies revealed that EOD and LOD showed differences in imaging tests as the disease progressed. For example, in the case of Alzheimer's disease, EOD caused a greater loss of cerebral synapses or severe infiltration of senile plaques and neurofibrillary tangles than LOD^[6-8]. Moreover, even the frontal and parietal lobes, as well as the temporal lobe, atrophied^[6-8]. Additionally, since EOD patients are more likely to have a family history of dementia than LOD patients, it is suspected that EOD is affected by genetic predisposition more than LOD^[7]. However, these imaging tests are not the ideal way to identify the onset of dementia in the early stages because they can only be used to diagnose dementia accurately by skilled medical personnel after dementia has progressed to some extent. Moreover, previous studies^[9] that examined the

characteristics of EOD mostly evaluated Alzheimer's dementia. The demographic and neuropsychological characteristics of early-onset Parkinson dementia (EOPD) are relatively unknown.

In summary, despite the frequent progression from Parkinson's disease (PD) to Parkinson dementia (PDD), the basis to diagnose EOPD in the early stages is still insufficient. Currently, it is impossible to detect EOPD in the early stages just by using the cognitive screening test that is simply and commonly conducted for all types of dementia patients in South Korean public health centers. Although several biomarker candidates have been suggested based on the cerebrospinal fluid (CSF) test (e.g., A β 1-42 and total tau) for diagnosing dementia in the early stages^[10], the CSF test is not versatile because it causes pain in examinees, thus examinees tend to reject the test, and the reliability of it cannot be tested; these are limitations of this test. In other words, because a range of factors (e.g., gender, education level, and depression) affect EOPD^[11-13], it would be necessary to develop a prediction model by applying PD motor and non-motor symptoms and sociodemographic indices^[11,12] in addition to cognitive characteristics^[14]. Byeon^[15] argued that previous studies^[16,17] were limited to the exploration of individual risk factors because they used regression models for predicting dementia and variables were quite limited because the prediction models mainly included neuropsychological tests. Therefore, there are limitations to developing a highly reliable model to predict EOPD using individual (single) indicators, such as PD symptoms and neuropsychological tests. In order to develop an accurate prediction model, it is necessary to develop a comprehensive model that includes sociodemographic indices, PD motor symptoms, PD non-motor symptoms, rapid eye movement (REM) sleep behavior disorder, and neuropsychological indices.

Recent studies have used machine learning algorithms as a method to predict a high disease risk group^[18,19]. Machine learning is a process of analyzing relationships and rules in data to extract valuable information from the data. Random decision forest (RF) has been used widely; it produces many decision trees using an ensemble algorithm to overcome the limitations of overfitting and predicts target variables by combining them^[20].

We are not aware of any published RF-based machine learning studies to analyze EOPD prediction capability by considering sociodemographic factors, PD motor symptoms, PD non-motor symptoms, REM sleep disorder, and neuropsychological profiles, together with cognitive function. This study explored the prediction accuracy of sociodemographic factors, PD motor symptoms, PD non-motor symptoms, and REM sleep disorder for diagnosing EOPD using a large-scale PD registry dataset.

MATERIALS AND METHODS

Participants

This study was performed by analyzing the Parkinson's Disease Epidemiology in Korea (PDEPI-Korea) multicenter registry data provided by the National Biobank of Korea (NB-Korea, No. KBN-2019-005). The study was approved by the Research Ethics Review Board of the NB-Korea (No. KBN-2019-005) and the Korea Centers for Disease Control and Prevention (Korea-CDC, No. KBN-2019-1327). The NB-K was founded in 2008 upon the approval of the Ministry of Health and Welfare due to the necessity of managing bio data systematically at the national level. It has been managed by the Korea CDC. The ultimate goal of the NBK is to promote biomedical research and public health. Please see Byeon^[20] for details on the data source.

PD was diagnosed by a psychiatrist according to the diagnostic criteria of the United Kingdom Parkinson's Disease Society Brain Bank^[21]. In this study, PDD was defined as patients who met the diagnostic criteria of probable PDD, suggested by the Movement Disorder Society Task Force^[22]. This study excluded patients who had other causative diseases, such as hydrocephalus and vascular Parkinsonism, determined from magnetic resonance imaging. This study analyzed 342 PD patients [66 EOPD patients and 276 PD patients with normal cognition (PD-NC)] who were younger than 65 years. Sample size calculations based on power analysis are shown in Figure 1. As the minimum number of samples calculated based on power analysis was 210 (group 1 = 105, group 2 = 105) with significance level (α) = 0.05, effect size $d = 0.5$ and power of test $(1-\beta) = 0.95$ on the standard of normal distribution, the number of samples ($n = 342$) in our study was appropriate.

Measurement

The outcome variable was defined as the presence of EOPD (yes or no), a binary

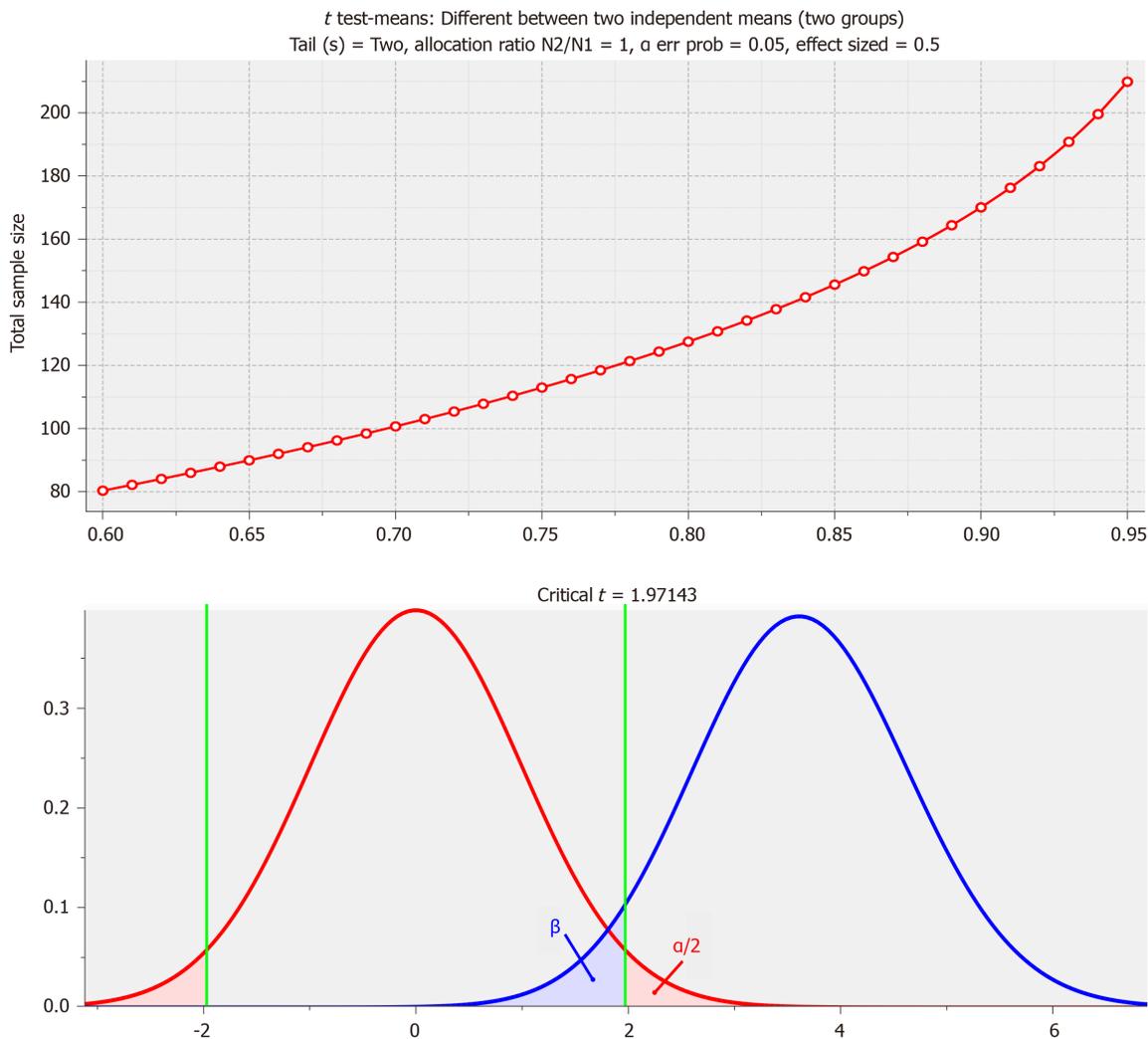


Figure 1 Sample size calculations.

variable, from a diagnosis by a neurologist. The explanatory variables included age; gender; education level (middle school graduate and below, or high school graduate and above); dominant hand (left hand or right hand); family PD history (yes or no); family dementia history (yes or no); pack-years (non-smoking, 1-20, 21-40, 41-60, or ≥ 61 pack-years); coffee-drinking (yes or no); coffee drinking period (no, ≤ 5 , 6-9, or ≥ 10 years); mean coffee intake per day (no, ≤ 1 , 2-3, or ≥ 4 cups); pesticide exposure recognition (never, currently not exposed but exposed previously, or currently exposed to pesticide); disease history (manganese poisoning, carbon monoxide poisoning, encephalitis, traumatic brain injury, stroke, alcoholism, diabetes, hyperlipidemia, hypertension, and/or atrial fibrillation); PD-related motor signs (tremor, akinesia/bradykinesia, postural instability, and/or late motor complications); neuropsychological characteristics determined by assessments such as the Korean Mini Mental State Examination (K-MMSE)^[23], the Korean Montreal Cognitive Assessment (K-MoCA)^[24], the sum of boxes in Clinical Dementia Rating (CDR)^[25], the global CDR score^[26], Korean Instrumental Activities of Daily Living (K-IADL)^[27], the total score of Untitled Parkinson’s Disease Rating (UPDRS)^[28], the motor score of UPDRS^[29], Hoehn and Yahr staging (H&Y staging)^[30], and the Schwab & England Activities of Daily Living scale (Schwab & England ADL)^[31]; and REM sleep behavior disorders. The definitions of the explanatory variables are shown in Table 1.

Development and evaluation of EOPD prediction model

The EOPD prediction model was developed using a RF algorithm and the accuracy of the developed model was compared with the naive Bayesian model and discriminant analysis. All analyses were performed using R version 3.5.2 (Foundation for Statistical Computing, Vienna, Austria).

RF is an ensemble classifier that randomly learns multiple decision trees and is a

Table 1 Measurement and definition of variables

Variable	Measurement	Characteristics
Sociodemographic factors	Age	Continuous variable
	Gender	Male or female
	Education	Middle school graduate and below or high school graduate and above
	Mainly used hand	Left hand, right hand, or both hands
	Family dementia history	Yes or No
	Family PD history	Yes or No
	Environmental factors	Exposure to pesticide
Health behaviors	Pack-years	Non-smoking, 1-20, 21-40, 41-60, or ≥ 61 pack-years
	Coffee-drinking	Yes or No
	Mean coffee intake per day (cups/d)	No, ≤ 1 , 2-3, or ≥ 4 cups
	Coffee drinking period (yr)	No, ≤ 5 , 6-9, or ≥ 10
Disease history	Carbon monoxide poisoning	Yes or No
	Manganese poisoning	Yes or No
	Encephalitis	Yes or No
	Traumatic brain injury	Yes or No
	Stroke	Yes or No
	Alcoholism	Yes or No
	Diabetes	Yes or No
	Hypertension	Yes or No
	Hyperlipidemia	Yes or No
	Atrial fibrillation	Yes or No
Neuropsychological characteristics	Total score of KMMSE	Continuous variable
	Total score of KMoCA	
	Global CDR score	
	Sum of boxes in CDR	
	K-IADL	
	Total score of UPDRS	
	Motor score of UPDRS	
	H&Y staging	
	Schwab & England ADL	
	Sleep behavior disorders	REM sleep behavior disorders
Exercise characteristics related to PD (PD related motor signs)	Tremor	Yes or No
	Rigidity	Yes or No
	Bradykinesia	Yes or No
	Postural instability	Yes or No
	LMC	Yes or No

PD: Parkinson's disease; KMMSE: Korean Mini Mental State Examination; KMoCA: Korean Montreal Cognitive Assessment; CDR: Clinical Dementia Rating; K-IADL: Korean Instrumental Activities of Daily Living; UPDRS: Score of Untitled Parkinson's Disease Rating; H&Y staging: Hoehn and Yahr staging; Schwab & England ADL: Schwab & England Activities of Daily Living scale; REM: Rapid eye movement; LMC: Late motor complications.

machine learning method based on the meta-learning of decision trees. It consists of a training stage composing many decision trees and a test stage that classifies or predicts when an input vector is entered.

The ensemble form of training data can be expressed as Forest $F = \{f_1, \dots, f_n\}$. The distributions obtained from the decision trees of each forest were averaged by the number (T) of decision trees and were then classified. For combining the predictors of each sample, the average was used when the target variable was a continuous variable, and the majority vote was used when it was a categorical variable (Figure 2).

RF is similar to bagging in that it improves stability by combining decision trees generated from multiple bootstrap samples, based on the majority rule. However, it is conceptually different from bagging because it uses explanatory variables, which are randomly selected in each bootstrap sample. The RF can be theoretically free from overfitting because it contains randomness for both features and learning instances. Moreover, it is not much affected by noise or outliers and it is more accurate than other machine learning methods, such as decision trees. The accuracy of RF increases when the number of trees increases. However, it may suffer from an elbow point, indicating a steep decrease in slope. Moreover, each tree is more likely to have a more complex structure when non-critical explanatory variables are selected. Consequently, this study used the grid search method that can minimize problems such as elbow point by considering mtry number ($n_estimator$), indicating the number of candidates for explanatory variables among RF hyperparameters in advance. The procedure of developing an RF-based prediction model is presented in Figure 3.

Comparison of model prediction accuracy

This study selected an algorithm with the best model performance as the final model by comparing the overall accuracy of RF, discriminant analysis, and the naive Bayesian model. Moreover, this study showed the variable importance of the final model. A partial dependence plot was presented to visually confirm the marginal effects of an input variable with the highest importance on a response variable. The function of partial dependence is given in the following Equation.

In the above equation, $p_1(x, x_{ic})$ is $\Pr(Y = 1)$, calculated from a specific value of an interest variable (x) and a fixed value of the remaining predictor (x_{ic}). This probability is calculated as the ratio classified as $Y = 1$ category in the corresponding random decision tree. In other words, partial dependence and the log odds of the logic model share the same concept, and it is the mean after calculating the log odds from all observations i .

RESULTS

General characteristics of the participants

The general characteristics of 342 participants with PD were analyzed (Table 2). The mean age of the subjects was 57.3 years old ($SD = 5.7$). The initial age at diagnosis of PD was 56.7 years old ($SD = 5.9$). Smokers made up 88.6% of the participants, subjects with a family history of PD were 5.1% of the participants, and subjects with a family history of dementia were 7.4% of the participants. It was found that 19.3% of the subjects had EOPD.

Development of the EOPD prediction model using RF

This study changed mtry values, presenting the number of explanatory variables to be used in the decision tree constituting RF, from 3 to 13, and selected values with the smallest error of Out-Of-Bag. The changes in the error of Out-Of-Bag are presented in Table 3. The optimal mtry to be applied in this study was 4, showing the lowest error rate (10.5%). When n_tree , the number of tree generations, and mtry were set as 500 and 4, respectively, the final RF model of this study had an overall accuracy of 89.5%.

Selection of the final EOPD prediction model

The overall accuracy of the RF was 89.5%, and it was higher than that of both discriminant analysis (78.3%) and the naive Bayesian model (85.8%). Therefore, the RF was assumed to be the most accurate prediction model among EOPD prediction models, and it was selected as the final prediction model. In Figure 3, the black line indicates the changes in each error rate against 500 bootstrap samples. Figure 4 shows that the changes in error rate became relatively stable when the number of bootstrap samples exceeded 60. Additionally, the multidimensional scaling plot of RF, which

Table 2 General characteristics of the subjects (*n* = 342)

Characteristics	<i>n</i> (%)
Age, mean ± SD (yr)	57.3 ± 5.7
K-MMSE, mean ± SD	25.6 ± 4.0
K-MoCA, mean ± SD	21.2 ± 5.1
Global CDR score, mean ± SD	0.4 ± 0.3
Sum of boxes in CDR, mean ± SD	1.5 ± 1.7
K-IADL, mean ± SD	1.0 ± 2.4
Total score of UPDRS, mean ± SD	41.3 ± 21.8
Motor score of UPDRS, mean ± SD	23.1 ± 11.1
H&Y staging, mean ± SD	2.3 ± 0.6
Schwab & England ADL, mean ± SD	77.5 ± 15.0
Gender	
Male	174 (50.9)
Female	168 (49.1)
Education	
Middle school graduate and below	195 (57.0)
High school graduate and above	147 (43.0)
Handness	
Right	318 (93.0)
Left	15 (4.4)
Both hands	9 (2.6)
Family PD history	
No	279 (94.9)
Yes	15 (5.1)
Family dementia history	
No	264 (92.6)
Yes	21 (7.4)
Smoking (pack year)	
1-20	18 (7.9)
21-40	9 (2.6)
41-60	3 (0.9)
61+	303 (88.6)
Coffee consumption	
No	174 (50.9)
Yes	168 (49.1)
Carbon monoxide poisoning	
No	294 (93.3)
Yes	21 (6.7)
Traumatic brain injury	
No	306 (97.1)
Yes	9 (2.9)
Diabetes	

No	276 (82.3)
Yes	60 (17.7)
Hypertension	
No	249 (73.5)
Yes	90 (26.5)
Hyperlipidemia	
No	303 (89.4)
Yes	36 (10.6)
Atrial fibrillation	
No	336 (99.1)
Yes	3 (0.9)
Tremor	
No	120 (36.0)
Yes	213 (64.0)
Rigidity	
No	24 (7.2)
Yes	309 (92.8)
Bradykinesia	
No	21 (6.3)
Yes	312 (93.7)
Postural instability	
No	159 (50.5)
Yes	156 (49.5)
REM sleep behavior disorders	
No	195 (61.3)
Yes	123 (38.7)
Late motor complications	
Only ON-OFF/Wearing OFF	57 (17.9)
Only levodopa-induced dyskinesia	12 (3.8)
Both ON-OFF/Wearing OFF and levodopa-induced dyskinesia are present	48 (15.1)
Both ON-OFF/Wearing OFF and levodopa-induced dyskinesia are absent	201 (63.2)
Depression	
No	147 (67.1)
Yes	72 (32.9)

K-MMSE: Korean Mini Mental State Examination; K-MoCA: Korean Montreal Cognitive Assessment; CDR: Clinical Dementia Rating; K-IADL: Korean Instrumental Activities of Daily Living; UPDRS: Score of Untitled Parkinson’s Disease Rating; H&Y staging: Hoehn and Yahr staging; Schwab & England ADL: Schwab & England Activities of Daily Living; PD: Parkinson’s disease; REM: Rapid eye movement.

visualizes the classification results through a two-dimensional diagram, is presented in [Figure 5](#).

Importance of variables in the final EOPD prediction model

The normalized importance of variables in the RF model, the final model, is presented in [Figure 6](#) and [Table 4](#). In this model, K-MMSE score, K-MoCA score, sum of boxes in CDR, global score of CDR, motor score of UPDRS, and K-IADL score were confirmed as the major variables with high weight for EOPD prediction. Among them, K-MMSE score was the most important factor in the final model.

Table 3 Error of Out-Of-Bag

Mtry (n)	Error of Out-Of-Bag
3	0.140
4	0.105
5	0.149
6	0.132
7	0.140
8	0.123
9	0.149
10	0.123
11	0.140
12	0.158
13	0.149

The partial dependence plot for K-MMSE, the most important variable in the EOPD prediction model, is presented in Figure 7. When the other factors (variables) were identical, the probability of the absence of EOPD tended to decrease as K-MMSE scores increased (Figure 6). In other words, it was confirmed that K-MMSE had the largest impact on EOPD prediction even after adjusting for other neuropsychological tests, PD symptoms, medical history, REM sleep disorder, depression, and sociodemographic factors.

DISCUSSION

Choosing a test with high feasibility and accuracy is critical to easily detect EOPD from PD in the point-of-care environment. It is required to comprehensively compare prediction accuracy for various predictors of EOPD, such as neuropsychological tests, lifestyle, sociodemographic factors, PD symptoms, depression, and REM sleep disorders. This study analyzed the prediction accuracy of various cognitive screening tests and neuropsychological profiles that could distinguish EOPD from PD using RF. The tests were ranked from greatest to least accurate as follows: K-MMSE score, K-MoCA score, sum of boxes in CDR, global score of CDR, motor score of UPDRS, and K-IADL score. It is noteworthy that the motor score of UPDRS, in addition to cognitive screening tests, was an important test in predicting EOPD. This is probably because participants with EOPD were more likely to show non-typical symptoms, such as movement problems, gait problems, and coordination problems^[22], and the motor score of UPDRS could comprehensively measure these PD motor symptoms.

In this study, K-MMSE score was the most important neuropsychological test for detecting EOPD. Moreover, the accuracy of K-MMSE was higher than that of K-MoCA. An essential factor in the diagnosis of EOPD is a decline in cognitive function that began after the onset of PD. This decline in cognitive function gradually progresses in various domains, such as executive function, memory, and visuospatial function. In particular, it has been reported that PDD patients experience impaired executive functions, reflecting the decrease in the ability to solve problems from the early stages of dementia^[32] as well as impaired visuospatial function^[33]. It is known that the impairment of other types of dementia, such as Alzheimer's disease, was significantly lower than that^[34]. K-MMSE and K-MoCA have been widely used as screening tests that simply compare the decline of various cognitive functions by types and comprehensively assess cognitive functions prior to in-depth tests in the point-of-care environment. The results of this study showed that the accuracy of K-MMSE was higher than that of K-MoCA when distinguishing EOPD from PD. Therefore, it is believed that if K-MMSE is given priority over other cognitive screening tests in order to distinguish EOPD from PD, the accuracy of detecting EOPD will be higher than conducting other screening tests first. However, further epidemiological studies will be needed to fully comprehend the results of the better accuracy of K-MMSE than that of K-MoCA while detecting EOPD, using the developed ensemble-based prediction

Table 4 The normalized importance of variables in the random forest model

Variables	Mean decrease Gini
K-MMSE	7.224
K-MoCA	2.992
Sum of boxes in CDR score	2.872
Global CDR score	2.304
Motor UPDRS	2.104
K-IADL	1.720
Total UPDRS	1.587
Schwab & England ADL	1.258
H&Y staging	1.040
Late motor complications	0.775
Consumption of coffee	0.699
Education level	0.527
Pack year	0.505
BDI	0.409
Tremor	0.338
Postural instability	0.338
Rigidity	0.331
Gender	0.255
REM sleep behavior disorders	0.249
Hypertension	0.168
Handness	0.151
Diabetes	0.146
Hyperlipidemia	0.129
Carbon monoxide poisoning	0.124
Family PD history	0.074
Family dementia history	0.066
Bradykinesia	0.034
Manganese poisoning	0.013
Traumatic brain injury	0.002
Atrial fibrillation	< 0.001

K-MMSE: Korean Mini Mental State Examination; K-MoCA: Korean Montreal Cognitive Assessment; CDR: Clinical Dementia Rating; K-IADL: Korean Instrumental Activities of Daily Living; UPDRS: Score of Untitled Parkinson's Disease Rating; H&Y staging: Hoehn and Yahr staging; Schwab & England ADL: Schwab & England Activities of Daily Living; PD: Parkinson's disease; REM: Rapid eye movement; BDI: Beck Depression Inventory.

model. Machine learning has the disadvantage of being unable to interpret the derived results, although it has better prediction accuracy than the traditional regression model. Therefore, future studies are required to develop explainable artificial intelligence models that have high prediction accuracy and are able to interpret results.

Another finding of this study was that the accuracy of RF was higher than that of both the naive Bayesian model and discriminant analysis. These results agreed with Byeon^[35], which showed that the ensemble algorithm was more accurate than a regression analysis or decision trees for predicting cognitive impairment in old age. RF has high prediction performance because it generates various decision trees from a number of bootstrap samples, and the possibility of overfitting is low^[35]. In particular,

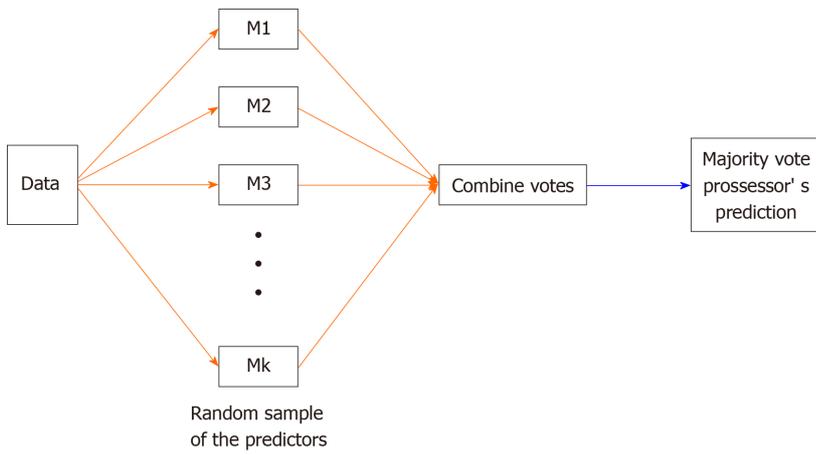


Figure 2 The random forest.

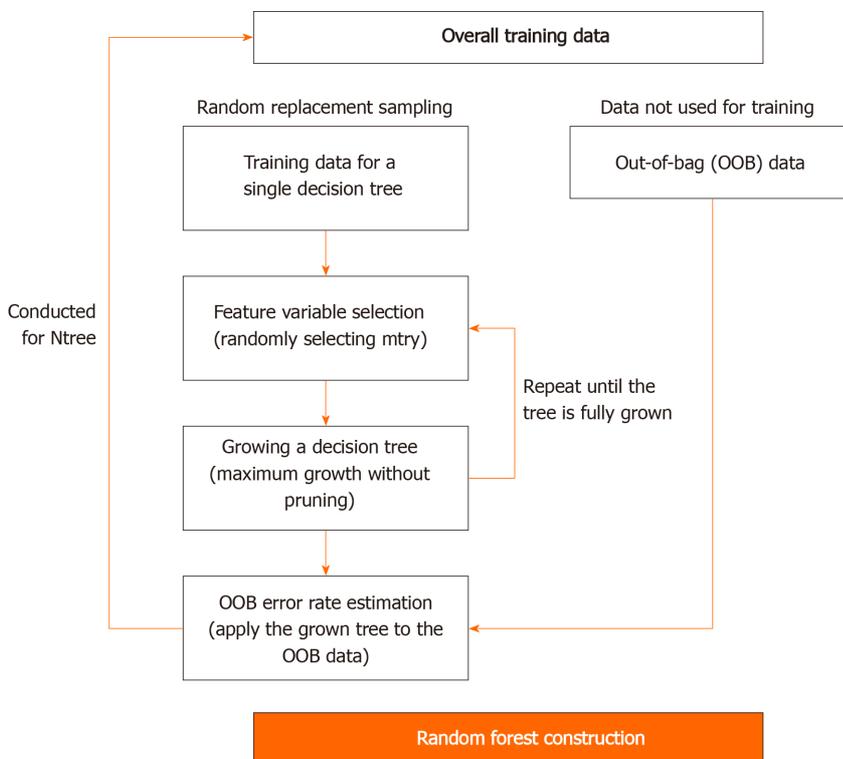


Figure 3 The development process of a random decision forest-based prediction model.

Equation:
$$f(x) = \frac{1}{n} \sum_{i=1}^n \log \left(\frac{p_1(x, x_{ic})}{p_0(x, x_{ic})} \right)$$

RF showed good predictive performance even when classifying binary variables using imbalanced disease data^[15,18]. Therefore, it is believed that, compared to traditional statistical techniques such as discriminant analysis, using RF will increase accuracy while exploring major variables, allowing us to predict EOPD.

The importance of this study was that it identified the prediction accuracy of sociodemographic factors, PD motor symptoms, PD non-motor symptoms, REM sleep disorder, and neuropsychological profiles for distinguishing EOPD from PD, using national examination data conducted by the National Biobank of Korea. The limitations of the study are as follows: (1) The data source of this study was the registry data of multiple institutions and subjects were not randomly sampled; (2) The prediction model did not include candidate markers, genetic information, or biomarkers; (3) Genes such as *PRKN* and *LRRK2* are known to be risk factors for PD

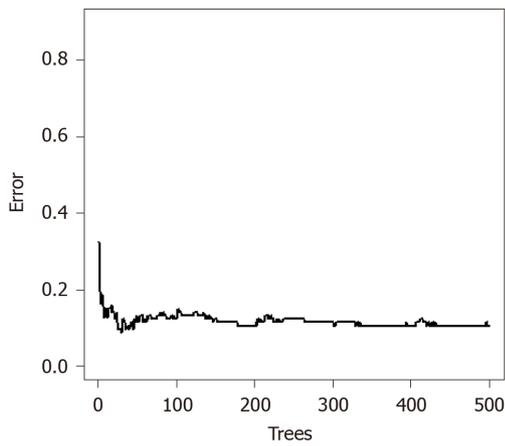


Figure 4 Error rate of the random forest model (500 trees).

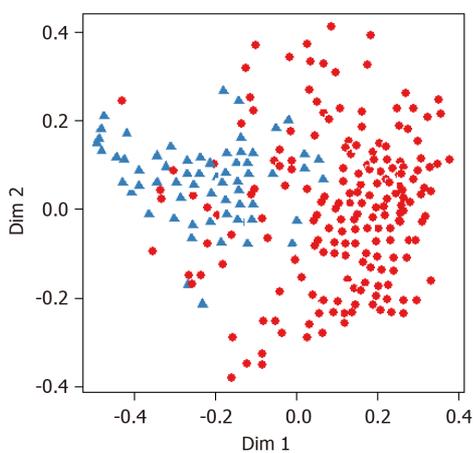


Figure 5 Multidimensional scaling plot of random forest (blue = early-onset Parkinson dementia and red = Parkinson's disease)

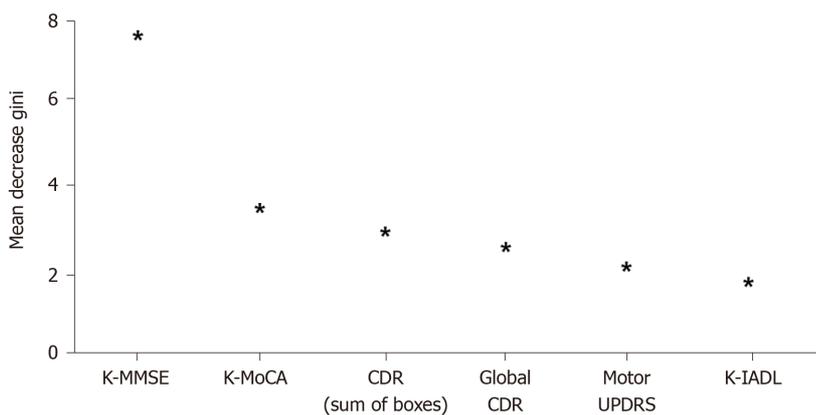


Figure 6 Importance of variables in the random forest-based early-onset Parkinson dementia prediction model (only the top six are presented). K-MMSE: Korean Mini Mental State Examination; K-MoCA: Korean Montreal Cognitive Assessment; CDR: Clinical Dementia Rating; K-IADL: Korean Instrumental Activities of Daily Living; UPDRS: Score of Untitled Parkinson's Disease Rating.

and highly related to cognitive functions; and (4) Even though administration of PD medicine could affect the results of cognitive tests, it was not considered as an input variable of the prediction model. Therefore, it is expected that it will be possible to derive more clinically meaningful results when a prediction model is developed by including genetic information or biomarkers in addition to neuropsychological tests. Furthermore, since PD medicine influences the expression of behavioral symptoms

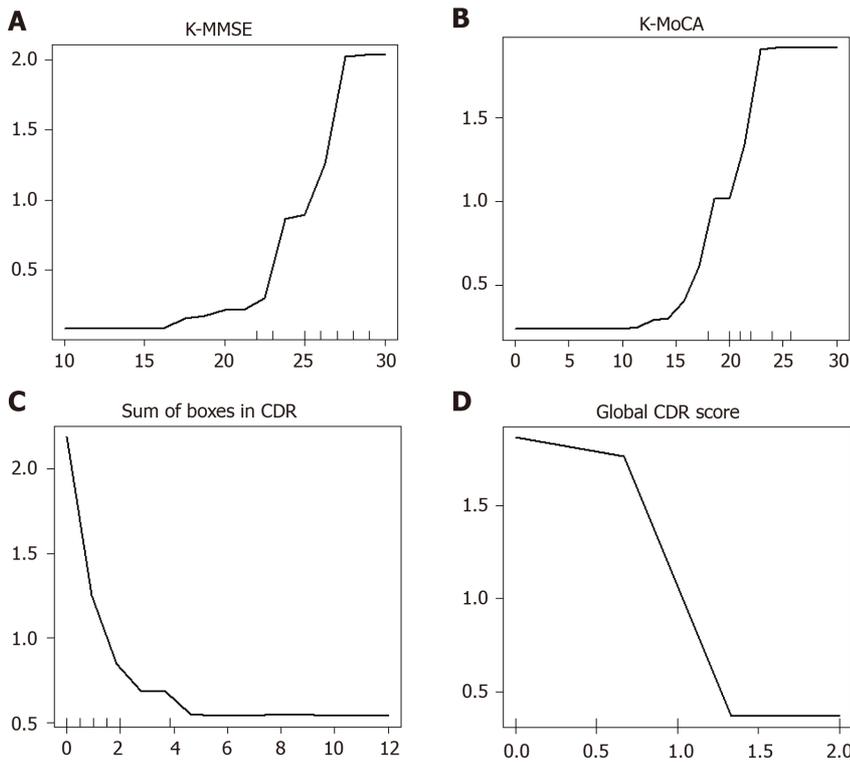


Figure 7 Partial dependence plot. K-MMSE: Korean Mini Mental State Examination; K-MoCA: Korean Montreal Cognitive Assessment; CDR: Clinical Dementia Rating.

and cognitive symptoms, it is necessary to investigate the application of it when developing an EOPD prediction model in the future.

CONCLUSION

It was found that Parkinson-related motor symptoms (*e.g.*, motor score of UPDRS) and instrumental daily performance (*e.g.*, K-IADL score), in addition to cognitive screening indicators (*e.g.*, K-MMSE score and K-MoCA score), were highly accurate predictors in EOPD prediction. Moreover, the accuracy of RF was higher than that of both the naive Bayesian model and discriminant analysis. This study showed the need for a customized screening test that can detect EOPD early using biomarkers or genetic big data.

ARTICLE HIGHLIGHTS

Research background

Despite the frequent progression from Parkinson's disease (PD) to Parkinson dementia, the basis to diagnose early-onset Parkinson dementia (EOPD) in the early stage is still insufficient.

Research motivation

It is limited to develop a highly-reliable model to predict EOPD using individual indicators such as PD symptoms and neuropsychological tests. In order to develop an accurate prediction model, it is necessary to develop a comprehensive model that includes sociodemographic indices, Parkinson's motor symptoms, Parkinson's non-motor symptoms, rapid eye movement (REM) sleep behavior disorder, and neuropsychological indices.

Research objectives

The objectives of our study were to explore the prediction accuracy of sociodemographic factors, Parkinson's motor symptoms, Parkinson's non-motor

symptoms, and REM sleep disorder for diagnosing EOPD using PD multicenter registry data.

Research methods

This study was performed by analyzing the Parkinson's Disease Epidemiology multicenter registry data provided by the National Biobank of Korea. This study analyzed 342 Parkinson patients (66 EOPD patients and 276 PD patients with normal cognition, younger than 65 years). The EOPD prediction model was developed using a random forest algorithm and the accuracy of the developed model was compared with the naive Bayesian model and discriminant analysis.

Research results

When the factors of EOPD were compared using "normalized importance of variables", the Korean Mini Mental State Examination score was the most important factor of EOPD. Also, the accuracy of random decision forest was higher than that of naive Bayesian model and that of discriminant analysis.

Research conclusions

It is believed that using random forest will increase accuracy while exploring major variables allowing us to predict EOPD, compared to traditional statistical techniques such as discriminant analysis.

Research perspectives

It is necessary to develop a customized screening test that can early detect EOPD using biomarkers or genetic big data.

ACKNOWLEDGEMENTS

The authors wish to thank the NB-Korea for providing the raw data.

REFERENCES

- 1 **Meagher DJ**, O'Connell H, Leonard M, Williams O, Awan F, Exton C, Tenorio M, O'Connor M, Dunne CP, Cullen W, McFarland J, Adamis D. Comparison of novel tools with traditional cognitive tests in detecting delirium in elderly medical patients. *World J Psychiatry* 2020; **10**: 46-58 [PMID: 32399398 DOI: 10.5498/wjp.v10.i4.46]
- 2 **Chiu M**, Wesson V, Sadavoy J. Improving caregiving competence, stress coping, and mental well-being in informal dementia carers. *World J Psychiatry* 2013; **3**: 65-73 [PMID: 24255878 DOI: 10.5498/wjp.v3.i3.65]
- 3 **Kim DW**. Clinical Research Center for Dementia of South Korea (CREDOS) (or CRCD) (CREDOS). In: ClinicalTrials.gov [Internet]. U.S. National Library of Medicine. Available from: <https://clinicaltrials.gov/ct2/show/NCT01198093> ClinicalTrials.gov Identifier: NCT01198093
- 4 **Wing Chi Au L**, Wong A, Abrigo J, Yuen YP, Yim Lung Leung E, Chung Tong Mok V. Early-onset dementia in Chinese: Demographic and etiologic characteristics. *Neurology Asia* 2019; **24**: 139-146 Available from: [https://www.neurology-asia.org/articles/neuroasia-2019-24\(2\)-139](https://www.neurology-asia.org/articles/neuroasia-2019-24(2)-139)
- 5 **Marceaux JC**, Soble JR, O'Rourke JF, Swan AA, Wells M, Amuan M, Sagiraju HKR, Eapen BC, Pugh MJ. Validity of early-onset dementia diagnoses in VA electronic medical record administrative data. *Clin Neuropsychol* 2020; **34**: 1175-1189 [PMID: 31645200 DOI: 10.1080/13854046.2019.1679889]
- 6 **Koedam EL**, Lauffer V, van der Vlies AE, van der Flier WM, Scheltens P, Pijnenburg YA. Early-versus late-onset Alzheimer's disease: more than age alone. *J Alzheimers Dis* 2010; **19**: 1401-1408 [PMID: 20061618 DOI: 10.3233/JAD-2010-1337]
- 7 **Koss E**, Edland S, Fillenbaum G, Mohs R, Clark C, Galasko D, Morris JC. Clinical and neuropsychological differences between patients with earlier and later onset of Alzheimer's disease: A CERAD analysis, Part XII. *Neurology* 1996; **46**: 136-141 [PMID: 8559362 DOI: 10.1212/wnl.46.1.136]
- 8 **Hvidsten L**, Engedal K, Selbæk G, Wyller TB, Bruvik F, Kersten H. Quality of Life in People with Young-Onset Alzheimer's Dementia and Frontotemporal Dementia. *Dement Geriatr Cogn Disord* 2018; **45**: 91-104 [PMID: 29694972 DOI: 10.1159/000487263]
- 9 **Johnen A**, Pawlowski M, Duning T. Distinguishing neurocognitive deficits in adult patients with NP-C from early onset Alzheimer's dementia. *Orphanet J Rare Dis* 2018; **13**: 91 [PMID: 29871644 DOI: 10.1186/s13023-018-0833-3]
- 10 **Parnetti L**, Chiasserini D, Bellomo G, Giannandrea D, De Carlo C, Qureshi MM, Ardah MT, Varghese S, Bonanni L, Borroni B, Tambasco N, Eusebi P, Rossi A, Onofri M, Padovani A, Calabresi P, El-Agnaf O. Cerebrospinal fluid Tau/ α -synuclein ratio in Parkinson's disease and degenerative dementias. *Mov Disord* 2011; **26**: 1428-1435 [PMID: 21469206 DOI: 10.1002/mds.23670]
- 11 **Laws KR**, Irvine K, Gale TM. Sex differences in cognitive impairment in Alzheimer's disease. *World J Psychiatry* 2016; **6**: 54-65 [PMID: 27014598 DOI: 10.5498/wjp.v6.i1.54]

- 12 **Glatt SL**, Hubble JP, Lyons K, Paolo A, Tröster AI, Hassanein RE, Koller WC. Risk factors for dementia in Parkinson's disease: effect of education. *Neuroepidemiology* 1996; **15**: 20-25 [PMID: 8719045 DOI: 10.1159/000109885]
- 13 **Reid WG**, Hely MA, Morris JG, Broe GA, Adena M, Sullivan DJ, Williamson PM. A longitudinal of Parkinson's disease: clinical and neuropsychological correlates of dementia. *J Clin Neurosci* 1996; **3**: 327-333 [PMID: 18638897 DOI: 10.1016/s0967-5868(96)90028-4]
- 14 **Byeon H**. Is the Random Forest Algorithm Suitable for Predicting Parkinson's Disease with Mild Cognitive Impairment out of Parkinson's Disease with Normal Cognition? *Int J Environ Res Public Health* 2020; **17** [PMID: 32290134 DOI: 10.3390/ijerph17072594]
- 15 **Byeon H**. Application of Machine Learning Technique to Distinguish Parkinson's Disease Dementia and Alzheimer's Dementia: Predictive Power of Parkinson's Disease-Related Non-Motor Symptoms and Neuropsychological Profile. *J Pers Med* 2020; **10** [PMID: 32354187 DOI: 10.3390/jpm10020031]
- 16 **Alcalay RN**, Mejia-Santana H, Tang MX, Rakitin B, Rosado L, Ross B, Verbitsky M, Kisselev S, Louis ED, Comella CL, Colcher A, Jennings D, Nance MA, Bressman S, Scott WK, Tanner C, Mickel SF, Andrews HF, Waters CH, Fahn S, Cote LJ, Frucht SJ, Ford B, Rezak M, Novak K, Friedman JH, Pfeiffer R, Marsh L, Hiner B, Siderowf A, Ottman R, Clark LN, Marder KS, Caccappolo E. Self-report of cognitive impairment and mini-mental state examination performance in PRKN, LRRK2, and GBA carriers with early onset Parkinson's disease. *J Clin Exp Neuropsychol* 2010; **32**: 775-779 [PMID: 20182943 DOI: 10.1080/13803390903521018]
- 17 **Rosness TA**, Barca ML, Engedal K. Occurrence of depression and its correlates in early onset dementia patients. *Int J Geriatr Psychiatry* 2010; **25**: 704-711 [PMID: 20069586 DOI: 10.1002/gps.2411]
- 18 **Morales DA**, Vives-Gilabert Y, Gómez-Ansón B, Bengoetxea E, Larrañaga P, Bielza C, Pagonabarraga J, Kulisevsky J, Corcuera-Solano I, Delfino M. Predicting dementia development in Parkinson's disease using Bayesian network classifiers. *Psychiatry Res* 2013; **213**: 92-98 [PMID: 23149030 DOI: 10.1016/j.psychres.2012.06.001]
- 19 **Cai B**, Broder MS, Chang E, Yan T, Metz DC. Predictive factors associated with carcinoid syndrome in patients with gastrointestinal neuroendocrine tumors. *World J Gastroenterol* 2017; **23**: 7283-7291 [PMID: 29142475 DOI: 10.3748/wjg.v23.i40.7283]
- 20 **Byeon H**. Exploring the Predictors of Rapid Eye Movement Sleep Behavior Disorder for Parkinson's Disease Patients Using Classifier Ensemble. *Healthcare (Basel)* 2020; **8** [PMID: 32369941 DOI: 10.3390/healthcare8020121]
- 21 **Hughes AJ**, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* 1992; **55**: 181-184 [PMID: 1564476 DOI: 10.1136/jnnp.55.3.181]
- 22 **Dubois B**, Burn D, Goetz C, Aarsland D, Brown RG, Broe GA, Dickson D, Duyckaerts C, Cummings J, Gauthier S, Korczyn A, Lees A, Levy R, Litvan I, Mizuno Y, McKeith IG, Olanow CW, Poewe W, Sampaio C, Tolosa E, Emre M. Diagnostic procedures for Parkinson's disease dementia: recommendations from the movement disorder society task force. *Mov Disord* 2007; **22**: 2314-2324 [PMID: 18098298 DOI: 10.1002/mds.21844]
- 23 **Kang Y**, Na DL, Hahn S. A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients. *J Korean Neurol Assoc* 1997; **15**: 300-308
- 24 **Kang Y**, Park J, Yu KH, Lee BC. The validity of the Korean-Montreal Cognitive Assessment (K-MoCA) as a screening test for both MCI and VCI. Conference Abstract: The 20th Annual Rotman Research Institute Conference, The frontal lobes, Toronto, Canada, March 22-26, 2010 [DOI: 10.3389/conf.fnins.2010.14.00148]
- 25 **Cho MJ**, Bae JN, Suh GH, Hahn BJ, Kim JK, Lee DW, Kang MH. Validation of geriatric depression scale, Korean version (GDS) in the assessment of DSM-III-R major depression. *J Korean Neuropsychiatr Assoc* 1999; **38**: 48-63
- 26 **Choi SH**, Na DL, Lee BH, Hahn DS, Jeong JH, Yoon SJ, Yoo KH, Ha CK, Han IW. Estimating the validity of the Korean version of expanded clinical dementia rating (CDR) scale. *J Korean Neurol Assoc* 2001; **19**: 585-591
- 27 **Kang SJ**, Choi SH, Lee BH, Kwon JC, Na DL, Han SH. The reliability and validity of the Korean Instrumental Activities of Daily Living (K-IADL). *J Korean Neurol Assoc* 2002; **20**: 8-14
- 28 **Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease**. The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations. *Mov Disord* 2003; **18**: 738-750 [PMID: 12815652 DOI: 10.1002/mds.10473]
- 29 **Richards M**, Marder K, Cote L, Mayeux R. Interrater reliability of the Unified Parkinson's Disease Rating Scale motor examination. *Mov Disord* 1994; **9**: 89-91 [PMID: 8139610 DOI: 10.1002/mds.870090114]
- 30 **Hoehn MM**, Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology* 1967; **17**: 427-442 [PMID: 6067254 DOI: 10.1212/wnl.17.5.427]
- 31 **Schwab RS**, England ACJ. Projection technique for evaluating surgery in Parkinson's disease. In: Gillingham FJ, Donaldson IML, editors. Third Symposium on Parkinson's Disease. Edinburgh, Scotland: E & S Livingstone; 1969: 152-157
- 32 **Gnanalingham KK**, Byrne EJ, Thornton A, Sambrook MA, Bannister P. Motor and cognitive function in Lewy body dementia: comparison with Alzheimer's and Parkinson's diseases. *J Neurol Neurosurg Psychiatry* 1997; **62**: 243-252 [PMID: 9069479 DOI: 10.1136/jnnp.62.3.243]
- 33 **Bosboom JL**, Stoffers D, Wolters ECh. Cognitive dysfunction and dementia in Parkinson's disease. *J Neural Transm (Vienna)* 2004; **111**: 1303-1315 [PMID: 15480840 DOI: 10.1007/s00702-004-0168-1]
- 34 **Starkstein SE**, Sabe L, Petracca G, Chemerinski E, Kuzis G, Merello M, Leiguarda R. Neuropsychological and psychiatric differences between Alzheimer's disease and Parkinson's disease with dementia. *J Neurol Neurosurg Psychiatry* 1996; **61**: 381-387 [PMID: 8890777 DOI: 10.1136/jnnp.61.4.381]
- 35 **Byeon H**. A prediction model for mild cognitive impairment using random forests. *Int J Adv Comput Sci App* 2015; **6**: 8-12 [DOI: 10.14569/IJACSA.2015.061202]



Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: bpgoffice@wjgnet.com
Help Desk: <https://www.f6publishing.com/helpdesk>
<https://www.wjgnet.com>

