

# Mild cognitive impairment, dementia, and cognitive dysfunction screening using machine learning

Daehyuk Yim, Tae Young Yeo and  
Moon Ho Park 

## Abstract

**Objective:** To develop a machine learning algorithm to identify cognitive dysfunction based on neuropsychological screening test results.

**Methods:** This retrospective study included 955 participants: 341 participants with dementia (dementia), 333 participants with mild cognitive impairment (MCI), and 341 participants who were cognitively healthy. All participants underwent evaluations including the Mini-Mental State Examination and the Montreal Cognitive Assessment. Each participant's caregiver or informant was surveyed using the Korean Dementia Screening Questionnaire at the same visit. Different machine learning algorithms were applied, and their overall accuracies, Cohen's kappa, receiver operating characteristic curves, and areas under the curve (AUCs) were calculated.

**Results:** The overall screening accuracies for MCI, dementia, and cognitive dysfunction (MCI or dementia) using a machine learning algorithm were approximately 67.8% to 93.5%, 96.8% to 99.9%, and 75.8% to 99.9%, respectively. Their kappa statistics ranged from 0.351 to 1.000. The AUCs of the machine learning models were statistically superior to those of the competing screening model.

**Conclusion:** This study suggests that a machine learning algorithm can be used as a supportive tool in the screening of MCI, dementia, and cognitive dysfunction.

## Keywords

Machine learning, cognitive dysfunction, mild cognitive impairment, dementia, screening, diagnostic tool

Date received: 2 February 2020; accepted: 3 June 2020

---

Department of Neurology, Korea University Ansan Hospital, Ansan, Republic of Korea

---

## Corresponding author:

Moon Ho Park, Department of Neurology, Korea University Ansan Hospital, 123 JeokGeum-ro, DanWon-gu, Ansan-si, GyeongGi-do 15355, Republic of Korea.  
Email: [kmmse@naver.com](mailto:kmmse@naver.com)



## Introduction

Medical research involves the analysis of numerous factors, and the interactions between these factors should not be overlooked or ignored. However, classical statistical analyses are limited for evaluating these multifactorial complexities. Machine learning is an algorithm that can learn patterns from multifactorial complex data without relying on conventional statistical assumptions, and it plays an increasingly essential role in the field of medical research.<sup>1-4</sup>

Cognitive dysfunction has very diverse etiological factors, and their interactions may contribute to its pathophysiological complexity. Mild cognitive impairment (MCI) and dementia are the most representative neurodegenerative disorders of cognitive dysfunction. Cognitive dysfunction has been evaluated using various clinical data, including data from patient evaluations, caregiver interviews, and other clinical evaluations. Because cognitive dysfunction has multiple causal factors, clinical data and the numerous interactions among these factors should be considered.

The Mini-Mental State Examination (MMSE)<sup>5-8</sup> and Montreal Cognitive Assessment (MoCA)<sup>8-10</sup> are neuropsychological screening tests that are widely used to evaluate patients. Additionally, the Korean Dementia Screening Questionnaire (KDSQ)<sup>8,11-13</sup> is commonly used in caregiver interviews in Korea. In the present study, a prediction model was developed to help screen for cognitive dysfunction. This model included patient evaluations (MMSE and MoCA scores), caregiver or informant interviews (KDSQ results), and clinical evaluations (including basic demographic data). To evaluate the usefulness of the prediction model using a machine learning algorithm, we constructed and compared these prediction models. If we can apply machine learning to cognitive

dysfunction to identify features that are not revealed using classical statistical methods, we will demonstrate the potential and utility of applying machine learning to the clinical field.

## Materials and methods

### Participants

This was a retrospective cross-sectional study of consecutive patients who visited a memory clinic at a university hospital in the Republic of Korea and were referred for neuropsychological screening. We analyzed participants with dementia (dementia), participants with mild cognitive impairment (MCI), and participants who were cognitively healthy (controls). A consensus diagnosis was determined using the standardized clinical criteria for MCI<sup>14</sup> and dementia.<sup>15</sup> MCI and dementia subtypes were not analyzed in this study. Controls did not meet the criteria for MCI or dementia but were recruited and assessed in a manner identical to that used for the patients with MCI and dementia.<sup>12</sup> The controls were cognitively and functionally normal, independent, and fulfilled the health-screening exclusion criteria.<sup>16</sup>

The consensus diagnoses of a geriatric physician and neuropsychologist were used to determine each subject's clinical status on the basis of clinical evaluations. The exclusion criteria included preexisting conditions that might affect participants' cognitive performance, such as intellectual disability, drug or substance abuse, and severe psychiatric illness. All participants who were accompanied by a caregiver or informant were included. The informants were participants' spouses or relatives who lived in the same household and had no psychiatric or neurological disease themselves. Each informant needed to see the participant at least 3 days per week for more than 4 hours per visit to ensure that

they had an adequate understanding of the participant's condition.

All participants underwent the MMSE, MoCA, and KDSQ examinations on the same day. The results of the MMSE, MoCA, and KDSQ were not available during the consensus diagnosis.

### *Neuropsychological screening tests*

The MMSE, MoCA, and KDSQ were administered as neuropsychological screening tests in this study. The MMSE is the most commonly used test for the screening of cognitive impairment and can be performed in a relatively short time.<sup>5,6</sup> Possible scores range from 0 to 30 points, where higher scores indicate better cognitive function. The MMSE is the most appropriate test for detecting moderate and severe cognitive dysfunction.<sup>6-8</sup> The MMSE examines the following six cognitive domains: orientation in time, orientation in place, memory registration, memory recall, attention and calculation, and language and visuospatial function.

The MoCA is the most widely used screening test for cognitive dysfunction, including MCI and the early stages of dementia.<sup>8-10</sup> The MoCA has higher sensitivity than the MMSE for detecting early-stage cognitive decline.<sup>8,9,17,18</sup> It evaluates visuospatial, naming, attention, language, abstract, memory, and orientation abilities. Possible scores range from 0 to 30 points, where higher scores indicate better cognitive function.

The KDSQ is an informant-based questionnaire that addresses changes in cognitive performance over the previous year in older people.<sup>11</sup> The KDSQ has been widely used in Korea because of its ease of use and culturally specific adaptation, and it has a high validity and reliability for dementia screening in older people.<sup>8,11-13</sup> The KDSQ assesses memory function, other cognitive functions, and the ability to

perform complex tasks in daily life. The KDSQ contains 15 items, each rated on a three-point scale: 0 (no change), 1 (sometimes/occasional change), and 2 (often/frequent change); a higher score indicates a poorer function.

### *Other clinical evaluations*

Demographic data (age and sex) and information regarding years of education were collected from participants and informants. All participants were evaluated based on their medical history, physical and neurological examination results, laboratory test results, brain imaging findings, and a neuropsychological battery. The neuropsychological battery was used with the Korean version of the assessment packet developed by the Consortium to Establish a Registry for Alzheimer's Disease.<sup>19</sup>

### *Statistical analysis*

The screening model was intended to identify the ability to screen for cognitive dysfunction when given only basic information, including age, sex, and education level. This model represents the accuracy of prediction when there is no information from neuropsychological screening tests, and was used as a comparison criterion for the other models. The screening model was evaluated using binary logistic regression (LR).

The machine learning models were evaluated using data from neuropsychological screening tests, including the MMSE, MoCA, and KDSQ, as well as basic information. The interaction algorithm or patterns among neuropsychological screening tests, including each subtest score as well as the total scores, were included in the analysis. The machine learning models were applied using several algorithms, including LR, penalized binary logistic regression (PLR), linear support vector machine (LSVM), linear discriminant

analysis (LDA), decision tree (DT), radial basis function kernel support vector machine (rSVM), random forest (RF), gradient boosting (GBM), and neural network (NN).

To verify each model, participants were reclassified into two groups for the application of all algorithms. Discriminations that were frequently encountered in clinical practice were evaluated, such as MCI from control, dementia from control, and cognitive dysfunction (MCI + dementia) from control.

The data were divided into a training dataset that was used for model construction, and a test dataset that was used to evaluate prediction performance. Both the training and test datasets were constructed such that they had the same rate of cognitive status. Each algorithm used repeated iterative cross-validation to determine the hyperparameter that determined the efficiency of the learning process, which used the minimized error in a prediction model. The prediction performance, which was calculated by the training dataset, was tested using the test dataset.

To measure performance, the overall accuracies and kappa values (Cohen's kappa) were evaluated. Overall accuracy was expressed as the percentage agreement with cognitive status, and was used to represent basic reliability. The kappa value could be used to correct the unbalanced distribution of the two groups, and represented moderate agreement if it was  $\geq 0.4$ . Receiver operating characteristic (ROC) curves and the area under the ROC curve (AUC) were used to evaluate the ability to discriminate between groups. Pairwise comparisons of AUCs were performed to assess the statistical significance of the difference between each pair of AUCs.<sup>20</sup> Because pairwise comparisons were performed in 36 pairs, a Bonferroni correction was used, and  $P < 0.001$  was considered significant.

Values were presented as the mean (standard deviation) or number (percentage)

unless otherwise indicated. Statistical tests were two-tailed, and  $\alpha$  was set at 0.05. R version 3.4.4 (www.r-project.org) and its suitable packages were used to perform all statistical analyses and modeling in this study.<sup>21,22</sup> The R packages 'caret', 'glmnet', 'randomForest', and 'gbm' were used to analyze the machine learning models.

### *Ethics approval*

This study was approved by the Institutional Review Board of Korea University Ansan Hospital (2018AS0187). The requirement to obtain informed consent was waived for the following reasons: this was a retrospective review and all participant records were anonymized and de-identified prior to analysis. The waiver is not inconsistent with the national law, and the research involved no more than minimal risk to the participants. The research could not practicably be performed without the waiver or alteration, and the waiver or alteration does not adversely affect the rights and welfare of the participants. No data in this paper reveal the identity of the participants.

### **Results**

The demographic and clinical characteristics of the study population are summarized in Table 1. A total of 955 participants were analyzed: 341 participants in the dementia group, 333 participants in the MCI group, and 341 participants in the control group. The mean age of the subjects was 70.13 ( $\pm 10.32$ ) years, and the mean length of education was 7.78 ( $\pm 4.86$ ) years. Fifty-eight percent of participants were women. There were no significant differences in age, sex, or level of education between the MCI, dementia, and control groups.

The results of the neurocognitive screening tests, including the MMSE, MoCA, and KDSQ, are summarized in Table 2. The mean total scores of the MMSE, MoCA,

**Table 1.** Demographic and clinical characteristics.

	Cognitive dysfunction			Statistics	P-value
	MCI ( <i>n</i> = 333)	Dementia ( <i>n</i> = 341)	Control ( <i>n</i> = 281)		
Age (years)	70.43 (9.36)	70.45 (11.32)	69.41 (10.12)	$H(2) = 2.62$	0.270
Sex (female)	191 (57.4%)	197 (57.8%)	166 (59.1%)	$\chi^2 = 0.20$ ( <i>df</i> = 2)	0.906
Education (years)	7.56 (4.68)	7.55 (5.04)	8.33 (4.80)	$H(2) = 4.27$	0.118

Values are presented as the mean (standard deviation) or number (percentage).

**Table 2.** The sub-scores of neuropsychological screening tests and their comparison between groups.

	Cognitive dysfunction			Cohen's <i>d</i>	P-value
	MCI ( <i>n</i> = 333)	Dementia ( <i>n</i> = 341)	Control ( <i>n</i> = 281)		
<b>MMSE</b>					
Total	24.84 (3.75)	18.20 (12.11)	26.94 (2.44)	0.192	<0.001
Time	4.38 (0.94)	2.47 (1.47)	4.76 (0.53)	0.468	<0.001
Place	4.83 (0.44)	3.95 (1.18)	4.88 (0.32)	0.240	<0.001
Memory	2.99 (0.09)	2.80 (0.57)	3.00 (0.00)	0.071	<0.001
Attention	3.21 (1.53)	1.51 (1.45)	3.90 (1.18)	0.340	<0.001
Recall	1.65 (0.95)	0.59 (0.80)	2.11 (0.83)	0.354	<0.001
Others	7.80 (1.36)	6.31 (1.99)	8.28 (0.95)	0.235	<0.001
<b>MoCA</b>					
Total	19.41 (5.67)	11.17 (5.80)	23.93 (4.13)	0.496	<0.001
Visuospatial	2.92 (1.42)	1.55 (1.04)	3.90 (1.24)	0.375	<0.001
Naming	2.48 (0.86)	1.78 (1.19)	2.94 (0.24)	0.225	<0.001
Attention	4.33 (1.68)	2.42 (1.96)	5.20 (1.11)	0.333	<0.001
Language	1.72 (0.93)	1.11 (0.90)	2.29 (0.83)	0.220	<0.001
Abstract	1.18 (0.73)	0.68 (0.73)	1.42 (0.70)	0.153	<0.001
Memory	1.47 (1.45)	0.43 (0.89)	2.45 (1.54)	0.278	<0.001
Orientation	5.33 (1.02)	3.21 (1.66)	5.74 (0.60)	0.462	<0.001
<b>KDSQ</b>	5.00 (3.64)	15.77 (7.52)	2.58 (2.58)	0.555	<0.001

MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; KDSQ, Korea Dementia Screening Questionnaire.

and KDSQ were 23.08 (8.54), 17.80 (7.48), and 8.13 (7.75), respectively. The total scores and subscores of the MMSE and MoCA and total score of the KDSQ were significantly different between the control, MCI, and dementia groups. Cohen's *d* ranged from 0.071 to 0.555. The total KDSQ score ( $d=0.555$ ) had the largest effect size, and the memory subscore of

the MMSE ( $d=0.071$ ) had the smallest effect size.

Using the screening model that included age, sex, and level of education, the overall accuracies for MCI, dementia, and cognitive dysfunction were 55.5%, 55.6%, and 70.6%, respectively. The overall accuracies of the machine learning models for MCI, dementia, and cognitive dysfunction were

**Table 3.** The performance of screening models created using different machine learning algorithms.

Machine learning algorithms	MCI vs. control	Dementia vs. control	Cog dys vs. control
	Overall accuracy (kappa)	Overall accuracy (kappa)	Overall accuracy (kappa)
Screening model	55.5% (0.051)	55.6% (0.030)	70.6% (0.000)
Logistic regression	68.2% (0.359)	97.2% (0.943)	82.2% (0.551)
Penalized logistic regression	72.7% (0.463)	96.8% (0.935)	81.9% (0.590)
Linear SVM	67.8% (0.351)	97.2% (0.943)	80.8% (0.522)
Linear discriminant analysis	70.6% (0.412)	95.2% (0.903)	82.7% (0.576)
Decision tree	75.1% (0.499)	97.2% (0.943)	75.8% (0.637)
Radial basis function kernel SVM	72.2% (0.436)	97.6% (0.951)	82.4% (0.552)
Random forest	80.8% (0.618)	97.2% (0.943)	89.2% (0.747)
Gradient boosting	93.5% (0.869)	99.9% (1.000)	95.5% (0.891)
Neural network	76.7% (0.530)	97.6% (0.951)	85.6% (0.640)

Accuracy is presented as a percentage with Cohen's kappa in parentheses. MCI, mild cognitive impairment; Cog dys, cognitive dysfunction; SVM, support vector machine.

approximately 67.8% to 93.5%, 96.8% to 99.9%, and 75.8% to 95.5%, respectively (Table 3). Their kappa statistics ranged from 0.351 to 1.000.

The ROC curves and their AUCs were compared between the screening model and the machine learning models. All machine learning models were significantly different from the screening model (Figure 1). The AUCs of RF and GB were significantly larger than those of the other machine learning models (Figure 1). The GB model had the highest AUCs for MCI, dementia, and cognitive dysfunction.

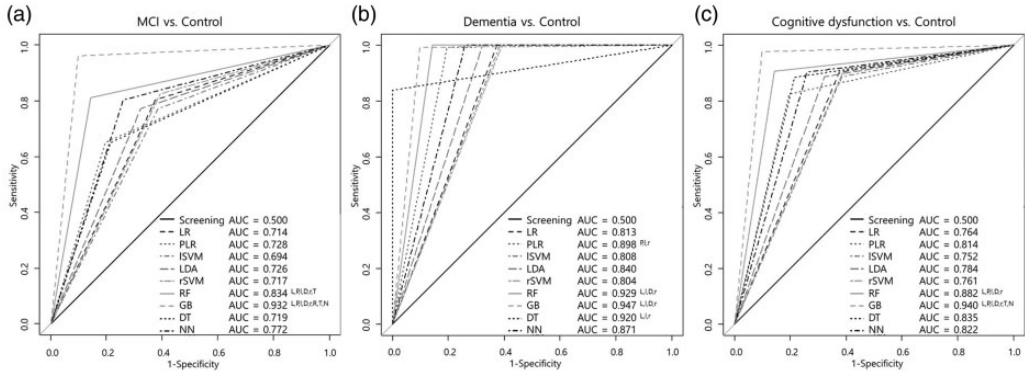
## Discussion

In the present study, we applied machine learning models to neuropsychological screening tests that are widely used in clinical practice, and constructed a hypothetical model to screen for MCI, dementia, and cognitive dysfunction. To effectively use clinical data from neuropsychological screening tests, our results suggest that machine learning models that can extract complex patterns from clinical data are useful. Using the proposed machine

learning models, the highest overall accuracy for screening was 93.5% for MCI, 99.9% for dementia, and 95.5% for cognitive dysfunction.

Cognitive dysfunction has been clinically evaluated using medical history, neurological examination results, and biomarkers, including brain imaging findings. Recently, other biomarkers, such as amyloid beta or tau, have been suggested to be useful for evaluating cognitive dysfunction, but data for such biomarkers are not easily accessible. In current clinical practice, neuropsychological screening tests are performed in most clinics. However, difficulties in the clinical interpretation of neuropsychological data have increased the need for computational techniques.

In machine learning approaches, adequate feature selection is an important task for creating a classification model that can successfully interpret data, with reduced variance and improved classification accuracy.<sup>23</sup> The present study evaluated the application of machine learning models to neuropsychological screening test results and basic demographic information. The data used in this



**Figure 1.** ROC curves for the screening of each cognitive dysfunction according to different machine learning methods. Comparison of the power of the ROC curve of different machine learning models in predicting (a) MCI versus control, (b) dementia versus control, and (c) cognitive dysfunction versus control. Using different line styles, the AUCs of the different machine learning models are presented as values. Superscript letters indicating the first letter of each machine learning method's name (or second letter, in the case of LDA [D] and DT [T]) show that the AUCs of RF and GB are significantly higher than those of other machine learning methods ( $P < 0.001$ ). ROC, receiver operating characteristic; MCI, mild cognitive impairment; AUC, area under the ROC curve; LR, binary logistic regression; PLR, penalized binary logistic regression; ISVM, linear support vector machine; LDA, linear discriminant analysis; rSVM, radial basis function kernel support vector machine; RF, random forest; GB, gradient boosting; DT, decision tree; NN, neural network.

study were variables obtained from the MMSE, MoCA, and KDSQ, which are widely used in clinical practice.<sup>5–13</sup> The MMSE was developed as a brief screening tool to provide a quantitative assessment of cognitive dysfunction and is one of the most frequently used bedside screening tools for dementia.<sup>5,6</sup> The MoCA is the most widely used screening test for MCI and early-stage dementia and is another very frequently used bedside screening tool for cognitive dysfunction.<sup>8–10</sup> The KDSQ is a dementia screening questionnaire that is completed by informants or caregivers and can screen for both the early and late stages of dementia.<sup>8,11–13</sup> The tools used in the present study can therefore screen for conditions ranging from MCI or early-stage dementia to late-stage dementia.

The neuropsychological screening tools in this study adequately cover different cognitive domains. Traditionally, neuropsychological assessment is performed to

examine several cognitive domains, including orientation, attention, executive function, visuospatial ability, language, and memory.<sup>24</sup> The MMSE covers multiple cognitive domains, such as orientation, memory registration and recall, attention and calculation, and language and visuospatial function.<sup>5,6</sup> The MoCA also covers multiple cognitive domains, such as short-term memory, visuospatial abilities, executive function, attention/concentration and working memory, language, and orientation.<sup>8–10</sup> The KDSQ consists of three areas, comprising memory function, other cognitive functions, and instrumental activities of daily living.<sup>8,11–13</sup>

In the current study, we demonstrated that the proposed machine learning models were more accurate than the screening model, which indicates the advantage of using machine learning algorithms to find patterns by simultaneously considering several variables. Because machine learning

considers multidimensional interactions between variables, it does not need to summarize a large number of variables and is not bound by the limitation of a single verification using each variable.<sup>1-4</sup> Machine learning models were originally designed to analyze large, complex datasets. Thus, machine learning has become a useful methodology for processing vast amounts of data that have already been obtained in medical research, and is now actively used in research into neurodegenerative diseases.<sup>25-27</sup> The present study revealed that our proposed machine learning methods had high overall accuracies, indicating that they are appropriate for screening cognitive dysfunction. Additionally, some machine learning models, like GB and RF, had excellent predictive performance. Moreover, this study demonstrated that these machine learning models can screen not only for cognitive dysfunction, but also for MCI and dementia, which will be very useful in daily clinical practice.

Several limitations should be noted in our study. First, it was subject to all of the limitations inherent to a retrospective study design. In addition, there may be some degree of selection bias in this retrospective study. A prospective study is therefore warranted to validate our results. Second, the risk of overfitting is higher when the sample size is small, as in the present study. Although we used both training and test datasets, it should be noted that the accuracy of machine learning algorithms may be inflated when a complex algorithm is used with a small sample size. Similarly, if the machine learning models become too complex, the risk of overfitting increases, while the ability to generalize the models to new data decreases.<sup>28</sup> Third, the proposed model is only applicable for differentiating cognitive status, and cannot predict temporal stage or prognosis. Fourth, this study evaluated only basic information, including age, sex, and education level, as the real primary clinical field. However,

other comorbidities and laboratory parameters can affect cognitive function.<sup>29</sup> Thus, a diagnostic test rather than a screening test should be analyzed with these variables. Fifth, the scope of this study is limited because it evaluated data from a single study center. The generalizability of a hypothetical model depends on the inclusiveness of the training dataset. That is, a hypothetical model that was trained using data from a single center or a small sample size may not be suitable to apply to subjects from another population or to a large sample. A useful model therefore requires data collected from a range of subjects among a large sample size.

## **Conclusions**

When screening for cognitive dysfunction, difficulties in the clinical interpretation of neuropsychological data have prompted the use of machine learning tools. The present study suggests that a machine-learning-based approach can be a valuable tool to screen for cognitive dysfunction. To increase the applicability of the hypothetical model, so it can then be used as a clinical tool to screen for cognitive status, it is necessary to validate the model using a large comprehensive dataset. Furthermore, even if the analysis of neuropsychological data gives significant results, additional data from biomarkers, positron emission tomography, magnetic resonance imaging (MRI), functional MRI, or genetic analysis may also contribute to screening and diagnosis.

## **Declaration of conflicting interest**

The authors declare that there is no conflict of interest.

## **Funding**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.



## ORCID iD

Moon Ho Park  <https://orcid.org/0000-0002-4892-3475>

## References

- Handelman GS, Kok HK, Chandra RV, et al. eDoctor: machine learning and the future of medicine. *J Intern Med* 2018; 284: 603–619.
- Deo RC. Machine learning in medicine. *Circulation* 2015; 132: 1920–1930.
- Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev* 1959; 3: 210–229.
- Ahmed MR, Zhang Y, Feng Z, et al. Neuroimaging and machine learning for dementia diagnosis: recent advancements and future prospects. *IEEE Rev Biomed Eng* 2019; 12: 19–33.
- Folstein MF, Folstein SE and McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; 12: 189–198.
- Han C, Jo SA, Jo I, et al. An adaptation of the Korean mini-mental state examination (K-MMSE) in elderly Koreans: demographic influence and population-based norms (the AGE study). *Arch Gerontol Geriatr* 2008; 47: 302–310.
- Trivedi D. Cochrane review summary: Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Prim Health Care Res Dev* 2017; 18: 527–528.
- Rosli R, Tan MP, Gray WK, et al. Cognitive assessment tools in Asia: a systematic review. *Int Psychogeriatr* 2016; 28: 189–210.
- Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005; 53: 695–699.
- Kang Y, Park J, Yu KH, et al. A reliability, validity, and normative study of the Korean-Montreal Cognitive Assessment (K-MoCA) as an instrument for screening of vascular cognitive impairment (VCI). *Korean J Clin Psychol* 2009; 28: 549–562.
- Yang DW, Cho BL, Chey JY, et al. The development and validation of Korean dementia screening questionnaire (KDSQ). *J Korean Neurol Assoc* 2002; 20: 135–141.
- Yeo TY, Kim K and Park MH. Conversion from the Korean dementia screening questionnaire to the informant questionnaire on cognitive decline in the elderly. *Neuropsychiatr Dis Treat* 2019; 15: 3205–3211.
- Lee SJ, Han JH, Hwang JW, et al. Screening for normal cognition, mild cognitive impairment, and dementia with the Korean dementia screening questionnaire. *Psychiatry Investig* 2018; 15: 384–389.
- Petersen RC, Doody R, Kurz A, et al. Current concepts in mild cognitive impairment. *Arch Neurol* 2001; 58: 1985–1992.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders- Fourth Edition, text revision*. Washington, DC: American Psychiatric Press, 2000: 133–155.
- Christensen KJ, Multhaup KS, Nordstrom S, et al. A cognitive battery for dementia: development and measurement characteristics. *Psychol Assess* 1991; 3: 168–174.
- Hoops S, Nazem S, Siderowf AD, et al. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology* 2009; 73: 1738–1745.
- Freitas S, Simoes MR, Alves L, et al. Montreal Cognitive Assessment (MoCA): validation study for vascular dementia. *J Int Neuropsychol Soc* 2012; 18: 1031–1040.
- Lee JH, Lee KU, Lee DY, et al. Development of the Korean version of the Consortium to Establish a Registry for Alzheimer’s Disease Assessment Packet (CERAD-K): clinical and neuropsychological assessment batteries. *J Gerontol B Psychol Sci Soc Sci* 2002; 57: P47–P53.
- DeLong ER, DeLong DM and Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; 44: 837–845.
- R Core Team. *The R project for statistical computing*. [www.r-project.org](http://www.r-project.org) (accessed 02 February 2020).

22. Ihaka R and Gentleman R. R: a language for data analysis and graphics. *J Comput Graph Stat* 1996; 5: 299–314.
23. Saeys Y, Inza I and Larranaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics* 2007; 23: 2507–2517.
24. Salmon DP and Bondi MW. Neuropsychological assessment of dementia. *Annu Rev Psychol* 2009; 60: 257–282.
25. Battista P, Salvatore C and Castiglioni I. Optimizing neuropsychological assessments for cognitive, behavioral, and functional impairment classification: a machine learning study. *Behav Neurol* 2017; 2017: 1850909.
26. Gurevich P, Stuke H, Kastrup A, et al. Neuropsychological testing and machine learning distinguish Alzheimer's disease from other causes for cognitive impairment. *Front Aging Neurosci* 2017; 9: 114.
27. Valliani AA, Ranti D and Oermann EK. Deep learning and neurology: a systematic review. *Neurol Ther* 2019; 8: 351–365.
28. Foster KR, Koprowski R and Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research - commentary. *Biomed Eng Online* 2014; 13: 94.
29. Livingston G, Sommerlad A, Orgeta V, et al. Dementia prevention, intervention, and care. *Lancet* 2017; 390: 2673–2734.