Dement Neurocogn Disord. 2021 Oct;20(4):70-79
https://doi.org/10.12779/dnd.2021.20.4.70
pISSN 1738-1495·eISSN 2384-0757

DND Dementia and Neurocognitive Disorder

Original Article

Check for updates

Comparison of RCF Scoring System to Clinical Decision for the Rey Complex Figure Using Machine-Learning Algorithm

Chanda Simfukwe [1], **Seong Soo An** [2], **Young Chul Youn** [1]

[1]Department of Neurology, Chung-Ang University Hospital, Seoul, Korea
[2]Department of Bionano Technology, Gachon University, Seongnam, Korea

OPEN ACCESS

Received: Jun 29, 2021
Revised: Oct 7, 2021
Accepted: Oct 20, 2021

Correspondence to
Seong Soo An
Department of Bionano Technology, Gachon University, 1342 Seongnam-daero, Sujeong-gu, Seongnam 13120, Korea.
E-mail: seong.an@gmail.com

Young Chul Youn
Department of Neurology, Chung-Ang University Hospital, 102 Heukseok-ro, Dongjak-gu, Seoul 06973, Korea.
E-mail: neudoc@gmail.com

ORCID iDs
Chanda Simfukwe
https://orcid.org/0000-0003-2845-8016
Young Chul Youn
https://orcid.org/0000-0002-2742-1759

Funding
The research was funded by the Gachon University Research Fund of 2019 (GCU-2019-0815) and the National Research Foundation of Korea (NRF-2020R1A2B5B01002463).

Conflict of Interest
The authors have no financial conflicts of interest.

ABSTRACT

Background and Purpose: Interpreting the Rey complex figure (RCF) requires a standard RCF scoring system and clinical decision by clinicians. The interpretation of RCF using clinical decision by clinicians might not be accurate in the diagnosing of mild cognitive impairment (MCI) or dementia patients in comparison with the RCF scoring system. For this reason, a machine-learning algorithm was used to demonstrate that scoring RCF using clinical decision is not as accurate as of the RCF scoring system in predicting MCI or mild dementia patients from normal subjects.
Methods: The RCF dataset consisted of 2,232 subjects with formal neuropsychological assessments. The RCF dataset was classified into 2 datasets. The first dataset was to compare normal vs. abnormal and the second dataset was to compare normal vs. MCI vs. mild dementia. Models were trained using a convolutional neural network for machine learning. Receiver operating characteristic curves were used to compare the sensitivity, specificity, and area under the curve (AUC) of models.
Results: The trained model's accuracy for predicting cognitive states was 96% with the first dataset (normal vs. abnormal) and 88% with the second dataset (normal vs. MCI vs. mild dementia). The model had a sensitivity of 85% for detecting abnormal with an AUC of 0.847 with the first dataset. It had a sensitivity of 78% for detecting MCI or mild dementia with an AUC of 0.778 with the second dataset.
Conclusions: Based on this study, the RCF scoring system has the potential to present more accurate criteria than the clinical decision for distinguishing cognitive impairment among patients.

Keywords: Machine Learning; Rey-Osterrieth Complex Figure; Mild Cognitive Impairment; Dementia; Neuropsychological Test

INTRODUCTION

Neuropsychological tests can help obtain information on the functional integrity and structure of the human brain. These tests play a core role in accessing patients with mild cognitive impairment (MCI) and dementia.[1] MCI is an intermediate stage between the expected cognitive decline of normal aging and the serious decline of dementia. Patients with

Author Contributions
Conceptualization: Youn YC; Funding acquisition: An SS, Youn YC; Investigation: Youn YC; Methodology: Simfukwe C; Project administration: Youn YC; Resources: Youn YC; Supervision: An SS, Youn YC; Validation: Youn YC; Writing - original draft: Simfukwe C, An SS, Youn YC; Writing - review & editing: An SS, Youn YC.

MCI might have substantial limitations for daily activities.[2] Dementia is a clinical syndrome characterized by a progressive cognitive decline that can interfere with one's ability to function independently in daily activities.[3]

Rey complex figure (RCF) is a widely used neuropsychological instrument to evaluate the cognitive function and access constructional ability (**Fig. 1**).[4] RCF was first proposed by a Swiss psychologist Andre Rey in 1941. Paul-Alexandre Osterrieth further standardized the scoring system to an 18-point scale scoring system and added child norms in 1944.[5] RCF has been used as a neuropsychological test for visual perception and long-term visual memory both in clinical and different research environments. The order and accuracy in which the RCF is copied and drawn from the recall can provide information on the location and extent of neuropsychological disorder.[6] Studies using RCF have revealed visual memory disturbances in individuals with dementia. Poorer copying of a figure by a given patient in comparison with that by normal healthy controls could suggest Alzheimer's disease (AD).[7] It has been reported that RCF is one of neuropsychological tests with adequate sensitivity and specificity for discriminating patients with different neuropsychological conditions such

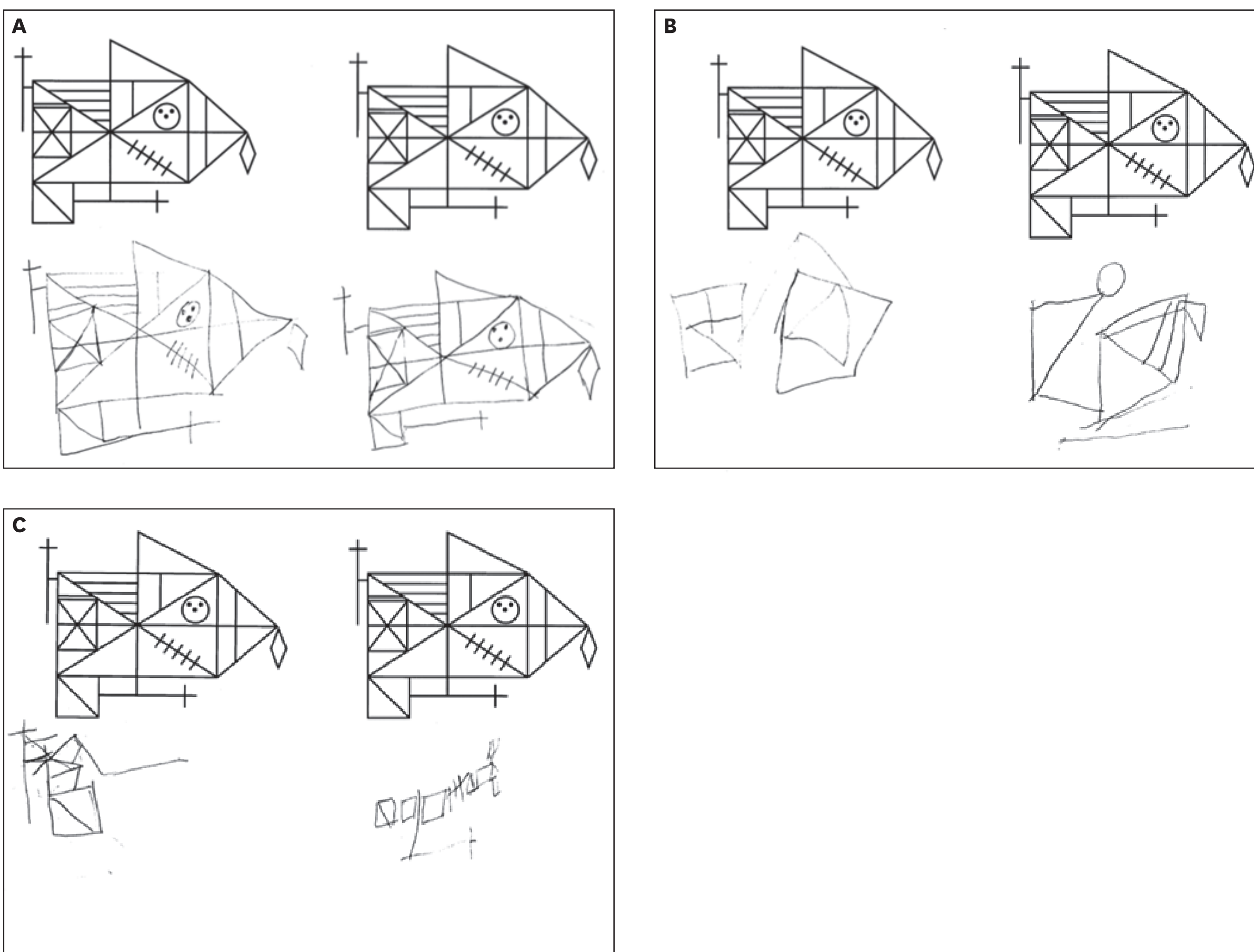

Fig. 1. Rey complex figure drawing samples for participants. (A) Normal subjects, (B) abnormal and mild cognitive impairment subjects, (C) mild dementia subjects.

 https://doi.org/10.12779/dnd.2021.20.4.70

Table 1. CDR scaling criteria

Characteristics	Scale
Normal	0
MCI	0.5
Mild dementia	1.0
Moderate dementia	2.0
Severe dementia	3.0

CDR: clinical dementia rating, MCI: mild cognitive impairment.

as dementia, MCI, depression, and mild head injury.[8] Many clinicians have used RCF for diagnosing MCI and dementia to improve their clinical decisions.

The RCF scoring system for participants can be performed in 2 ways: the standard RCF scoring system and clinical decision based on clinical dementia rating (CDR).[8,9] The CDR is evaluated using the Korean Screening Questionnaire (KDSQ) on a scale of 0–3 (**Table 1**).[10,11] The KDSQ is a dementia screening questionnaire that identifies early dementia patients.[11] In the 36-point version of the RCF scoring system, the figure is split into eighteen identifiable areas.[6,8] Each line category is considered separately and marked on the accuracy of its position and the exhibited distortion with scales shown in Appendix 1. The clinical decision was a major criterion for differentiating MCI and mild dementia subjects. The RCF score depends on the accuracy of the drawing by the participant. Clinical decision is based on the CDR scale that depends on KDSQ results of each participant. The RCF tool can only help inform clinicians with the diagnosis of MCI or mild dementia. The final diagnosis of patients with MCI or mild dementia is determined by clinicians.[9]

Machine-learning algorithms could analyze large and complex medical datasets.[12,13] Many machine-learning algorithms have been used in the detection, prediction, diagnosis, and classification of diseases, such as cancer and other chronic diseases.[14] Other studies have reported a fully automated scoring algorithm for RCF by comparing scoring results of the algorithm to results of manual scoring.[7] The algorithm consisting of a cascade machine learning model was trained on manual scores to extract 18 segments of the RCF. The algorithm performance was high, but not strictly equivalent to manual scoring. The previous study concluded that scoring the RCF using machine learning algorithm could save time in clinical practices. The objective of the current study was to compare the RCF scoring system with clinical decision using a machine-learning algorithm in the prediction of cognitive states.

METHODS

TensorFlow (https://www.tensorflow.org) was used to train machine learning models with an artificial neural network algorithm to distinguish subjects diagnosed as cognitively impaired from normal subjects. The first model consisted of normal vs. abnormal subjects. They were grouped according to the RCF scoring system. The second model consisted of normal vs. MCI vs. mild dementia groups based on the clinical decision from clinicians. The dataset was obtained from the Chung-Ang University Hospital (IRB No. C2012049[744]) Department of Neurology, South Korea. The convolutional neural network (CNN) in Tensorflow was implemented for machine learning. CNNs are the most common networks used with image classification developed for the machine-learning algorithm using Python language in Tensorflow, which could take in an assigned image input with learnable weights and biases and then combine both feature extraction and classification.[15] Tensorflow software was developed by Google with a library for machine learning with Python computer language.[16]

https://doi.org/10.12779/dnd.2021.20.4.70

Participants

The study was carried out with a total number of 2, 232 participants. It was a hospital-based cohort study to assess the occurrence and risk factors of cognitive disorders in the elderly. The dataset was composed of participants categorized into 5 diagnostic groups depending on the RCF scoring system and clinical decision. These groups were categorized as following. Normal and abnormal were based on the value of threshold cutoff point of 16% depending on the calculated average value from RCF copy test score (Appendix 1), age, and education level. Normal, MCI, and mild dementia groups were based on clinical decisions depending on the KDSQ.[11] Participants were grouped using CDR scale (**Table 1**).

Criteria for differentiating normal and abnormal groups were based on the value of threshold cutoff point of 16% averaged from the RCF copy test score, age, and education level of each participant from the study as follows: 1) participants with an average score of ≥16% were grouped as normal subjects; 2) participants with an average score of <16% were grouped as abnormal subjects. Criteria for differentiating normal, MCI, and mild dementia groups were based on the clinical decision using the CDR scale of participant from the study as follows: 1) those with CDR of 0 were grouped as normal subjects, 2) those with CDR of 0.5 were grouped as MCI subjects, 3) those with CDR of 1 were grouped as mild dementia subjects.

Subjects with medical history, such as territorial brain tumor, hydrocephalus, encephalitis, severe head trauma, and current or past neurological or psychiatric illnesses, such as schizophrenia and epilepsy, were excluded from the study. A total of 2,209 participants were selected and inputted for the machine-learning. The dataset was sorted according to the criteria for each class. Age, sex, and education level were also considered during group classifications of each subject.

Models training and statistical analyses

The first step of model training involved in organizing the dataset into 5 groups based on their criteria: (normal vs. abnormal) and (normal vs. MCI vs. mild dementia). The dataset was prepared for 2-way classification (normal vs. abnormal) and 3-way classification (normal vs. MCI vs. mild dementia) with respect to RCF scoring system and clinical decision. The dataset of normal vs. abnormal groups was used for the first model to predict abnormal from normal subjects. Normal vs. MCI vs. mild dementia groups were analyzed as a dataset for the second model to predict MCI and mild dementia from normal subjects. The MCI vs. mild dementia dataset was not evaluated for machine learning because this dataset was too small for training.

The current study was conducted to compare the scoring criteria between the RCF scoring system and clinical decision for their accuracy in predicting cognitively impaired patients from normal subjects through a machine-learning algorithm.

The second step of the model training involved splitting data and training machine learning models with an artificial neural network using CNN in Tensorflow on Colab cloud platform (https://colab.research.google.com/), which required to go through the following pre-processing steps. Data were imported from participants with '.png' format. The "validation_split" function from "tf.keras.preprocessing.image_dataset_from_directory" was used to randomly split data into training and test datasets. The train size was 0.70, which indicated the percentage of the data withheld for the training set. The validation dataset was comprised of the remaining 30%.

https://doi.org/10.12779/dnd.2021.20.4.70

The training dataset was used for further model training using CNN in TensorFlow. The trained model was then applied to an algorithm with Keras library. From Keras library, the "Keras. Models" function was loaded to the developed model from TensorFlow. The model then predicted subjects with abnormal, MCI, and mild dementia from normal subjects. The "NumPy_asarray" function was used to convert the RCF test image and applied in the algorithm using "one_hot encoding" to evaluate the accuracy of prediction of the model.

This artificial neural network consisted of 5 convolutional and max-pooling layers. A dropout layer was inserted before connecting it to a fully connected neural network. To improve the prediction accuracy and prevent overfitting, 2–3 out of 10 weights were connected to the next layer. Since datasets were relatively small for machine learning, ImageDataGenerator from Keras library was used for image augmentation of RCF images. ImageDataGenertor was an augmentation technique that could be used to artificially expand the size of the training dataset by creating modified versions of images in the dataset.[17] A dropout rate of 0.2–0.3 was used in the activation layer. The cost was minimized with the "adam" optimizer method and calculated using "Sparse_Categorical_Crossentropy". A batch size of 20 and 40–72 epochs was applied to model training. The optimal dropout rates and epochs were found and adjusted during the model training.

The model training in Tensorflow was performed once with the datasets, which were grouped according to their criteria. The fourth step involved calculating the sensitivity, specificity, and AUC of the model with the ROC curve by conducting logistics regression using TensorFlow (https://www.tensorflow.org). This process was performed for both the first dataset (normal vs. abnormal) and the second dataset (normal vs. MCI vs. mild dementia) separately.

Ethics statement

The ethics committee of the Chung-Ang University Hospital approved this study(IRB No. C2012049[744]).

RESULTS

After obtaining raw data from the Department of Neurology, these raw data seemed to be imbalanced. In the first dataset, there were 1,296 and 913 normal and abnormal cases, respectively. There were 447 males and 849 females in the normal group of subjects and 375 males and 538 females in the abnormal group of subjects. In the second dataset, were 1,649, 453, and 107 normal, MCI, and mild dementia subjects, respectively. Gender distributions in the second dataset were as follows: 606 males and 1,043 females in normal subjects; 179 males and 274 females in MCI subjects; and 59 males and 48 females in mild dementia subjects (**Tables 2** and **3**).

Demographics of the dataset

From a total number of 2,232, 2,209 subjects were allocated to training and validation datasets. The first model had a training dataset of 1,296 normal subjects and a validation dataset of 913 abnormal subjects. The second model had a training dataset of 800 normal subjects with 150 MCI and 40 mild dementia patients.

Based on the 36-point RCF scoring system, the accuracy of the first model for predicting normal and abnormal subjects was 96%. In the second model with clinical decision criteria based on CDR scale, the prediction accuracy for normal, MCI, and mild dementia was

https://doi.org/10.12779/dnd.2021.20.4.70

Table 2. Demographics of subjects included in this study and their classifications in the training and testing sets according to normal and abnormal group following the 36-point RCF scoring system

Characteristic	Normal*	Abnormal†
No. of subjects (n=2,209)	1,296	913
Age	71.27±9.65	72.53±10.24
Males (females)	447 (849)	375 (538)
Education	8.89±4.92	8.24±5.44
RCF score‡	57.60±21.78	2.95±4.25
Training	1,296	913
Testing	450	330

Data are shown as mean±standard deviation.
RCF: Rey complex figure.
*Normal, normal subjects with RCF score ≥16%; †Abnormal, abnormal subjects with RCF score <16%; ‡$p<0.05$.

Table 3. Demographics of subjects included in this study and their classification in the training and testing sets according to normal, MCI, and mild dementia following clinical decision based on the CDR scale

Characteristic	Normal*	MCI†	Mild dementia‡
No. of subjects (n=2,209)	1,649	453	107
Age	70.84±9.55	74.77±10.18	78.09±9.24
Males (females)	606 (1,043)	179 (274)	59 (48)
Education	8.99±5.04	7.32±5.27	6.78±5.18
Training	1,640	450	107
Testing	800	150	40

Data are shown as mean±standard deviation.
MCI: mild cognitive impairment, CDR: clinical dementia rating.
*Normal, normal subjects classified with the clinical decision based on CDR scale of 0; †MCI, MCI subjects classified with clinical decision based on CDR scale of 0.5; ‡Mild dementia, mild dementia subjects classified with the clinical decision based on CDR scale of 1.0.

88%. The first model had a sensitivity of 85% for detecting abnormal subjects and the second model had a sensitivity of 78% for detecting MCI and dementia cases. The receiver characteristic curve (ROC) of the 5 predictors of the 2 models is shown in **Fig. 2**. The first model (normal vs. abnormal) had a larger area under the curve (AUC) of 0.847 in comparison with the second model (normal vs. MCI vs. mild dementia) with an AUC of 0.778.

DISCUSSION

The goal of this study was to develop a machine-learning algorithm to differentiate the scoring method for RCF to achieve a higher accuracy in diagnosing cognitive state of a subject. The appropriate scoring criteria for RCF were verified as the standard scoring system since the accuracy for predicting normal and abnormal subjects was 96%. The validation dataset accuracy was at 90% in comparison with the clinical decision scoring system based on CDR scale by clinicians, which had an accuracy of 88% for predicting subjects with normal, MCI, and mild dementia with an accuracy of 85% using the validation dataset.

The first model revealed a higher sensitivity, a higher AUC, and a lower specificity in comparison to the second model. Results showed that the standard 36-point RCF scoring system was an appropriate scoring system for RCF test compared to clinical decision scoring method in the detection of cognitive states of the subject. Generally, screening tools would require a higher sensitivity in order not to miss patients with dementia even with lower specificity.[18] Several reports have suggested the implementation of RCF neuropsychological tests in assessing patients with AD. Previous study results have indicated significant differences between 381 normal subjects from clinical decision classification by CDR of 0 and 137 patients with AD classified with clinical decision of CDR of 0.5. All subjects could

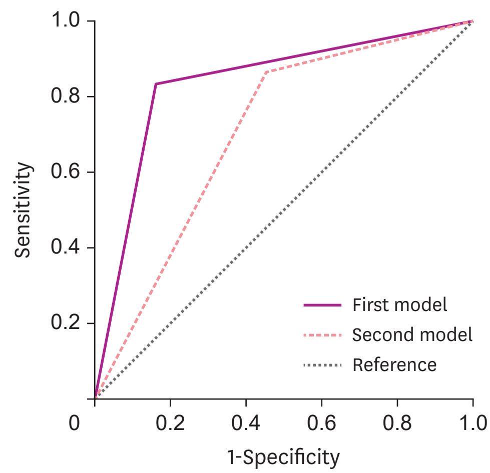

Fig. 2. ROC curves of 2 predictors' first model and second model. ROC curves were obtained by conducting a logistic regression using Tensorflow on the dataset of the first model (normal vs. abnormal) and the second model (normal vs. mild cognitive impairment vs. mild dementia).
ROC: receiver operating characteristic.

perform the full set of RCF tasks. Subjects with CDR 0 revealed a significant step-by-step learning effect of drawing the RCF, while subjects with CDR 0.5 did not show a significant step-by-step learning effect.[19] This result demonstrated that RCF could be as a useful cognitive tool to differentiate a cognitively impaired patient from a normal subject.

The classified datasets for training might not be enough. Overfitting occurred during the training. A data augmentation technique called ImageDataGenerator using the flow from the directory method was carried out to avoid overfitting. ImageDataGenertor was an augmentation technique for artificially expanding the size of the training dataset by creating modified versions of images in the dataset from the Keras library.[20] Keras is a great high-level library that could allow the creation of powerful machine-learning models.[21]

Previous studies that involved diagnosing MCI based on the CDR scale demonstrated that CDR scoring criteria could not accurately diagnose patients with MCI. Grundman et al.[22] have analyzed 769 patients with MCI diagnosed using clinical decision based on CDR scale, 107 cognitively normal elderly controls, 122 patients with very mild AD rated as having a CDR of 0.5, and 183 patients with mild AD rated as having a CDR of 1.0. Patients met operational criteria for amnestic MCI. Controls were individuals who had a CDR of 0. They reported that the mean of the CDR sum of boxes of the CDR 0.5 AD group (n=122) was 1.7 times higher than that for individuals in the MCI group (n=769) whom the doctor diagnosed without using the CDR scale.[22] These results indicate that the classification of patients depending on the clinical decision based on CDR might not be accurate in the detection of MCI. The current method of analysis with machine learning from the second model supported their observations.

There are also other limitations with global CDR scores for MCI diagnosis when we consider disease progression. When MCI participants with global CDR scores of 0.5 were divided into groups with relatively intact or impaired ratings on the ADL subscale of the CDR, MCI participants with impaired ADL were more likely than those with intact ADL to progress to a clinical diagnosis of probable AD within the next 2 years.[23] These findings indicate that individuals with global CDR scores of 0.5 can manifest heterogeneous states of cognitive impairment ranging from healthy elders to mild AD.

RCF as a screening tool with an RCF scoring system was easier to utilize in comparison to the clinical decision scoring system depending on the CDR scale. The clinical decision scoring system showed several limitations. For example, the clinician must be skilled and the interview should be conducted face to face in screening patients, which required much time and effort.[11] Therefore, the RCF scoring system for RCF as a screening tool might be more appropriate in primary care practice.

The purpose of this study was to evaluate which criteria between RCF scoring system and clinical decision for RCF might be more accurate in diagnosing abnormal, MCI, and dementia using a neural network algorithm for machine learning. Our results indicated that the first trained model (normal vs. abnormal) based on the RCF scoring system had high accuracy, sensitivity, and specificity in predicting abnormal subjects compared to the second trained model (normal vs. MCI vs. dementia) based on clinical decision in predicting MCI and dementia from normal subjects. Therefore, the machine learning trained model on RCF can be used as a screening tool by neuropsychologists for diagnosing abnormal, MCI, and dementia from normal subjects.

In the current study, the first model (normal vs. abnormal) categorized with RCF scoring system criteria could predict abnormal subjects from normal, but the second model (normal vs. MCI vs. mild dementia) categorized with clinical decision criteria did not predict the severity of MCI or dementia from normal subjects. The sensitivity and specificity of the first model were high than those of the second model in the prediction. These results suggest that clinical decisions might not be accurate in diagnosing MCI or mild dementia. For this reason, neuropsychologists and other researchers should use the RCF scoring system for diagnosing cognitive states of subjects. The machine learning models developed in this study could assist neuropsychological decisions in diagnosing cognitive states of subjects.

ACKNOWLEDGEMENTS

We would like to thank the Department of Neurology at Chung-Ang University Hospital for providing the necessary tools to make this research successful.

REFERENCES

1. Harvey PD. Clinical applications of neuropsychological assessment. Dialogues Clin Neurosci 2012;14:91-99.
PUBMED | CROSSREF
2. Petersen RC, Caracciolo B, Brayne C, Gauthier S, Jelic V, Fratiglioni L. Mild cognitive impairment: a concept in evolution. J Intern Med 2014;275:214-228.
PUBMED | CROSSREF
3. Duong S, Patel T, Chang F. Dementia: what pharmacists need to know. Can Pharm J (Ott) 2017;150:118-129.
PUBMED | CROSSREF
4. Shin MS, Park SY, Park SR, Seol SH, Kwon JS. Clinical and empirical applications of the Rey-Osterrieth complex figure test. Nat Protoc 2006;1:892-899.
PUBMED | CROSSREF
5. Corwin J, Bylsma FW. Psychological examination of traumatic encephalopathy. Clin Neuropsychol 2007;7:3-21.
CROSSREF
6. Canham R, Smith S, Tyrrell A, editors. Automated scoring of a neuropsychological test: the Rey Osterrieth complex figure. Proceedings of the 26th Euromicro Conference; 2000 Sep 5-7; Maastricht, Netherlands. Piscataway, NJ: IEEE; 2002. 406-413 p.

7. Vogt J, Kloosterman H, Vermeent S, Van Elswijk G, Dotsch R, Schmand B. Automated scoring of the Rey-Osterrieth complex figure test using a deep-learning algorithm. Arch Clin Neuropsychol 2019;34:836-836.
CROSSREF

8. Zhang X, Lv L, Min G, Wang Q, Zhao Y, Li Y. Overview of the complex figure test and its clinical application in neuropsychiatric disorders, including copying and recall. Front Neurol 2021;12:680474.
PUBMED | CROSSREF

9. Kim KW, Lee SY, Choi J, Chin J, Lee BH, Na DL, et al. A comprehensive evaluation of the process of copying a complex figure in early- and late-onset Alzheimer disease: a quantitative analysis of digital pen data. J Med Internet Res 2020;22:e18136.
PUBMED | CROSSREF

10. Morris JC. The clinical dementia rating (CDR): current version and scoring rules. Neurology 1993;43:2412-2414.
PUBMED | CROSSREF

11. Youn YC, Choi SH, Shin HW, Kim KW, Jang JW, Jung JJ, et al. Detection of cognitive impairment using a machine-learning algorithm. Neuropsychiatr Dis Treat 2018;14:2939-2945.
PUBMED | CROSSREF

12. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019;20:e262-e273.
PUBMED | CROSSREF

13. Kang MJ, Kim SY, Na DL, Kim BC, Yang DW, Kim EJ, et al. Prediction of cognitive impairment via deep learning trained with multi-center neuropsychological test data. BMC Med Inform Decis Mak 2019;19:231.
PUBMED | CROSSREF

14. Caballé-Cervigón N, Castillo-Sequera JL, Gómez-Pulido JA, Gómez-Pulido JM, Polo-Luque ML. Machine learning applied to diagnosis of human diseases: a systematic review. Appl Sci 2020;10:5135.
CROSSREF

15. Van HP, Eun JR. A High-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets. Applied Sciences. 2019;9:4500.
CROSSREF

16. Li Y, Li B, Jiao B. Research and Implementation of CNN based on TensorFlow. IOP Conf Series Mater Sci Eng 2019;490:042022.
CROSSREF

17. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data 2019;6:60.
CROSSREF

18. Carpenter CR, Bassett ER, Fischer GM, Shirshekan J, Galvin JE, Morris JC. Four sensitive screening tools to detect cognitive dysfunction in geriatric emergency department patients: brief Alzheimer's Screen, Short Blessed Test, Ottawa 3DY, and the caregiver-completed AD8. Acad Emerg Med 2011;18:374-384.
PUBMED | CROSSREF

19. Kasai M, Meguro K, Hashimoto R, Ishizaki J, Yamadori A, Mori E. Non-verbal learning is impaired in very mild Alzheimer's disease (CDR 0.5): normative data from the learning version of the Rey-Osterrieth complex figure test. Psychiatry Clin Neurosci 2006;60:139-146.
PUBMED | CROSSREF

20. Mikołajczyk A, Grochowski M, editors. Data augmentation for improving deep learning in image classification problem. International Interdisciplinary PhD Workshop (IIPhDW); 2018 May 9-12; Swinoujscie, Poland. Piscataway, NJ: IEEE; 2018. 117-122 p.

21. Ketkar N. Introduction to Keras. In: Ketkar N, editor. Deep Learning with Python. Berkeley: Apress, 2017;97-111.

22. Grundman M, Petersen RC, Ferris SH, Thomas RG, Aisen PS, Bennett DA, et al. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. Arch Neurol 2014;61:59-66.
PUBMED | CROSSREF

23. Chang YL, Bondi MW, McEvoy LK, Fennema-Notestine C, Salmon DP, Galasko D, et al. Global clinical dementia rating of 0.5 in MCI masks variability related to level of function. Neurology 2011;76:652-659.
CROSSREF

Appendix 1. Rey complex figure: scoring criteria.
Score 2: accurately drawn and correctly placed; score 1: accurately drawn and incorrectly placed, or inaccurately drawn and correctly placed; score 0.5: inaccurately drawn but recognizable, and incorrectly placed; score 0: inaccurately drawn and unrecognizable or omitted, and incorrectly placed.