ORIGINAL RESEARCH

Detection of cognitive impairment using a machine-learning algorithm

Young Chul Youn[1]
Seong Hye Choi[2]
Hae-Won Shin[1]
Ko Woon Kim[3]
Jae-Won Jang[4]
Jason J Jung[5]
Ging-Yuek Robin Hsiung[6]
SangYun Kim[7]

[1]Department of Neurology, College of Medicine, Chung-Ang University, Seoul, South Korea; [2]Department of Neurology, Inha University College of Medicine, Incheon, South Korea; [3]Department of Neurology, Chonbuk National University Medical School and Hospital, Chonbuk, South Korea; [4]Department of Neurology, Kangwon National University Hospital, Chuncheon, South Korea; [5]Department of Computer Engineering, Chung-Ang University, Seoul, South Korea; [6]Division of Neurology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada; [7]Department of Neurology, Seoul National University College of Medicine and Seoul National University Bundang Hospital, Seoul, South Korea

Correspondence: SangYun Kim
Department of Neurology,
Seoul National University College of Medicine and Seoul National University Bundang Hospital, 82, Gumi-ro 173, Bundang-gu, Seongnam-si, Gyeonggi-do 463-707, Seoul, South Korea
Tel +82 31 787 7462
Fax +82 31 787 4059
Email neuroksy@snu.ac.kr

Purpose: The Mini-Mental State Examination (MMSE) is one of the most frequently used bedside screening measures of cognition. However, the Korean Dementia Screening Questionnaire (KDSQ) is an easier and more reliable screening method. Instead, other clinical variables and raw data were used for this study without the consideration of a cutoff value. The objective of this study was to develop a machine-learning algorithm for the detection of cognitive impairment (CI) based on the KDSQ and the MMSE.

Patients and methods: The original dataset from the Clinical Research Center for Dementia of South Korea study was obtained. In total, 9,885 and 300 patients were randomly allocated to the training and test datasets, respectively. We selected up to 24 variables including sex, age, education duration, diabetes mellitus, and hypertension. We trained a machine-learning algorithm using TensorFlow based on the training dataset and then calculated its accuracy using the test dataset. The cost was calculated by conducting a logistic regression.

Results: The accuracy of the model in predicting CI based on the KDSQ only, the MMSE only, and the combination of the KDSQ and MMSE was 84.3%, 88.3%, and 86.3%, respectively. For the KDSQ, the sensitivity for detecting CI was 91.50% and the specificity for detecting normal cognition (NL) was 59.60%. The sensitivity of the MMSE was 94.35%, and the specificity was 59.62%. When combining the KDSQ and the MMSE, the sensitivity for detecting CI was 91.5% and the specificity for detecting NL was 61.5%.

Conclusion: The algorithm predicting CI based on the MMSE is superior. However, the KDSQ can be administered more easily in clinical practice and the algorithm using KDSQ is a comparable screening tool.

Keywords: dementia, mild cognitive impairment, machine learning, TensorFlow, Mini-Mental State Examination, dementia questionnaire

Introduction

The prevention of dementia is one of the highest priorities for public health, and a predictive test is needed for its early intervention. Recently reported and several ongoing prevention trials have focused on the issues of screening of at-risk dementia patients or dementia prediction.[1–3] Several dementia prediction models have been proposed.[4–7]

The Mini-Mental State Examination (MMSE) was developed as a brief screening tool to provide a quantitative assessment of cognitive impairment (CI) and is one of the most frequently used bedside screening measures of cognition.[8] However, subsequent studies have suggested several limitations of its use. For example, it can overestimate impairments in those older than the age of 60 years and in those with less education.[9,10] The MMSE is also insensitive to CI from subcortical lesions in the right hemisphere

and frontal lobes.[11] These limitations are considered to have far-reaching implications for dementia screening.[12]

A questionnaire for dementia screening is a possible alternative. Compared to the MMSE, the Korean Dementia Screening Questionnaire (KDSQ) is a dementia screening questionnaire that does not have to be conducted in person and can be administered by an interviewer without specialized skills. It is not influenced by age or educational level. The KDSQ is sensitive in identifying early dementia patients, and its validity and reliability have previously been evaluated. Conducting a KDSQ is an easier and more reliable screening method (Table S1).[13]

Machine-learning models were originally designed to analyze large, complex medical datasets.[14] Machine-learning algorithms have been used to detect various diseases such as coronary artery disease and liver malfunction and select genes for cancer detection. Recently, machine-learning techniques for diagnosing dementia have been studied, the majority of which involved analyzing brain magnetic resonance imaging (MRI), positron emission tomography, and electroencephalography and some involved analyzing risk factors and gene data.[15–20] However, the intent of this current study was to identify cognitively impaired patients with memory problems using a large dataset of dementia screening test results and to investigate the utility of the KDSQ.

Our objective herein was to evaluate and comprehensively compare machine learning for the diagnosis of CI based on the KDSQ and MMSE. Significant differences are not expected to occur between machine learning using the KDSQ and that using the MMSE.

Patients and methods

We trained a machine-learning algorithm using TensorFlow (https://www.tensorflow.org/) to distinguish between the patients with CI and those with normal cognition (NL) based on the data that were obtained from the Clinical Research Center for Dementia of South Korea (CREDOS).[21] Some of the patients were tested using this algorithm to evaluate the accuracy rate. The design and protocol of this prospective study were approved by the institutional review board of Chung-Ang University Hospital (registration no I2007040 [98]). TensorFlow is an open source software library for machine learning developed by Google based on the python computer language.[22]

Participants

A total of 10,185 patients were selected from the CREDOS study. The CREDOS study was a prospective, multi-center,

hospital-based cohort study designed to assess the occurrence and risk factors of cognitive disorders. The "clinical diagnosis" dataset was composed of participants categorized as one of three diagnostic classes: cognitively normal, mild cognitive impairment (MCI), and Alzheimer's disease (AD). The cognitively normal class technically included subjective cognitive decline (SCD) because the CREDOS data were from a hospital-based study.

The inclusion criteria for SCD were as follows: 1) sustained subjective memory complaints; 2) normal general cognition (within 1 SD of the age- and education-adjusted norms of the Korean version of the MMSE [K-MMSE][23] and a score of >26); 3) intact activities of daily living (ADL); and 4) no abnormality (within 1 SD of age- and education-adjusted norms) on a comprehensive neuropsychological battery (Seoul Neuropsychological Screening Battery).[24,25]

The criteria for MCI in the CREDOS study were as follows: 1) the presence of memory complaints; 2) intact function in ADL; 3) objective CI (≥ 1 SD below age- and education-adjusted norms) in more than one cognitive domain on a comprehensive neuropsychological battery (Seoul Neuropsychological Screening Battery);[24,25] 4) a clinical dementia rating (CDR) of 0.5; and 5) not demented according to the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*)-*IV* criteria.

The patients with AD met the probable AD criteria proposed by the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's Disease and Related Disorders Association[26] as well as the *DSM-IV*.

Patients with clinical evidence of a stroke; structural lesions such as territorial infarction, intracranial hemorrhage, brain tumor, and hydrocephalus; and current or past neurological or psychiatric illnesses such as schizophrenia, epilepsy, brain tumor, encephalitis, and severe head trauma were excluded.

The original dataset we received from CREDOS had 786 variables for 10,185 patients. The data were mined for the objective variables only, and subjective and leading variables that would indicate CI were removed. That is, the variables represented by numbers or scores were adopted, and variables such as neuropsychiatric inventory obtained through caregiver's interviews and a Global Deterioration Score or CDR suggestive of CI were excluded. The missing values that were identified in the past medical history were ultimately enrolled. Ultimately, there were 24 variables (including the outcome variable). The following is a sample of the variables such as sex, age at the time of visit, education duration, diabetes mellitus (DM), hypertension (HT),

hypercholesterolemia, stroke history, 15-item score of the KDSQ, MMSE score, and the outcome variable.

Model training

We divided the patients into two groups: cognitively normal and cognitively impaired, the latter of which included MCI and dementia patients because this study was conducted to develop an algorithm to identify CI among patients who required further inquiry.

The first step in modeling the data comprised the following preprocessing steps. The cardinality variables were standardized. The variables (age at the time of visit, education duration, and MMSE score) were normalized as follows:

$$\chi' = \frac{\chi - \mu}{\sigma}$$

where χ' denotes standardized cardinality of variables, χ denotes raw cardinality of variables, μ denotes mean, and σ denotes SD.

In the second step, the dataset was then split randomly into train and test sets, and each set was developed to create feature (x_data) and outcome (y_data) variables. The third step involved model training with the train dataset using TensorFlow. The cost was calculated by a logistic regression. The fourth step involved calculation of the accuracy using the test dataset (Table S2).

Statistical analyses

The difference between CI and NL with regard to age, education duration, and MMSE and KDSQ scores was evaluated using a Student's *t*-test, and the sex ratio, frequency of DM and HT, and stroke history were evaluated using a chi-squared test.

The accuracies of the MMSE, KDSQ, and MMSE with KDSQ in the prediction of CI were measured by the frequency of correct estimations by the trained algorithm in the test dataset. Therefore, the ratio of correct answers among all patients in the test dataset was used.

The sensitivity and specificity of the algorithm and the receiver operating characteristic (ROC) curve were calculated using Microsoft Excel version 16.0 (Microsoft Corporation, Chicago, IL, USA) and SPSS version 23 (IBM Corporation, Armonk, NY, USA).

Results

After data mining of raw data from CREDOS, a brief inspection of the data revealed that it was an imbalanced dataset. Cases of CI were 5.5-fold more prevalent in the dataset; 84.7% of the cases had CI and 15.3% of the cases were cognitively normal. The gender distribution was 31.2% males and 68.8% females (Table 1).

In total, 9,885 and 300 patients were allocated to the training and test datasets, respectively. The test dataset was composed of 52 cognitively normal and 248 cognitively impaired patients. All were tested with the trained model.

The accuracy of the model for predicting CI based on the KDSQ only, MMSE only, and KDSQ and MMSE combined was 84.3%, 88.3%, and 86.3%, respectively. The KDSQ had a 91.5% sensitivity for detecting CI and a 59.6% specificity for detecting cognitively normal cases. The sensitivity and specificity of the MMSE were 94.35% and 59.62%, respectively. When combining the KDSQ and MMSE, the sensitivity for detecting CI was 91.5% and the specificity for detecting cognitively normal cases was 61.5%.

The ROC curves of the three predictors of the KDSQ, MMSE, and the combination of KDSQ and MMSE are provided in Figure 1. To discriminate CI from NL, the MMSE had the highest area under the ROC curve (0.770), followed by the combination of the KDSQ and MMSE (0.765) and the MMSE only (0.756).

Discussion

This study investigated the utility of an algorithm for predicting CI using a large dataset from the CREDOS study. It also verified the accuracy rate for predicting patients with CI.

The accuracies of the screening tests in this study are lower than those expected in machine learning.[27] However,

Table 1 Demographics of the dataset

	Sex (M:F)	Age (years)[a]	ED	MMSE[a]	KDSQ[a]	DM	HT	DL	Stroke
CI (n=8,627)	2,697:5,930	71.9±8.8	7.2±5.3	21.9±5.9	10.2±7.2	1,852	4,382	1,478	666
NL (n=1,558)	484:1,074	67.5±7.6	7.8±5.2	23.3±5.5	9.2±7.0	309	758	275	106
Total (N=10,185)	3,181:7,004	71.2±8.8	7.3±5.3	22.1±5.8	10.1±7.2	2,161	5,140	1,753	772

Notes: [a]*P*<0.01, statistical significance of the difference between CI and NL as determined by the Student's *t*-test. No statistical significance of frequency difference was found in sex, DM, HT, DL, and Stroke between patients with CI and NL with chi-squared test.
Abbreviations: CI, cognitive impairment; DL, dyslipidemia; DM, diabetes mellitus; ED, educational duration; F, female; HT, hypertension; KDSQ, Korean Dementia Screening Questionnaire; M, male; MMSE, Mini-Mental State Examination; NL, normal cognition; Stroke, stroke history.
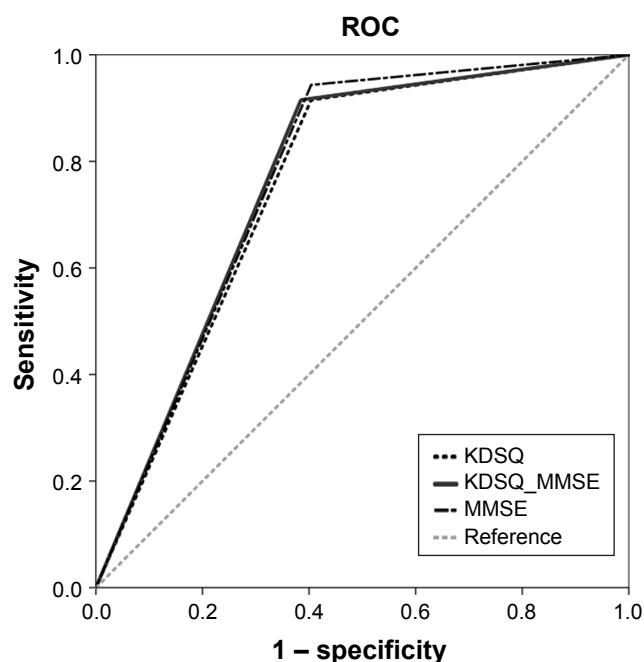
Figure 1 The ROC curve of three predictors of the KDSQ, MMSE, and the combination of KDSQ and MMSE.
Note: The ROC curves were obtained by conducting a logistic regression using TensorFlow on a dataset from the CREDOS.
Abbreviations: CREDOS, Clinical Research Center for Dementia of South Korea; KDSQ, Korean Dementia Screening Questionnaire; MMSE, Mini-Mental State Examination; ROC, receiver operating characteristic.

the sensitivity of the algorithm using the KDSQ to detect the CI was 91.5%, which is good for a screening tool. Even after adding the MMSE, the algorithm did not improve in the detection of CI; however, the specificity improved slightly. The improved specificity may be the reason that the accuracy improved. This algorithm exhibited a higher sensitivity and a lower specificity; however, there are possible implications from a dementia screening perspective. Screening tools should not miss the patients who have CI; therefore, they generally require a higher sensitivity even if this results in a lower specificity.[28]

The algorithm using the MMSE only exhibited the highest sensitivity and accuracy rates. The first reason is that this algorithm was trained with variables including age and education levels, which particularly affect the accuracy of the MMSE compared to that of the KDSQ. The second reason is that this outcome was caused by the inclusion of diagnostic criteria for NL, MCI, and AD in the CREDOS data subset, which already included an MMSE cutoff score between cognitively normal and MCI.

A previous study that investigated screening methods using the KDSQ and MMSE demonstrated that the combination of the KDSQ and MMSE had the highest area under the

ROC curve (0.784).[13] The area under the ROC curve of our algorithm was 0.765, which is similar to that of the previous study; however, the MMSE had the highest area under the ROC curve (0.770). The differences in the values between the current study and the previous study may be due to the previous study already having a cutoff value for the MMSE and KDSQ. Instead, other clinical variables and raw data were used for this study without the consideration of a cutoff value. In addition, this study included data from a substantial number of patients from the CREDOS.

As a screening tool, compared to the KDSQ, the MMSE has several limitations. For example, it must be conducted face-to-face, and the interviewer must be skilled. In addition, it is influenced by age and education level. Therefore, the KDSQ is easier to utilize, particularly in primary care practice.

Significant differences were not expected to occur between machine learning using the KDSQ without and with the MMSE for screening CI because there was no difference in the sensitivity level. The addition of the MMSE improved only the specificity for detecting NL.

The machine-learning algorithm using only the KDSQ and several clinical variables could be useful for screening patients with CI in primary care practice.

However, there are several limitations. The sensitivity of the model is >90%, but even as a screening tool for CI, the specificity is low and the accuracy is not sufficient to reach the level that is typically expected in machine learning. Additionally, the data provided by the CREDOS were an imbalanced dataset since patients with CI were 5.5-fold more prevalent. Furthermore, the model is only applicable to differentiating between cognitively normal persons and patients with MCI or dementia and cannot be used to differentiate dementia subtypes.

Conclusion

We trained and tested a machine-learning algorithm model of the KDSQ for distinguishing cognitively normal and cognitively impaired patients using the CREDOS data and suggest its possibility as a screening tool for CI.

Acknowledgment

Author contributions

All authors contributed toward data analysis and drafting and revision of the paper, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Shaik MA, Khoo CH, Thiagarajah AG, et al. Pilot Evaluation of a Dementia Case Finding Clinical Service Using the Informant AD8 for At-Risk Older Adults in Primary Health Care: A Brief Report. *J Am Med Dir Assoc*. 2016;17(7):673.e5–673.e8.
2. Mendonça MD, Alves L, Bugalho P. From Subjective Cognitive Complaints to Dementia: Who is at Risk? A Systematic Review. *Am J Alzheimers Dis*. 2016;31(2):105–114.
3. Bregman N, Regev K, Moore O, Giladi N, Ash E. A Simple Tool to Reach Populations at Risk for Developing Dementia and Alzheimer's Disease. *J Alzheimers Dis*. 2015;46(1):151–155.
4. Solomon A, Soininen H. Dementia: Risk prediction models in dementia prevention. *Nat Rev Neurol*. 2015;11(7):375–377.
5. Rondeau V, Mauguen A, Laurent A, Berr C, Helmer C. Dynamic prediction models for clustered and interval-censored outcomes: Investigating the intra-couple correlation in the risk of dementia. *Stat Methods Med Res*. 2017;26(5):2168–2183.
6. Lebedeva AK, Westman E, Borza T, et al. MRI-Based Classification Models in Prediction of Mild Cognitive Impairment and Dementia in Late-Life Depression. *Front Aging Neurosci*. 2017;9:13.
7. Stephan BC, Kurth T, Matthews FE, Brayne C, Dufouil C. Dementia risk prediction in the population: are screening models accurate? *Nat Rev Neurol*. 2010;6(6):318–326.
8. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189–198.
9. Crum RM, Anthony JC, Bassett SS, Folstein MF. Population-based norms for the Mini-Mental State Examination by age and educational level. *JAMA*. 1993;269(18):2386–2391.
10. Grigoletto F, Zappalà G, Anderson DW, Lebowitz BD. Norms for the Mini-Mental State Examination in a healthy population. *Neurology*. 1999;53(2):315–320.
11. Oh E, Kang Y, Shin JH, Yeon BK. A Validity Study of K-MMSE as a Screening Test for Dementia: Comparison Against a Comprehensive Neuropsychological Evaluation. *Dement Neurocognitive Disord*. 2010; 9:8–12.
12. Lancu I, Olmer A. The minimental state examination – an up-to-date review. *Harefuah*. 2006;145(9):687–690.
13. Choi SH, Park MH. Three screening methods for cognitive dysfunction using the Mini-Mental State Examination and Korean Dementia Screening Questionnaire. *Geriatr Gerontol Int*. 2016;16(2):252–258.
14. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med*. 2001;23(1):89–109.
15. Bryan RN. Machine Learning Applied to Alzheimer Disease. *Radiology*. 2016;281(3):665–668.
16. Chen R, Herskovits EH. Machine-learning techniques for building a diagnostic model for very mild dementia. *Neuroimage*. 2010;52(1):234–244.
17. Mathotaarachchi S, Pascoal TA, Shin M, et al. Identifying incipient dementia individuals using machine learning and amyloid imaging. *Neurobiol Aging*. 2017;59:80–90.
18. Scheubert L, Luštrek M, Schmidt R, Repsilber D, Fuellen G. Tissue-based Alzheimer gene expression markers-comparison of multiple machine learning approaches and investigation of redundancy in small biomarker sets. *BMC Bioinformatics*. 2012;13:266.
19. Shankle WR, Mania S, Dick MB, Pazzani MJ. Simple models for estimating dementia severity using machine learning. *Stud Health Technol Inform*. 1998;52(Pt 1):472–476.
20. Tohka J, Moradi E, Huttunen H. Alzheimer's Disease Neuroimaging I. Comparison of Feature Selection Techniques in Machine Learning for Anatomical Brain MRI in Dementia. *Neuroinformatics*. 2016;14: 279–296.
21. Park HK, Na DL, Han SH, et al. Clinical characteristics of a nationwide hospital-based registry of mild-to-moderate Alzheimer's disease patients in Korea: a CREDOS (Clinical Research Center for Dementia of South Korea) study. *J Korean Med Sci*. 2011;26(9):1219–1226.
22. Rampasek L, Goldenberg A. TensorFlow: Biology's Gateway to Deep Learning? *Cell Syst*. 2016;2(1):12–14.
23. Han C, Jo SA, Jo I, Kim E, Park MH, Kang Y. 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(3):302–310.
24. Ahn HJ, Chin J, Park A, et al. Seoul Neuropsychological Screening Battery-dementia version (SNSB-D): a useful tool for assessing and monitoring cognitive impairments in dementia patients. *J Korean Med Sci*. 2010;25(7):1071–1076.
25. Jahng S, Na DL, Kang Y. Constructing a Composite Score for the Seoul Neuropsychological Screening Battery-Core. *Dementia Neurocognitive Disord*. 2015;14(4):137–142.
26. Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol*. 2007;6(8):734–746.
27. Prashanth R, Dutta Roy S, Mandal PK, Ghosh S. High-Accuracy Detection of Early Parkinson's Disease through Multimodal Features and Machine Learning. *Int J Med Inform*. 2016;90:13–21.
28. Isella V, Mapelli C, Siri C, et al. Validation and attempts of revision of the MDS-recommended tests for the screening of Parkinson's disease dementia. *Parkinsonism Relat Disord*. 2014;20(1):32–36.

Supplementary materials

Table S1 KDSQ cognition

1. She/He does not know what month and day of the week it is today.
2. She/He cannot find her/his belongings that she/he left somewhere.
3. She/He repeats the same question.
4. She/He forgets to keep her/his word.
5. She/He goes to search for something somewhere, but returns without remembering her/his original purpose.
6. Due to difficulty in remembering something or someone, she/he hesitates to name it or the person.
7. Unable to understand dialog, she/he repeats the same question.
8. She/He has lost her/his way or wandered from place to place.
9. Her/His ability to calculate is worse than before (eg, It is impossible for the patient to calculate commodity prices or changes).
10. Her/His character has changed, compared to the past.
11. She/He is not as good at operating devices (washing machine, electric rice cooker, cultivator, etc.) as before.
12. She/He is not as good at arranging furnishings at home as before.
13. She/He cannot select appropriate clothing for the circumstances.
14. She/He has difficulty with traveling to destinations using mass transportation (difficulties caused by physical disturbances such as arthritis are excluded).
15. She/He fails to change dirty underwear or clothes.

Total score:/:/30
No=0; **Sometimes (or a little)**=1; **Often (or much)**=2

Abbreviation: KDSQ, Korean Dementia Screening Questionnaire.

Table S2 The TensorFlow code using CREDOS data to predict CI

```
import numpy as np
import tensorflow as tf

# each set was developed to create feature (x_data) and outcome (y_data) variables
xy = np.loadtxt('CRCD_KDSQ_MMSE_train.txt', unpack=True, dtype='float32')
x_data = xy[0:−1]
y_data = xy[−1]

# model training with the training dataset
X = tf.placeholder(tf.float32)
Y = tf.placeholder(tf.float32)

W = tf.Variable(tf.random_uniform([1, len(x_data)], −1.0, 1.0))

h = tf.matmul(W, X)
hypothesis = tf.div(1., 1. + tf.exp(−h))

cost = -tf.reduce_mean(Y * tf.log(hypothesis) + (1 − Y) * tf.log(1 − hypothesis))

a = tf.Variable(0.15)
optimizer = tf.train.GradientDescentOptimizer(a)
train = optimizer.minimize(cost)

# calculation of the accuracy with the test dataset
xy = np.loadtxt('CRCD_KDSQ_MMSE_test.txt', unpack=True, dtype='float32')
x_data = xy[0:−1]
y_data = xy[−1]

print (sess.run(hypothesis, feed_dict={X: x_data, Y: y_data}))

answer = tf.equal(tf.floor(hypothesis +0.4), Y)
accuracy = tf.reduce_mean(tf.cast(answer, "float"))
print ("Accuracy: ", accuracy.eval(session=sess, feed_dict={X: x_data, Y: y_data}))
sess.close()
```

Abbreviations: CI, cognitive impairment; CREDOS, Clinical Research Center for Dementia of South Korea.