Journal of Engineering and Applied Sciences 13 (Special Issue 2): 2885-2888, 2018
ISSN: 1816-949X

The Development of Mild Cognitive Impairment Prediction Model based on Random Forest with Focusing on Visuospatial Function and Language Ability

[1]Sunghyoun Cho, [2]Seonghun Yu, [1]Heekyung Jin and [3]Haewon Byeon
[1]Department of Physical Therapy, Nambu University, Gwangju, South Korea
[2]Gwng-ju Trauma Center, Department of Physical Therapy, Gwangju, South Korea
[3]Department of Speech Language Pathology, College of Health Science, Honam University, Gwangju, Republic of Korea

Abstract: This study identified the visuospatial and verbal characteristics of Mild Cognitive Impairment (MCI), developed a MCI prediction model based on the random forest technique with using the identified characteristics and provided basic data for diagnosing and intervening MCI. The subjects of study were 245 elderly over the age of 65 (174 healthy elderly and 71 elderly with MCI). Visuospatial function was assessed by using Rey-Osterrieth Complex Figure (RCFT-copy) and linguistic ability was assessed by using Korean Boston Naming test (K-BNT) and Korean-Mini Mental State Examination (K-MMSE). The prediction model was developed by using the random forest technique and its predictive power was compared with models developed by logistic regression analysis and decision tree. As the result of independent t-test, healthy elderly group had significantly higher mean value of RCFT-copy and K-BNT than MCI group. There was no significant difference, however in reading and writing ability. The prediction model developed by using the random forest technique showed that major predictors of it were RCFT-copy and K-BNT and the accuracy of it was 66.1%. Random forests had better predictive power than logistic regression analysis and decision tree. In this study, visuospatial ability and confrontation naming ability were significant predictors of MCI. In order to investigate the causal relationship between visuospatial functions and confrontational naming ability and MCI, longitudinal studies are required to be conducted in the future.

Key words: Mild cognitive impairment, visuospatial functions, confrontational naming ability, language ability, future, Korea

INTRODUCTION

Aging society is a society in which the elderly over the age of 65 occupy more than 7% of total population and when they are more than 20%, the society is classified as post-aged society. Korea already reached aging society with elderly population occupying 7.2% of the population as of 2000 and is expected to become a post-aged society in 2026. In addition, according to 2015 OECD report, life expectancy of Koreans is 81.8 years, showing that their lives continue more than 15 years after they enter old age (Anonymous, 2017).

With rapid aging trend continuing, interest in geriatric diseases has been on the increase as well. Especially, prevalence rate of dementia for the Korean elderly over the age of 65 was 9.18% (540.000 people) as of 2012 (MHW., 2014) which is higher than that of Canada, UK or India (Prince *et al*., 2013; Iachini *et al*., 2009). Dementia population in Korea is predicted to continuously increase with aging trend. According to survey on dementia prevalence rate conducted by the Ministry of Health and Welfare in 2012 (MHW., 2014), patients with dementia will increase to 1.27 million in 2030 and 2.71 million in 2050 which is a serious trend as the number is expected to double every 20 years.

Approximately 71.3% of dementia patients are known to have Alzheimer's disease (MHW., 2014). Though complete treatment method of Alzheimer's has not been developed yet, it has been reported that there are possibilities to prolong cognitive decline or improve symptoms if sustained interventions are performed with medication or cognitive therapy from the stage of Mild Cognitive Impairment (MCI) which is the previous stage before dementia (Iachini *et al*., 2009). Thus, MCI is clinically important in not only the early discovery of Alzheimer's disease but maximizing the treatment effect.

MCI is a state in which there is decline of cognitive functions but ability to perform daily living is preserved and it is defined as the intermediate stage to dementia in

Corresponding Author: Haewon Byeon, Department of Speech Language Pathology, College of Health Science, Honam University, Gwangju, Republic of Korea

the normal aging process (Petersen, 2004). Prevalence rate of MCI is reported as diverse as 16~21% (Petersen, 2004) and 15% of MCI cases progress into Alzheimer's (Petersen *et al*., 2001). Since, prevalence rate of MCI in Korea was reported to be 27.8% as of 2012 (MHW., 2014), interest in the treatment and prevention of MCI is gradually on the increase.

As patients with MCI are high-risk group for Alzheimer's, exact understanding of the characteristics of MCI is clinically vital because it not just becomes a critical strategy in the early prevention and intervention of MCI but also can protract the progress into Alzheimer's. Even though numerous studies have been conducted on the decline of cognitive functions of MCI, consistent result has not been drawn out (Lee *et al*., 2012; Mendez and Cummings, 2003; Dudas *et al*., 2005; Barbeau *et al*., 2004; Tsutsumimoto *et al*., 2014). This tendency to inconsistent results of the preceding studies implies that difference in cognitive characteristics between healthy elderly and patients with MCI may vary depending on the types of cognitive functions.

Recently, random forests have been used as a way to identify predictors of disease (Byeon, 2015; Byeon *et al*., 2016). Random forest is one of ensemble methods, specifically, designed for classification and prediction. It has excellent predictive power and accuracy because it is composed of multiple decision trees with having different characteristics due to randomness (Brown and Mues, 2012). This study identified the characteristics of the healthy senior citizen's and MCI's visuospatial and language functions, developed an MCI prediction model by using a random forest model based on them and provided basic data for diagnosing MCI in the early stage and intervening it.

MATERIALS AND METHODS

Subjects: The subjects of study were 245 elderly over the age of 65 (174 healthy elderly and 71 elderly with MCI) who used 3 health centers and 2 elderly day care centers in Suwon, Inchon and Hwasung City from September, 2014 through January, 2015 (Table 1). All participants signed a written informed consent form approved by the Institutional Review Board of the A University. Also, the present study was conducted in accordance with the ethical standards of the Declaration of Helsinki. Standards for the selection of subjects were as follows: first, those who were not diagnosed as having dementia by medical institutions: second, those who do not have neurogenic communication disorders such as aphasia: third, those who can read and write: fourth, those who do not have visual or auditory problems for tests: fifth, those who do not have moderate or severe depression: sixth, those who do not have neurological diseases such as Parkinson's or Huntington's disease. According to Petersen *et al*. (2001), MCI was defined as the elderly who subjectively complain about memory disorder have no problem performing daily life, maintain overall cognitive functions but have lower level of memory than those of the same age or education level and do not fit the standards of diagnosis for dementia.

Table 1: Demographic and clinical characteristics of the subjects, M±SD

Variables	HE (n = 174)	MCI (n = 71)
Sex (male/female)	68/106	26/45
Age (Years)*	67.1±2.2	68.5±2.5
K-MMSE*	25.4±2.0	23.1±2.1
CDR*	0.3±0.3	0.5±0.2
GDS*	2.5±0.4	3.1±0.6
K-GDS*	11.7±6.6	15.1±5.9

*p<0.05; HE = Healthy Elderly; MCI = Mild Cognitive Impairment; K-MMSE = Korean-Mini Mental State Examination; CDR = Clinical Dementa Rating; GDS = Global Deterioration Scale; K-GDS = Korean Version of Geriatric Depression Scale

Visuospatial functions were assessed by using Rey-Osterrieth Complex Figure (RCFT-copy) (Meyers and Meyers, 1995). RCFT-copy was composed of a total of 30 points with 18 items. Language functions were assessed with confrontational naming ability by using Korean Boston Naming test (K-BNT) (Kim and Na, 1997) while writing and reading abilities were assessed with 1 item by using Korean-Mini Mental State Examination (K-MMSE) (Kang, 2006).

Statistical analysis: Data were analyzed with SPSS 23.0 (IBM Inc., Chicago, IL) and significance level was defined as <0.05 in two-sided test. Visuospatial and language functions of healthy elderly and MCI were compared by using independent t-test and χ^2-test.

The random forest is a classification algorism developed by Breiman (2001). It is based on an ensemble method and it synthesizes predictions of multiple decision trees (Fig. 1). Decision tree determines split variable and split point at each node based on impurity metrics such as the GINI index and the entropy index. It is easier to interpret the results than the artificial neural network but it has weaker predictive power which is a major shortfall. To overcome this shortfall, the random forest uses the bootstrap method. The algorism of random forest is as follows.

First, extract bootstrap samples by extracting samples (n = N) from analytical data. Second, form a decision tree classifier of a bootstrap sample by using randomly selected variables (n = m) from independent variables (n = P). Third, repeat the abovementioned method and create decision trees (n = B). Fourth, calculate predictions by using the majority rule.

Logistic regression analysis is the most widely used classification model for analyzing and predicting independent variables. It determines where an independent variable is classified. Logistic regression analysis is similar to linear regression analysis in the aspect of analysis objectives and procedures. However, logistic regression analysis is used for categorical variables and their dependent variables are measured in the nominal scale which is different from linear regression analysis (Peng *et al*., 2002). Logistic regression analysis provides regression coefficient or odds ratio on the assumption of linearity, so, it is easy to interpret. However, it has limitations in predicting non-linear data.

Decision tree is an analysis method to classify and predict by schematizing decision rules into a tree shape (Fig. 2). It is easier to interpret than regression analysis because the classification and prediction procedures are represented by inference rules based on tree structure (Byeon, 2014).

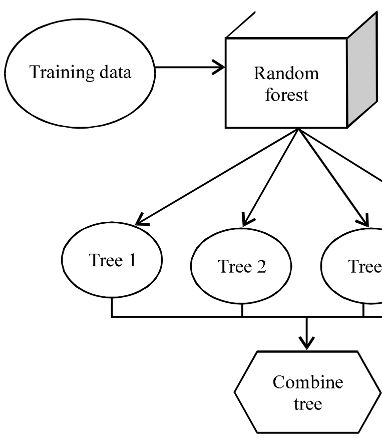

Fig. 1: Random forest

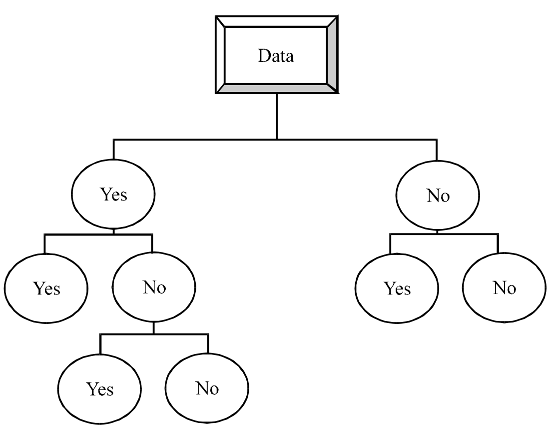

Fig. 2: Decision tree

RESULTS AND DISCUSSION

Comparison of visuospatial ability and verbal ability between the normal elderly group and the MCI group: As the result of analysis the characteristics of visuospatial and language functions of healthy elderly and MCI by using independent t-test and χ^2test (Table 2), there was significant difference in visuospatial functions and naming ability ($p<0.05$). Healthy elderly group had significantly higher mean value of RCFT-copy and K-BNT than MCI group ($p<0.05$). There was no significant difference, however in reading and writing ability.

Comparison of results and prediction model's accuracy: The MCI prediction model was developed through random forests and the predictive power of it was compared with those of logistic regression model and decision tree model (Table 3). Analysis results on training data showed that random forests had a very high accuracy (61.3%). On the other hand, the accuracy of decision tree model was 60.8% and that of logistic regression model was 56.5% which was the lowest.

In test data, the random forest model had the highest accuracy (66.1%) and logistic regression model showed the lowest accuracy (56.8%). The random forest model had the highest accuracy in both training and test data.

Predictions models were established for predicting MCI based on logistic regression, decision tree and random forests by using four explanatory variables (Table 3). Logistic regression, decision tree and random forests all indicated that RCFT-copy and K-BNT were major risk factors of MCI. The accuracy was 66.1, 61.5 and 56.8% for random forests, decision tree and logistic regression, respectively.

Table 2: Characteristics of visuospatial and language functions, M±SD

Variables	HE	MCI
RCFT-copy*	31.3±6.7	24.1±10.2
K-BNT*	41.4±6.3	34.1±10.8
K-MMSE-reading, n (%)		
Positive response	174 (100)	71 (100)
K-MMSE-writing, n (%)		
Positive response	172 (98.9)	70 (98.6)

*$p<0.05$; HE = Healthy Elderly; MCI = Mild Cognitive Impairment; RCFT = Rey-Osterrieth Complex Figure; K-BNT = Korean-Boston Naming Test; K-MMSE = Korean-Mini Mental State Examination

Table 3: Comparison of results and prediction model's accuracy

Data/Models	Accuracy (%)
Training data	
Logistic regression	56.5
Decision tree	60.8
Random forest	61.3
Test data	
Logistic regression	56.8
Decision tree	61.5
Random forest	66.1

Grasping clinical characteristics of MCI is crucial for early detection and intervention of dementia. As the result of comparison of visuospatial and language functions between healthy elderly and MCI, there were differences in visuospatial and naming ability except for reading and writing ability which corresponds with the results of numerous preceding studies.

Alzheimer's disease progresses by stage from normal state through MCI to dementia in the aging process and the most clearly visible damage in cognitive functions detected in early stage is memory disorder (Iachini *et al*., 2009). Other than memory disorder, however, problems of language functions occur such as verbal fluency, word finding and naming and visuospatial problems also take place such as getting lost in familiar places and experiencing loss of direction. It has been reported that early AD patients generally perform more poorly in visuospatial composition assignments than healthy elderly do.

In particular, since, not only do the elderly with MCI have significantly lower performance in verbal fluency and naming ability but also attention and naming have been reported to be significant predicting factors for the future onset of Alzheimer's disease, language functions have received much attention in distinguishing MCI.

In this study, however, there was no significant difference in reading and writing ability between healthy elderly and elderly with MCI. This result is assumed to be caused by following two reasons. First, there is a possibility that MCI or early stage of Alzheimer might have relatively well-preserved reading and writing ability. Preceding studies also reported that elderly with amnestic MCI which is the prior stage of dementia had the level of language functions such as reading similar to that of the healthy elderly (Jacova *et al*., 2007). It is known that reading and writing is anatomico-physiologically related to supramarginal gyrus or angular gyrus located in parietal lobe (Simon *et al*., 2002) which receives smaller damage than hippocampus in early stages of Alzheimer's disease.

Second, the result comes from the limitations of the test which examined reading and writing abilities. As this study used items of K-MMSE to test reading and writing ability, there were limitations in assessing detailed level of language functions of the subjects.

This study established an MCI prediction model and confirmed that random forest had the best predictive algorithm compared with decision tree or regression analysis. The results were presumed to be because the random forest technique is an ensemble method based on decision tree and it was classified based on the independent set of random vectors. The previous study reported that decision tree models had good explanatory power with lower predictive power and the stability of the model was low when there were only a few explanatory variables (Siroky, 2009). Therefore, it is believed that random forest models which has good predictive power and is stable will be effective for analyzing medical examination data which generally has a few explanatory variables.

CONCLUSION

This study developed a prediction model based on random forests and confirmed that RCFT-copy and K-BNT are main predictors of MCI. In order to investigate the causal relationship between visuospatial functions and confrontational naming ability and MCI, longitudinal studies are required to be conducted in the future.

LIMITATIONS

The limitations of this study are as follows. First, as the study used items of K-MMSE, there were restrictions in assessing detailed level of language functions of the subjects. Second, this study did not classify the sub-categories of MCI. Thus, studies are required in the future which classify MCI into 4 categories of Amnestic MCI single domain, Amnestic MCI multiple domain, Non-Amnestic MCI single domain and Non-Amnestic MCI multiple domain and compare their characteristics of cognition and memory.

ACKNOWLEDGEMENT

The researchers would like to thank the subjects who participated in this study.

REFERENCES

Anonymous, 2017. Health status. Organisation for Economic Co-operation and Development (OECD), Paris, France. http://stats. oecd. org/index. aspx?DataSetCode= HEALTH_STAT.

Barbeau, E., M. Didic, E. Tramoni, O. Felician and S. Joubert *et al*., 2004. Evaluation of visual recognition memory in MCI patients. Neurol., 62: 1317-1322.

Breiman, L., 2001. Random forests. Mach. Learn., 45: 5-32.

Brown, I. and C. Mues, 2012. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Syst. Applic., 39: 3446-3453.

Byeon, H., 2014. Verbal and visuospatial memory deficits in mild cognitive impairment: A cross-sectional study. Intl. Inf. Inst. Inf., 17: 6319-6324.

Byeon, H., 2015. A prediction model for mild cognitive impairment using random forests. Intl. J. Adv. Comput. Sci. Appl., 6: 8-12.

Byeon, H., D. Lee and S. Cho, 2016. Assessment for the model predicting of the cognitive and language ability in the mild dementia by the method of data-mining technique. Assess., 7: 75-79.

Dudas, R.B., F. Clague, S.A. Thompson, K.S. Graham and J.R. Hodges, 2005. Episodic and semantic memory in mild cognitive impairment. Neuropsychologia, 43: 1266-1276.

Iachini, T., A. Iavarone, V.P. Senese, F. Ruotolo and G. Ruggiero, 2009. Visuospatial memory in healthy elderly, AD and MCI: A review. Curr. Aging Sci., 2: 43-59.

Jacova, C., A. Kertesz, M. Blair, J.D. Fisk and H.H. Feldman, 2007. Neuropsychological testing and assessment for dementia. Alzheimer's Dementia, 3: 299-317.

Kang, Y., 2006. A normative study of the Korean Mini-Mental State Examination (K-MMSE) in the elderly. Korean J. Psychol., 25: 1-12.

Kim, H. and D.L. Na, 1997. Korean-Boston Naming Test. Haji Commercial Company, Seoul, South Korea.

Lee, L.K., S. Shahar, A.V. Chin, N.A.M. Yusoff, N. Rajab and S.A. Aziz, 2012. Prevalence of gender disparities and predictors affecting the occurrence of Mild Cognitive Impairment (MCI). Arch. Gerontol. Geriatrics, 54: 185-191.

MHW., 2014. Survey on dementia prevalence rate 2012. Ministry of Health and Welfare, Sejong, Korea.

Mendez, M.F. and J.L. Cummings, 2003. Dementia: A Clinical Approach. Butterworth-Heinemann, Boston, California,.

Meyers, J.E. and K.R. Meyers, 1995. Rey Complex Figure Test and Recognition Trial Professional Manual. PAR, Inc., Lake Magdalene, Florida.

Peng, C.Y., K.L. Lee and G.M. Ingersoll, 2002. An introduction to logistic regression analysis and reporting. J. Educ. Res., 96: 3-14.

Petersen, R.C., 2004. Mild cognitive impairment as a diagnostic entity. J. Internal Med., 256: 183-194.

Petersen, R.C., R. Doody, A. Kurz, R.C. Mohs and J.C. Morris *et al*., 2001. Current concepts in mild cognitive impairment. Arch. Neurol., 58: 1985-1992.

Prince, M., R. Bryce, E. Albanese, A. Wimo, W. Ribeiro and C.P. Ferri, 2013. The global prevalence of dementia: A systematic review and metaanalysis. Alzheimers Dementia, 9: 63-75.e2.

Simon, O., J.F. Mangin, L. Cohen, D.L. Bihan and S. Dehaene, 2002. Topographical layout of hand, eye, calculation and language-related areas in the human parietal lobe. Neuron, 33: 475-487.

Siroky, D.S., 2009. Navigating random forests and related advances in algorithmic modeling. Stat. Surv., 3: 147-163.

Tsutsumimoto, K., T. Doi, H. Shimada, H. Makizako and K. Uemura *et al*., 2014. Self-reported exhaustion is associated with small life space in older adults with mild cognitive impairment. J. Phys. Ther. Sci., 26: 1979-1983.