Exploring Parkinson's Disease Predictors based on Basic Intelligence Quotient and Executive Intelligence Quotient

Haewon Byeon

Department of Medical Big Data

College of AI Convergence, Inje University

Gimhae 50834, Gyeonsangnamdo, South Korea

Abstract—It is important to identify the risk factors of dementia and prevent them for the health of patients and caregivers. This study (1) explored sampling methods that could minimize overfitting due to data imbalance using a data-level approach, (2) developed nine ensemble learning models for predicting Parkinson's Disease–Mild Cognitive Impairment (PD-MCI) ((undersampling, oversampling, and SMOTE) × (boosting, bagging, and random forest)=9), and (3) compared the accuracies, sensitivities, and specificities of these models to understand the prediction performance of the developed models. We examined 368 subjects: 320 healthy elderly people (≥60 and ≤74 years old) without Parkinson's disease (168 men and 152 women) and 48 subjects with PD-MCI (20 men and 28 women). This study used the Cognition Scale for Olde Adults (CSOA), which could measure cognitive functions comprehensively while considering age and education level, to determine the specific cognitive level of the subject. Our study developed nine prediction models ((undersampling, oversampling, and SMOTE) × (boosting, bagging, and random forest)=9) for developing a model to predict PD-MCI based on basic intelligence quotient and executive intelligence quotient. The analysis results showed that a random forest classifier with SMOTE had the best prediction performance with a sensitivity of 69.2%, a specificity of 75.7%, and a mean overall accuracy of 74.0%. In this final model, digit span test-backward, stroop test-interference trial, verbal memory test-delayed recall, verbal fluency test, and confrontation naming test were identified as the key variables with high weight in predicting PD-MCI. The results of this study implied that a random forest classifier with SMOTE could produce models with higher accuracy than a bagging classifier with SMOTE or a boosting classifier with SMOTE when analyzing imbalanced data.

Keywords—*Undersampling; oversampling; SMOTE; random forest; Parkinson's disease–mild cognitive impairment*

I. INTRODUCTION

The prevalence of dementia is rapidly increasing in South Korea along with the increase of the elderly population [1]. The National Dementia Epidemiology Survey conducted by the Ministry of Health and Welfare in 2012 showed that the dementia prevalence of the elderly (≥65 years old) in 2012 was 9.18% and the number of dementia patients was 540,755 (155,955 men and 384,800 women) [2]. The survey predicted that the prevalence of dementia in old age will increase up to 13.17% in 2050 [2]. Dementia is a stressful disease for both patients and their families because the overall cognitive function of adults who have achieved normal cognitive development declines, the patients have to struggle against dementia for a long time, and symptoms gradually worsen [3]. Therefore, it is important to identify the risk factors of dementia and prevent them for the health of patients and caregivers [4].

Especially, it is critical to screening dementia as soon as possible from the viewpoint of geriatric medicine. Dementia is known as an irreversible disease that is difficult to cure after it occurs [5]. However, thanks to the rapid development of molecular biology, many studies [6,7,8] have continuously reported that cholinergic enzyme inhibitors such as donepezil can delay the progress of dementia or inhibit the decline of cognitive function. As a result, the perception of dementia treatment has been shifted and early detection of high dementia risk groups has emerged as an important topic. Consequently, if we can detect high dementia risk groups sooner, it will be possible to provide professional counseling on the prognosis and help people establish a better health plan in old age.

Before the onset of dementia, the preclinical phase can last from five to seven years [9]. If appropriate therapeutic interventions are provided during this period, it is possible to delay the development of dementia for about 5 years [10]. Therefore, recent studies have focused on detecting the preclinical phase, particularly mild cognitive impairment (MCI), which is known as a middle stage between normal aging and dementia, as soon as possible [11]. Nevertheless, compared to studies on the MCI of Alzheimer's dementia, much fewer studies have identified the risk factors of Parkinson's disease–mild cognitive impairment (PD–MCI) [12]. Moreover, it has rarely evaluated the relationship between neuropsychological tests and PD-MCI using machine learning [13].

Over the past decade, many studies have widely utilized ensemble learning, a supervised learning algorithm, for classifying and predicting the complex risk factors of diseases [14,15,16]. Although ensemble learning is known to be more accurate than conventional decision trees [17], when a prediction model is developed using binomial categorical imbalanced data, the recall and precision of it are highly likely to decrease because the classification can be biased into major classes. In particular, in the case of disease data, since the

number of patients is generally smaller than that of healthy people, data imbalanced problems are more likely to occur [18,19]. Therefore, a sampling technique for processing imbalanced data is additionally needed to overcome the prediction error due to class imbalance in disease data. Previous studies [20,21,22] suggested using oversampling, undersampling, and synthetic minority over-sampling technique (SMOTE) to improve the classification performance for imbalanced data. This study (1) explored sampling methods that could minimize overfitting due to data imbalance using a data-level approach, (2) developed nine ensemble learning models for predicting PD-MCI ((undersampling, oversampling, and SMOTE) × (boosting, bagging, and random forest)=9), and (3) compared the accuracies, sensitivities, and specificities of these models to understand the prediction performance of the developed models.

Construction of this study is as follows: Section II explains subjects, measurements, a data-level approach for improving classification performance of imbalanced data, and analyzed variables. Section III compares the results of developed nine prediction model ((undersampling, oversampling, and SMOTE) × (boosting, bagging, and random forest)). Lastly, Section IV presents conclusion and direction for future studies.

II. METHODS

A. Subjects

This study examined 368 subjects: 320 healthy elderly people (≥ 60 and ≤ 74 years old) without Parkinson's disease (168 men and 152 women) and 48 subjects with PD-MCI (20 men and 28 women). In this study, patients with Parkinson's disease were defined as patients diagnosed with idiopathic Parkinson's disease according to the diagnostic criteria of the United Kingdom Parkinson's Disease Society Brain Bank. The criteria for selecting healthy elderly were (1) those who received at least 24 points from the Korean version of Mini-Mental State Examination (K-MMSE)[23], a normal range, (2) those who did not have any impairment in vision and hearing for performing cognitive tests, and (3) those who did not have a history of stroke, Parkinson's disease, or dementia.

G-Power version 3.1.9.6 (Universität Mannheim, Mannheim, Germany) was used to conduct a power test for the sample size of this study. When the number of predictors was 18, significance level (alpha) was 0.05, power (1-B) was 0.95, and the effect size (f2) was 0.15, the minimum sample size was estimated as 213. Therefore, the sample size of this study (n=373) exceeded the recommended sample size for testing the statistical significance (Fig. 1 & 2).
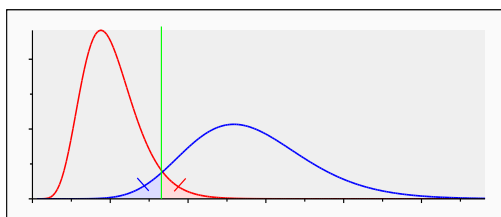

Fig. 1. Results of Power Test.

Fig. 2. Results of Calculating the Appropriate Sample Size to Test Statistical Significance.

B. Measurements

This study used the Cognition Scale for Olde Adults (CSOA)[24], which could measure cognitive functions comprehensively while considering age and education level, to determine the specific cognitive level of the subject. The CSOA is a standardized cognitive test that can comprehensively measure the cognitive functions of the elderly who are suspected to suffer from cognitive impairment or dementia. The CSOA is composed of stroop simple trial, stroop interference trial, digit span test-forward, digit span test-backward, general information, verbal fluency test, confrontation naming test, Rey complex figure test-copy, recognition, immediately recall, and delayed recall. Among them, stroop simple trial, digit span test-forward, general information, confrontation naming test, and delayed recognition were defined as basic intelligence quotient. Stroop interference trial, digit span test-backward, verbal fluency test, Rey complex figure test-copy, immediately recall, and delayed recall were defined as executive intelligence quotient. The sum of basic intelligence quotient and executive intelligence quotient was defined as full-scale intelligence quotient. Kim (2011)[25] reported that the reliability of the CSOA (Cronbach alpha) was 0.932. This study converted the raw scores of 10 sub-tests into standard scores with an average of 100 and a standard deviation of 15 and used them for machine learning. The composition of the CSOA's sub-tests is presented in Fig. 3.
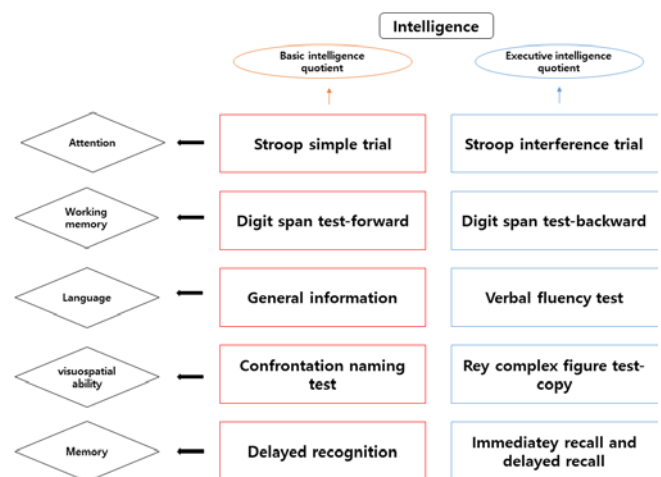

Fig. 3. Composition of basic Intelligence Quotient and Executive Intelligence Quotient in the CSOA.

C. Definitions of Variables

Digit Span Test: When the tester calls out a number, the test subject repeats it immediately after listening to it. There are two-digit span tests: digit span test-forward and digit span test-backward. Each test starts with an item with a short-length number and the length gradually increases as the test progresses. Each raw score is the sum of all items, and it ranges from 0 to 14 points.

Stroop Test: Stroop test consists of a simple trial and an interference trial. The simple trial measures the reaction time takes to tell the colors of 24 circles. The interference trial measures the reaction time to tell the color of a word that describes a color (for example, if "yellow" is written in red, the correct response is "red"). A higher score indicates better reaction sense.

Verbal Memory Test: The verbal memory test is a test that comprehensively calculates the Memory Function Index using 10 picture cards. It is conducted in the order of immediate recall, delayed recall, and recognition. The delayed recall shall be conducted 15-20 after performing the immediate recall. The recognition shall be carried out immediately after completing the delayed recall. The raw score shall be calculated by summing the immediate recall, delayed recall, and recognition, and the score ranges from 0 to 50 points.

General Information: It is a series of questions and answers, and these questions ask common sense. It consists of 20 questions, and each question is one point. Therefore, the total score ranges from 0 to 20 points. A higher score indicates better common sense.

Verbal Fluency Test: It is composed of two trials. The subject shall state nouns in the animal category as many as possible in the first trial and nouns in the crop category as many as possible in the second trial. The time limit for each trial is 1 minute. The raw score is calculated by summing the number of correct responses in the first and second trials. A higher score refers to a better visuospatial function and verbal fluency.

RCFT: Rey Complex Figure (RCF) is to test a subject by copying a figure. Copying is defined as a visuospatial ability, and recalled drawing is defined as visuospatial memory. RCF is evaluated by scoring 18 elements. Each element shall be evaluated by considering the shape and position of each figure, and the original score ranges from 0 to 36 points. A higher score indicates a better visuospatial ability and visuospatial memory.

Confrontation Naming Test: This test asks a subject to read a drawing of an object and say the name (noun) of it. It consists of 24 items, and the range of the raw score is 0 to 24 points. A higher score indicates better confrontation naming ability.

Explanatory variable: Explanatory variables were education level ("middle graduation or below" or "high school graduation or above"), gender (male or female), age, living with a spouse (living together, bereavement/separated, or single), economic activity (yes or no), subjective stress (yes or no), mean monthly household income (<₩1.5 million, ₩1.5-3.0 million, or ≥₩3.0 million), smoking (non-smoking or smoking), drinking (non-drinking or drinking), MMSE-K, verbal memory test, stroop test, general information, digit span test, RCFT, confrontation naming test, verbal fluency test, total score of activities of daily living (ADL), and total score of instrumental activities of daily living (IADL).

D. A Data-level Approach for Improving Classification Performance of Imbalanced Data

The results of this study showed that 86.9% (n=320) of the subjects were healthy without suffering from PD-MCI and those suffering from PD-MCI were 13.1% (n=48), indicating that the data was imbalanced. A classifier that learns from binomial categorical imbalanced data, which have a large difference between the size of a major group and that of a minor group, tends to have a classification biased toward the majority group. Therefore, it classifies the majority of the data into the major group to severely reduce the classification accuracy of the minor. In other words, a prediction model developed from unbalanced data can have a higher overall accuracy, but it is highly likely to show a low precision and recall for a minor group. This study used undersampling [26], oversampling [27], and SMOTE methods [28] as data-level approaches to improve the classification performance of binomial categorical imbalanced data.

Undersampling is a method of overcoming the data imbalance issue by randomly removing samples falling in a major class. The undersampling can save time for constructing a model by reducing the amount of data, but it has a disadvantage of losing information [20,29]. Oversampling is a method of overcoming the data imbalance issue by randomly replicating samples falling in a minor class [30].

The oversampling technique takes more time to build a model because the sample size increases, and it may cause an overfitting issue because it copies samples in a minor class [22,31]. The SMOTE finds n nearest neighbors in a minor class of a certain datum in the minor class. Afterward, it draws a line between the datum and the nearest neighbor and randomly generates data along the line until these randomly generated data become synthetic [32].

E. Development of Prediction Models and Evaluation of Prediction Performance

This study developed nine prediction models ((undersampling, oversampling, and SMOTE) × (boosting, bagging, and random forest)=9) for developing a model to predict PD-MCI based on basic intelligence quotient and executive intelligence quotient. The prediction performance of the developed models was tested by using 5-fold cross-validation. Since the ensemble algorithm has randomness, when the ensemble model was reiterated, seed #12468 was always used. In all ensemble models, the number of decision trees (ntree) was set to 100.

The prediction performance of the developed models was compared by using the accuracy, sensitivity, and specificity of each model. Accuracy indicates the rate of predicting the outcome correctly. Sensitivity refers to the rate of predicting PD-MCI as PD-MCI. Specificity means the rate of predicting a healthy elderly person without PD-MCI and a healthy

elderly person without PD-MCI. This study compared the prediction performance of models and defined that the best prediction model was a model with the highest accuracy while sensitivity and specificity were at least 0.6. The best model was selected as the final model for predicting PD-MCI. All analyzes were performed using R version 4.0.2 (Foundation for Statistical Computing, Vienna, Austria).

III. RESULTS

A. Comparing the Accuracy of the Developed Prediction Models

The accuracy, sensitivity, and specificity of the nine prediction models are presented in Fig. 4, 5, and 6, respectively. The analysis results showed that a random forest classifier with SMOTE had the best prediction performance with a sensitivity of 69.2%, a specificity of 75.7%, and a mean overall accuracy of 74.0%. On the other hand, a boosting classifier with undersampling had the worst performance among the nine prediction models with a sensitivity of 51.8%.
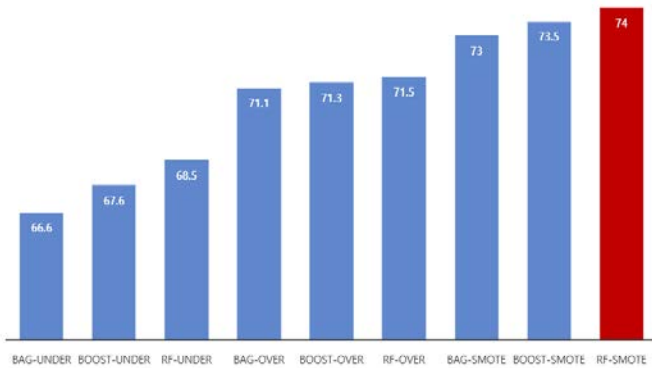

Fig. 4. Accuracy of the Nine Classifier Ensembles.

RF-SMOTE=random forest classifier with SMOTE; RF-OVER=random forest classifier with oversampling; RF-UNDER=random forest classifier with undersampling; BAG-SMOTE=bagging classifier with SMOTE; BAG-OVER=bagging classifier with oversampling; BAG-UNDER=bagging classifier with undersampling; BOOST-SMOTE=boosting classifier with SMOTE; BOOST-OVER=boosting classifier with oversampling; BOOST-UNDER=boosting classifier with undersampling.
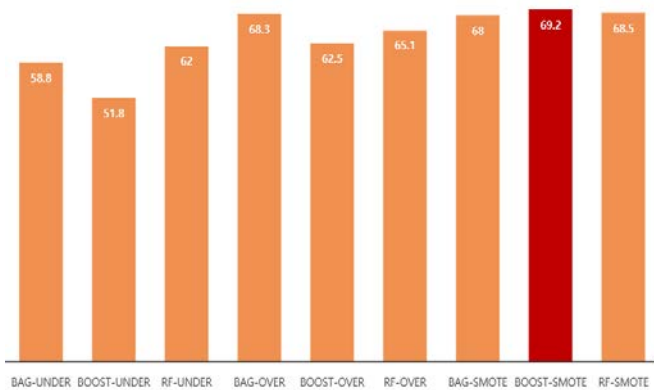

Fig. 5. Sensitivity of the Nine Classifier Ensembles.

RF-SMOTE=random forest classifier with SMOTE; RF-OVER=random forest classifier with oversampling; RF-UNDER=random forest classifier with undersampling; BAG-SMOTE=bagging classifier with SMOTE; BAG-OVER=bagging classifier with oversampling; BAG-UNDER=bagging classifier with undersampling; BOOST-SMOTE=boosting classifier with SMOTE; BOOST-OVER=boosting classifier with oversampling; BOOST-UNDER=boosting classifier with undersampling.
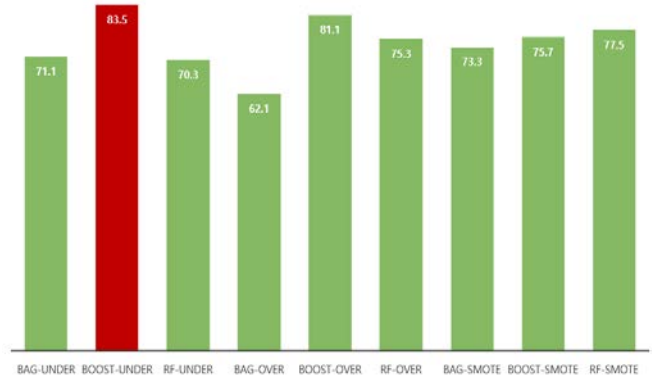

Fig. 6. Specificity of the Nine Classifier Ensembles.

RF-SMOTE=random forest classifier with SMOTE; RF-OVER=random forest classifier with oversampling; RF-UNDER=random forest classifier with undersampling; BAG-SMOTE=bagging classifier with SMOTE; BAG-OVER=bagging classifier with oversampling; BAG-UNDER=bagging classifier with undersampling; BOOST-SMOTE=boosting classifier with SMOTE; BOOST-OVER=boosting classifier with oversampling; BOOST-UNDER=boosting classifier with undersampling.

B. Importance of Variables for PD-MCI Classification in the Final Model

The normalized importance of the variables of the final model (random forest classifier with SMOTE) is presented in Fig. 7. In this model, digit span test-backward, stroop test-interference trial, verbal memory test-delayed recall, verbal fluency test, and confrontation naming test were identified as the key variables with high weight in predicting PD-MCI. Among them, digit span test-backward was the most important variable in a random forest classifier with SMOTE.
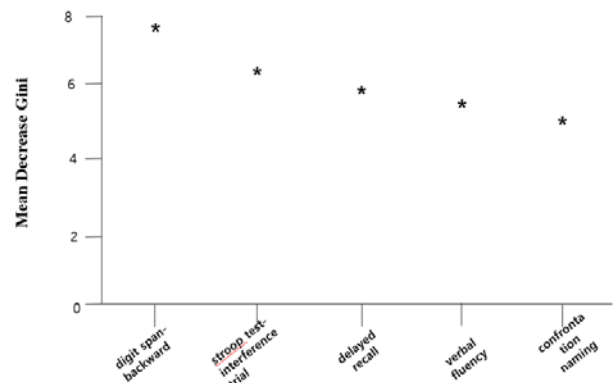

Fig. 7. The Importance of Variables in the Random Forest Classifier with SMOTE-based PD-MCI Prediction Model (Only the Top 5 are Presented).

IV. Discussion

This study compared the prediction performance of nine ensemble learning models ((undersampling, oversampling, and SMOTE) × (undersampling-boosting, bagging, and random forest)=9) for predicting PD-MCI. The results of this study showed that the random forest classifier with SMOTE was the best model (sensitive=69.2%, specificity=75.7%, and mean overall accuracy=74.0%). The result of this study agreed with the results of previous studies [16, 33] showing that random forest based models were superior to other machine learning algorithms for predicting diseases. Particularly, this study developed models by applying oversampling, undersampling, and SMOTE as data-level approaches for improving the classification performance of imbalanced data. It is noteworthy that the accuracy of a random forest classifier with SMOTE was better than that of other learning machine algorithms and ensemble models with SMOTE, oversampling-random forest, or undersampling-random forest.

V. Conclusion

The results of this study implied that a random forest classifier with SMOTE could produce models with higher accuracy than a bagging classifier with SMOTE or a boosting classifier with SMOTE when analyzing imbalanced data. Additional studies are needed to compare the accuracy by using various datasets from diverse fields to prove the prediction performance of a random forest classifier with SMOTE.

Acknowledgment

References.

[1] M. J. Cho, The prevalence and risk factors of dementia in the Korean elderly. Health and welfare policy forum, vol. 2009, no. 10, pp. 43-48, 2009.

[2] G. J. Yoon, The 2021 outlook for healthcare policy. Health and welfare policy forum, vol. 2021, no. 1, pp. 9-22, 2021.

[3] J. Hashimie, S. K. Schultz, and J. T. Stewart, Palliative care for dementia: 2020 update. Clinics in Geriatric Medicine, vol. 36, no. 2, pp. 329-339, 2020.

[4] H. Byeon, Development of depression prediction models for caregivers of patients with dementia using decision tree learning algorithm. International Journal of Gerontology, vol. 13, pp. 314-319, 2019.

[5] M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, and C. P. Ferri, The global prevalence of dementia: a systematic review and metaanalysis. Alzheimer's & Dementia, vol. 9, no. 1, pp. 63-75, 2013.

[6] P. Celsis, Age-related cognitive decline, mild cognitive impairment or preclinical Alzheimer's disease?. Annals of Medicine, vol. 32, pp. 6-14, 2000.

[7] S. T. DeKosky, and K. Marek, Looking backward to move forward: early detection of neurodegenerative disorders. Science, vol. 302, no. 5646, pp. 830-834, 2003.

[8] A. Lleó, S. M. Greenberg, and J. H. Growdon, Current pharmacotherapy for Alzheimer's disease. Annual Review of Medicine, vol. 57, pp. 513-533, 2006.

[9] K. M. Langa, and D. A. Levine, The diagnosis and management of mild cognitive impairment: a clinical review. JAMA, vol. 312, no. 23, pp. 2551-2561, 2014.

[10] T. Luck, M. Luppa, S. Briel, and S. G. Riedel-Heller, Incidence of mild cognitive impairment: a systematic review. Dementia and Geriatric Cognitive Disorders, vol. 29, no. 2, pp. 164-175, 2010.

[11] J. M. Ellison, A 60-year-old woman with mild memory impairment: review of mild cognitive impairment. JAMA, vol. 300, no. 13, pp. 1566-1574, 2008.

[12] G. J. Geurtsen, J. Hoogland, J. G. Goldman, B. A. Schmand, A. I. Tröster, D. J. Burn, and I. Litvan, Parkinson's disease mild cognitive impairment: application and validation of the criteria. Journal of Parkinson's Disease, vol. 4, no. 2, pp. 131-137, 2014.

[13] H. Byeon, Predicting the severity of Parkinson's disease dementia by assessing the neuropsychiatric symptoms with an SVM regression model. International Journal of Environmental Research and Public Health, vol. 18, no. 5, pp. 2551, 2021.

[14] K. H. Miao, J. H. Miao, and G. Miao, Diagnosing coronary heart disease using ensemble machine learning. International Journal of Advanced Computer Science and Application, vol. 7, no. 10, pp. 30-39, 2016.

[15] A. K. Verma, S. Pal, and S. Kumar, Classification of skin disease using ensemble data mining techniques. Asian Pacific Journal of Cancer Prevention, vol. 20, no. 6, pp. 1887-1894, 2019.

[16] H. Byeon, Best early-onset Parkinson dementia predictor using ensemble learning among Parkinson's symptoms, rapid eye movement sleep disorder, and neuropsychological profile. World Journal of Psychiatry, vol. 10, no. 11, 245-259, 2020.

[17] H. Byeon, Evaluating the accuracy of models for predicting the speech acceptability for children with cochlear implants. International Journal of Advanced Computer Science and Applications, vol. 12, no. 2, pp. 25-29, 2021.

[18] H. He, and E. A. Garcia, Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263-1284, 2009.

[19] H. Byeon, Comparing the accuracy and developed models for predicting the confrontation naming of the elderly in South Korea using weighted random forest, random forest, and support vector regression. International Journal of Advanced Computer Science and Applications, vol. 12, no. 2, pp. 326-331, 2021.

[20] H. Byeon, Development of a physical impairment prediction model for Korean elderly people using synthetic minority over-sampling technique and XGBoost. International Journal of Advanced Computer Science and Applications, vol. 12, no. 1, pp. 36-41, 2021.

[21] R. Mohammadi, R. Javidan, M. Keshtgari, and N. Rikhtegar, SMOTE: an intelligent SDN-based multi-objective traffic engineering technique for telesurgery. IETE Journal of Research, vol. 1-11, 2021.

[22] H. Byeon, Predicting the depression of the South Korean elderly using SMOTE and an imbalanced binary dataset. International Journal of Advanced Computer Science and Applications, vol. 12, no. 1, pp. 74-79, 2021.

[23] Y. C. Kwon, Korean version of mini-mental state examination (MMSE-K). Journal of the Korean Neurological Association, vol. 1, pp. 123-135, 1989.

[24] H. K. Kim, and T. Y. Kim, Cognition scale for older adults; CSOA: manual. Neuropsy Incorporated, Daegu, 2007.

[25] Y. S. Kim, Diabetes and cognitive function in community-dwelling older adults. Journal of Korean Academy of Community Health Nursing, vol. 22, no. 4, pp. 377-388, 2011.

[26] S. J. Yen, and Y. S. Lee, Cluster-based under-sampling approaches for imbalanced data distributions. Expert Systems with Applications, vol. 36, no. 3, pp. 5718-5727, 2009.

[27] L. Abdi, and S. Hashemi, To combat multi-class imbalanced problems by means of over-sampling techniques. IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 1, 238-251, 2015.

[28] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.

[29] H. Guan, Y. Zhang, M. Xian, H. D. Cheng, and X. Tang, SMOTE-WENN: Solving class imbalance and small sample problems by oversampling and distance scaling. Applied Intelligence, vol. 51, no. 3, pp. 1394-1409, 2021.

[30] M. Wang, X. Yao, and Y. Chen, An imbalanced-data processing algorithm for the prediction of heart attack in stroke patients. IEEE Access, vol. 9, pp. 25394-25404, 2021.

[31] M. T. García-Ordás, C. Benavides, J. A. Benítez-Andrades, H. Alaiz-Moretón, and I. García-Rodríguez, Diabetes detection using deep learning techniques with oversampling and feature augmentation. Computer Methods and Programs in Biomedicine, vol. 202, pp. 105968, 2021.

[32] B. Chen, S. Xia, Z. Chen, B. Wang, and G. Wang, RSMOTE: A self-adaptive robust SMOTE for imbalanced problems with label noise. Information Sciences, vol. 553, pp. 397-428, 2021.

[33] A. Sarica, A, Cerasa, and A. Quattrone, Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Frontiers in Aging Neuroscience, vol. 9, pp. 329, 2017.