engineering proceedings

MDPI

Proceeding Paper

Development of a Predictive Model for Mild Cognitive Impairment in Parkinson's Disease with Normal Cognition Using Kernel-Based C5.0 Machine Learning Blending: Preliminary Research †

Haewon Byeon

Department of Medical Big Data, College of AI Convergence, Inje University, Gimhae 50834, Korea; bhwpuma@naver.com; Tel.: +82-10-7404-6969

† Presented at the 2nd International Electronic Conference on Applied Sciences, 15–31 October 2021; Available online: https://asec2021.sciforum.net/.

Abstract: This preliminary study mainly compared the performance for predicting mild cognitive impairment in Parkinson's disease (PDMCI) between single machine learning and hybrid machine learning. This study analyzed 185 patients with Parkinson's disease (75 Parkinson's disease) patients with normal cognition, and 110 patients with PDMCI. PDMCI, an outcome variable, was divided into "with PDMCI" and "with normal cognition" according to the diagnosis of the neurologist. This study used 48 variables (diagnostic data), including motor symptoms of Parkinson's disease, non-motor symptoms of Parkinson's disease, and sleep disorders, as explanatory variables. This study developed seven machine learning models using blending (three hybrid models (polydot + C5.0, vanilladot + C5.0, and RBFdot + C5.0) and four single machine learning models (polydot, vanilladot, RBFdot, and C5.0)). The results of this study showed that the RBFdot + C5.0 was the model with the best performance to predict PDMCI in Parkinson's disease patients with normal cognition (AUC = 0.88) among the seven machine learning models. We will develop interpretable machine learning using C5.0 in a follow-up study based on the results of this study.

Keywords: hybrid machine learning; blending approach; mild cognitive impairment in Parkinson's disease; SVM; C5.0

check for updates

Citation: Byeon, H. Development of a Predictive Model for Mild Cognitive Impairment in Parkinson's Disease with Normal Cognition Using Kernel-Based C5.0 Machine Learning Blending: Preliminary Research. *Eng. Proc.* **2021**, *11*, 18. https://doi.org/10.3390/ASEC2021-11147

Academic Editor: Nunzio Cennamo

Published: 15 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

It has been reported that mild cognitive impairment (MCI), known as the preclinical phase of dementia, may last up to seven years and appropriate therapeutic interventions in the MCI stage can delay the progression to dementia approximately five years [1]. As a result, many studies [2,3] have focused on detecting MCI, known as an intermediate stage between normal aging and Alzheimer's disease, as soon as possible. As longitudinal studies [4,5] on Parkinson's disease have reported that patients with Parkinson's disease frequently suffer from cognitive impairment, recent studies [6,7] have paid more attention to mild cognitive impairment in Parkinson's disease (PDMCI) as well as Alzheimer's MCI. Although PDMCI occurs frequently in patients with Parkinson's disease, the characteristics of PDMCI are known much less than those of Alzheimer's MCI and those of vascular MCI.

Although a number of previous studies [8,9] have reported that the most critical characteristic of PDMCI is executive function impairment due to frontal lobe dysfunction found at an early stage, it is hard to detect it only with the degree of executive function because early-stage MCI due to Alzheimer disease or vascular dementia shows executive function impairment [10]. In particular, since Parkinson's disease progresses slowly and symptoms appear little by little, patients and caregivers can perceive the cognitive problems caused by PDMCI as the cognitive frailty in the normal aging process. Therefore, it is hard to diagnose it early.

MCI is diagnosed based on an interview, evaluation of cognitive function through standardized neuropsychological tests, and brain imaging. However, brain imaging has limitations in its use for early diagnosis purposes because although it can detect the presence of cerebrovascular disease and brain atrophy, it can find them only when these symptoms are very advanced. Therefore, neuropsychological tests also evaluating cognitive function are known to be effective screening tests for detecting MCI early [11].

On the other hand, studies in the medical field have steadily predicted the risk probability or high-risk groups of a disease using data mining in recent years [12,13]. However, it is challenging to accurately predict diseases with single machine learning (learner). For example, the artificial neural network technique has the limitation of not being able to explain the derived results but it offers high prediction accuracy. On the other hand, the decision tree technique allows clinicians to easily interpret the results derived from it, but it is exposed to a higher overfitting risk than other machine learning algorithms such as SVM, the results of it can be altered by the type and order of input variables, and the accuracy of it can be lowered depending on them.

To overcome these limitations, a hybrid model combining Support Vector Machine (SVM) and decision tree model has been used recently to develop a model that has higher predictive power and explanatory power compared to single machine learning [14]. This study developed a PDMCI predictive model considering health behaviors, environmental factors, medical history, physical function, depression, and cognitive level using a hybrid model combining C-SVM and C5.0 and provided baseline data for the prevention and early management of Parkinson's disease. This preliminary study mainly compared the performance for predicting PDMCI between single machine learning and hybrid machine learning. We will develop interpretable machine learning using C5.0 in a follow-up study based on the results of this study.

2. Method

2.1. Data Source

It is a secondary data analysis study that analyzed Parkinson's Disease Epidemiologic (Parde) Data after receiving an approval (No. KBN-2019-005) from the Distribution Committee and an approval (No. KBN-2019-1327) from the Research Ethics Review Committee of the Korea Centers for Disease Control and Prevention and National Biobank of Korea. The design and administration of Parde data are described in detail elsewhere [12]. This study analyzed 185 patients with Parkinson's disease (75 Parkinson's disease) patients with normal cognition, and 110 patients with PDMCI.

2.2. Measurement

PDMCI, an outcome variable, was divided into "with PDMCI" and "with normal cognition" according to the diagnosis of the neurologist. This study used 48 variables (diagnostic data), including motor symptoms of Parkinson's disease, non-motor symptoms of Parkinson's disease, and sleep disorders, as explanatory variables.

2.3. Model Blending Based on Machine Learning

In this study, a PDMCI prediction model was developed using the blending approach (base model = SVM; meta model = C5.0). This study chose "C5.0" implemented by Kuhn et al. (2013) for the decision tree algorithm and "kernel-based machine learning (kernlab)" implemented by Karatzoglou et al. (2016) for the SVM to develop a PDMCI predictive model. The kernlab algorithm includes a polynomial kernel function (polydot), a linear kernel function (vanilladot), and a radial basis kernel function (RBFdot) that enable nonlinear SVM analysis. This study developed seven machine learning models using blending (three hybrid models (polydot + C5.0, vanilladot + C5.0, and RBFdot + C5.0) and four single machine learning models (polydot, vanilladot, RBFdot, and C5.0)). The structure of the blending model in this study is presented in Figure 1.

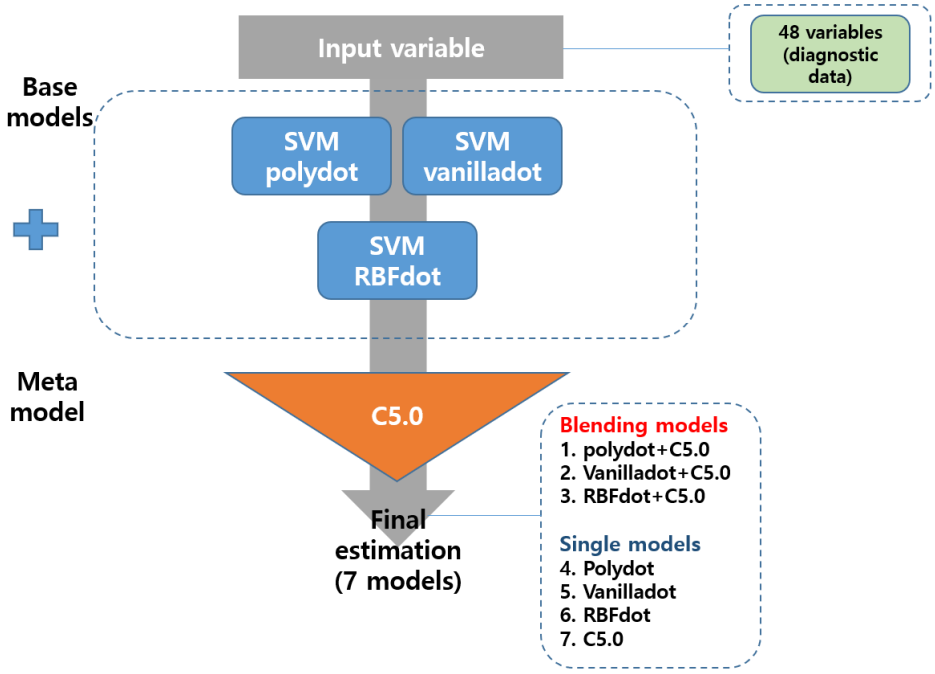

Figure 1. The structure of the prediction for PDMCI.

This study compared the predictive performance (general accuracy, F1-score, area under the curve (AUC), recall, precision) of these developed models using the 10-fold cross-validation method.

The performance evaluation of the machine learning model is different from the regression model and the classification model. In the case of the regression model, the average error value between the actual value and the predicted value is calculated, which is evaluated as MAE, MSE, and R^2. However, in the case of the classification model, the majority is calculated, which is calculated through general accuracy, F1-score, area under the curve (AUC), recall, and precision.

In this study, general accuracy, F1-score, AUC, recall, and precision are presented as performance evaluation indicators because the model is trained using binary data. In particular, recall and precision may have extreme values when data are unbalanced. Here, it is effective to use AUC or to use F1-score, which is a combination of recall and precision, as a performance evaluation. The formula of F1-SCORE is presented in Equation (1).

$$\mathrm{F1} = 2\frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \quad (1)$$

This study assumed that a model with the highest AUC was the best predictive performance. If the AUC was the same, a model with the highest F1-score was assumed as the optimal model.

3. Results

3.1. General Characteristics of Subjects

Among 185 patients with Parkinson's disease, 59.5% (108 subjects) had PDMCI. The results of chi-square test showed that PDMCI and Parkinson's disease patients with normal cognition had significantly different REM and RBD, Motor score of UPDRS, Total score of UPDRS, Global CDR, K-MoCA, K-MMSE, Sum of boxes in CDR, H&Y staging, K-IADL, and Schwab and England ADL ($p < 0.05$).

3.2. Comparing the Predictive Performance of Single Model and That of Blending Model

The results of this study showed that the RBFdot + C5.0 was the model with the best performance to predict PDMCI in Parkinson's disease patients with normal cognition

(AUC = 0.88) among the seven machine learning models. The AUC and F1-scores of the seven machine learning models analyzed in this study are presented in Figures 2 and 3, respectively.

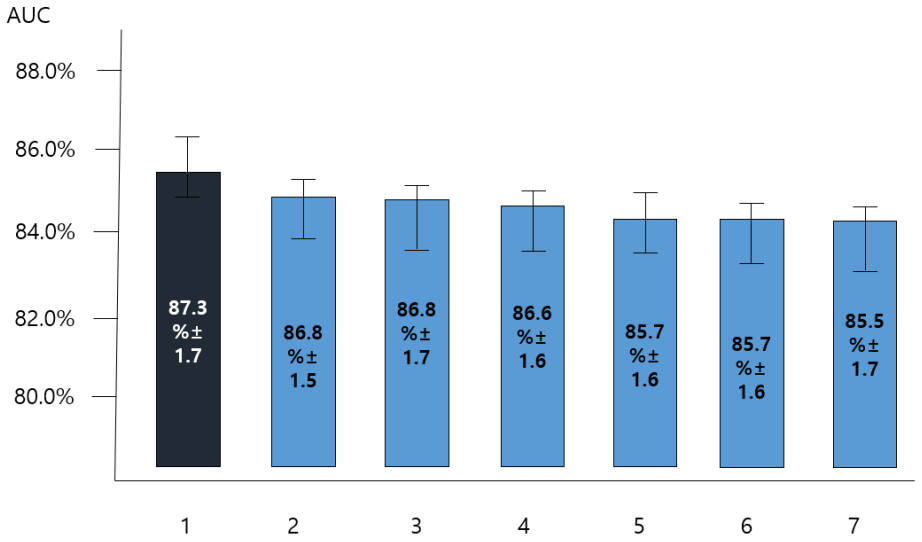

Figure 2. The comparison of AUC for seven machine learning models. 1 = RBFdot + C5.0; 2 = polydot + C5.0; 3 = vanilladot + C5.0; 4 = RBFdot + C5.0; 5 = C5.0; 6 = vanilladot; 7 = polydot.

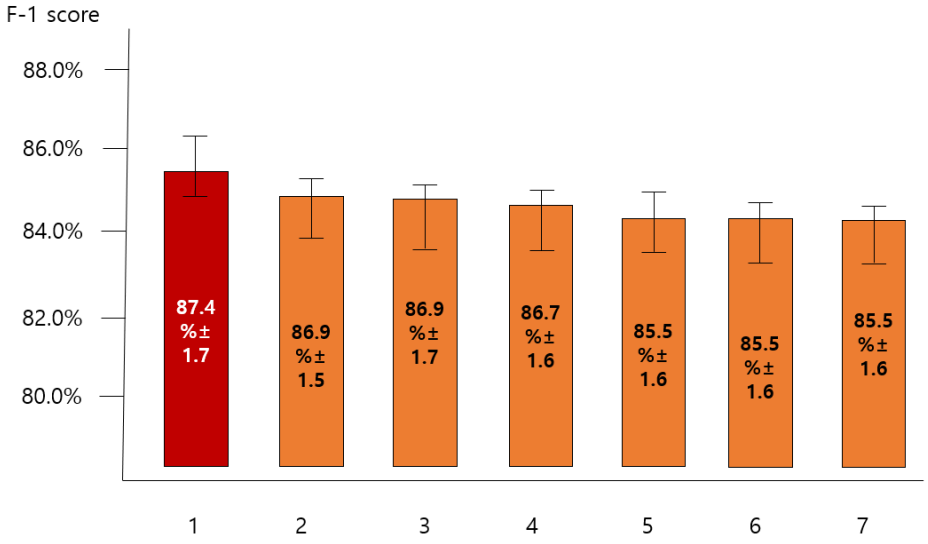

Figure 3. The comparison of F-1 score for seven machine learning models. 1 = RBFdot + C5.0; 2 = polydot + C5.0; 3 = vanilladot + C5.0; 4 = RBFdot + C5.0; 5 = C5.0; 6 = vanilladot; 7 = polydot.

4. Conclusions

The results of this study showed that the RBFdot + C5.0 was the model with the best performance to predict PDMCI in Parkinson's disease patients with normal cognition (AUC = 0.88) among the seven machine learning models. It is necessary to develop a customized screening program for detecting PDMCI in Parkinson's disease patients with normal cognition early based on the results of this study.

When developing a system to predict the morbidity of PDMCI from Parkinson's Disease with Normal Cognition in the future, it will be possible to predict more accurately with the RBFdot + C5.0 model proposed in this study than single machine learning such as SVM. We will develop a machine learning model that can explain the characteristics of high PDMCI risk groups based on the results of this study.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ASEC2021-11147/s1, S1: Development of a Predictive Model for Mild Cognitive Impairment in Parkinson's Disease with Normal Cognition Using Kernel-Based C5.0 Machine Learning Blending: Preliminary Research.

Funding: This research was funded by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, grant number 2018R1D1A1B07041091, 2021S1A5A8062526.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of National Biobank of Korea under Korea Centers for Disease Control and Prevention (protocol code KBN-2019-1327).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Brookmeyer, R.; Gray, S.; Kawas, C. Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset. *Am. J. Public Health* **1998**, *88*, 1337–1342. [CrossRef] [PubMed]
2. Gustavson, D.E.; Elman, J.A.; Sanderson-Cimino, M.; Franz, C.E.; Panizzon, M.S.; Jak, A.J.; Reynolds, C.A.; Neale, M.C.; Lyons, M.J.; Kremen, W. Extensive memory testing improves prediction of progression to MCI in late middle age. *Alzheimer's Dement. Diagn. Assess. Dis. Monit.* **2020**, *12*, e12004. [CrossRef] [PubMed]
3. Lim, S.J.; Lee, Z.; Kwon, L.N.; Chun, H.W. Medical health records-based Mild Cognitive Impairment (MCI) prediction for effective dementia care. *Int. J. Environ. Res. Public Health* **2021**, *18*, 9223. [CrossRef] [PubMed]
4. Monastero, R.; Cicero, C.E.; Baschi, R.; Davì, M.; Luca, A.; Restivo, V.; Zangara, C.; Fierro, B.; Zappia, M.; Nicoletti, A. Mild cognitive impairment in Parkinson's disease: The Parkinson's disease cognitive study (PACOS). *J. Neurol.* **2018**, *265*, 1050–1058. [CrossRef] [PubMed]
5. Domellöf, M.E.; Ekman, U.; Forsgren, L.; Elgh, E. Cognitive function in the early phase of Parkinson's disease, a five-year follow-up. *Acta Neurol. Scand.* **2015**, *132*, 79–88. [CrossRef] [PubMed]
6. Baiano, C.; Barone, P.; Trojano, L.; Santangelo, G. Prevalence and clinical aspects of mild cognitive impairment in Parkinson's disease: A meta-analysis. *Mov. Disord.* **2020**, *35*, 45–54. [CrossRef] [PubMed]
7. Wen, M.C.; Chan, L.L.; Tan, L.C.; Tan, E.K. Mild cognitive impairment in Parkinson's disease: A distinct clinical entity? *Transl. Neurodegener.* **2017**, *6*, 1–7. [CrossRef] [PubMed]
8. Martinez-Horta, S.; Kulisevsky, J. Mild cognitive impairment in Parkinson's disease. *J. Neural Transm.* **2019**, *126*, 897–904. [CrossRef] [PubMed]
9. Roheger, M.; Kalbe, E.; Liepelt-Scarfone, I. Progression of cognitive decline in Parkinson's disease. *J. Parkinsons Dis.* **2018**, *8*, 183–193. [CrossRef] [PubMed]
10. Zokaei, N.; Husain, M. Working Memory in Alzheimer's Disease and Parkinson's Disease. In *Processes of Visuospatial Attention and Working Memory*; Springer: Cham, Switzerland, 2019.
11. Biundo, R.; Weis, L.; Pilleri, M.; Facchini, S.; Formento-Dojot, P.; Vallelunga, A.; Antonini, A. Diagnostic and screening power of neuropsychological testing in detecting mild cognitive impairment in Parkinson's disease. *J. Neural Transm.* **2013**, *120*, 627–633. [CrossRef] [PubMed]
12. Byeon, H. Application of machine learning technique to distinguish Parkinson's disease dementia and Alzheimer's dementia: Predictive power of Parkinson's disease-related non-motor symptoms and neuropsychological profile. *J. Pers. Med.* **2020**, *10*, 31. [CrossRef] [PubMed]
13. Przybyszewski, A.W.; Chudzik, A.; Szlufik, S.; Habela, P.; Koziorowski, D.M. Comparison of different data mining methods to determine disease progression in dissimilar groups of Parkinson's patients. *Fundam. Inform.* **2020**, *176*, 167–181. [CrossRef]
14. Sun, L.; Fu, S.; Wang, F. Decision tree SVM model with fisher feature selection for speech emotion recognition. *Eurasip J. Audio Speech Music Process.* **2019**, *2019*, 1–14. [CrossRef]