aFeature Selection of EEG and Heart Rate Variability Indexes for Estimation of Cognitive Function in the Elderly

KENTAROU KANAI[†1] YURI NAKAGAWA[†1] MIDORI SUGAYA[†1]

Abstract: In recent years, the number of dementia patients has been increasing, and the number of residents in nursing homes has been rising, increasing the burden on nursing care workers. However, the current situation is that the cognitive level of elderly patients is not sufficiently assessed, and thus, caregivers are not able to respond to the patients accordingly. In this study, we aimed to estimate cognitive function by a machine learning model using simple electroencephalograph (EEG) and heart rate monitor data as a simple and objective method of estimating cognitive function, and to evaluate cognitive level using this model. However, there are many features that can be calculated from electroencephalography and heart rate monitors, and it is not clear which features should be machine-learned. Therefore, in this paper, we used feature selection to identify important features for building a model for estimating cognitive function in the elderly. The results showed that the importance of the HRV index was higher when the mutual information content was used, and the importance of the EEG index was higher when the random forest variable importance was used.

Keywords: feature-selection, cognitive, EEG, heart rate

1. Introduction

1.1 Background

According to a survey by Japan's Cabinet Office, the number of people with dementia was 18% of the 65-year-old elderly population in 2020 and is projected to increase to 25% by 2040 [1]. Currently, there is no exact cure for dementia [2], and early detection of dementia and delaying its progression are extremely important [3]. When dementia occurs, it affects the individual, family members, and a wide range of other people. Therefore, early detection methods such as fMRI and PET are provided in hospitals. While these methods can accurately measure brain information, they are expensive and time-consuming. Therefore, it is desirable to use a simpler method that can be used in outpatient diagnoses to understand the cognitive function of elderly people and to measure the progression of dementia. Currently, many tests such as interview, MRI (Magnetic Resonance Imaging), and neuropsychological tests such as MMSE (Mini Mental State Examination) are required to diagnose cognitive function of the elderly in large facilities such as hospitals. However, these methods require a large number of tests to diagnose cognitive function. Thus, while these diagnostic methods can accurately measure cognitive function, they are burdensome for the elderly and difficult to measure in a simple manner. As the number of people with dementia is expected to increase in the future, there is a need for a simple method to measure brain information and to understand cognitive function.

1.2 Related Research and Issues

A method for estimating cognitive function using machine learning from physiological information such as electroencephalogram (EEG) and heart rate variability(HRV) has been proposed as a simple method for understanding cognitive function in the elderly [4,5].

EEG and HRV are focused on because EEG is an electrical signal generated during brain activity and can be obtained noninvasively and inexpensively compared to other methods of obtaining brain information, such as fMRI and PET [4]. In addition, EEG is a useful measurement method for estimating cognitive function in the elderly because it directly reflects the neural function of the cerebral cortex [4]. Next, HRV refers to the interval of heartbeat time and can be measured easily and noninvasively among autonomic nervous system indexes [5]. In addition, the indexes that can be created from HRV have a significant relationship with cognitive function in the elderly, with almost all indexes showing a negative relationship [5], supporting its use in the evaluation of cognitive function.

For these reasons, EEG and HRV are considered to be useful physiological information for understanding cognitive function. Furthermore, there are research that use these data to understand cognitive function through machine learning and apply them to prediction and other purposes.

Je-Eon et al. used EEG and neuropsychological tests to understand cognitive functions [6]. Specifically, they constructed a machine learning model to measure the cognitive function of elderly people using the Verbal Fluency Test (VFT), a neuropsychological test, and an EEG that can acquire data from a single channel in the frontal lobe. The VFT is a test to measure cognitive function based on how many words in a specified category (e.g., animals, vehicles, etc.) can be said in 60 seconds, and is often used in outpatient diagnosis of dementia [7]. Je-Eon et al. constructed a machine learning model to estimate the cognitive function of the elderly using SVM (Support Vector Machine) with EEG as the explanatory variable and VFT results as the objective variable. The accuracy of the model was approximately 65%. One of the challenges of this research is that while a single channel of EEG acquired from the frontal lobe can be easily measured, it is difficult to obtain information on the entire brain. However, multi-channel electroencephalographs, which can obtain information on the entire brain, take as long as 30 minutes per fitting and are a significant burden on the elderly.

† 1 Shibaura Institute of Technology

Next, Eaman et al. constructed a machine learning model for estimating cognitive function using HRV [8]. As a result, the accuracy of the constructed machine learning model was about 82%. One of the issues in this research is that it is unclear whether HRV alone can be used for advanced measurement, because cognitive function is closely related to the brain, for example, dementia is a brain disease.

The problem with these research is that the accuracy of machine learning using EEG alone to understand the cognitive function of the elderly is low. In addition, when machine learning is used to understand the cognitive function of the elderly using only HRV, it is unclear whether advanced measurement can be made using only HRV, because cognitive function is closely related to the brain, as dementia is a disease of the brain.

Therefore, although there are many research that estimate the cognitive function of the elderly using EEG only or HRV only, there are few research that estimate the cognitive function of the elderly by multiple physiological indexes.

2. Objective・Proposal

To solve these problems, the purpose of this research was to estimate the cognitive function of the elderly in a simple and objective manner. In addition to HRV indexes, we also aim to estimate cognitive function using a machine learning model based on EEG.

However, there are many features that can be calculated from EEG and HRV indexes, and it is not clear which of them should be used for machine learning. For this reason, the objective of this research was to select the features. To achieve this, we used feature selection from the acquired EEG and HRV to identify important features for cognitive function estimation.

3. Experiment

In this research, an experiment was conducted to construct an estimation model of cognitive function of the elderly by obtaining the results of neuropsychological test, EEG and HRV of the elderly under normal conditions. The subjects of the experiment were seven elderly people, one male and six female, aged 70 years or older, residing in a nursing home.

In this experiment, EEG and heart rate monitors were used to acquire EEG and HRV. The EEG was measured using NeuroSky's MindWave mobile2, which can acquire EEG from one channel near the left frontal lobe. HRV was measured using Union Tool's myBeat heart rate monitor, which can be worn around the solar plexus below the heart. myBeat can measure HRV.

The Mini Mental State Examination (MMSE) was used as a neuropsychological test to assess cognitive function.

The MMSE is a neuropsychological test that asks memory and behavioral questions and calculates a score according to the number of questions answered. The test is scored on a 30-point scale, with a score of 28 or higher representing people without dementia, a score of 24 to 27 representing people with mild dementia, and a score of 23 or lower representing people with dementia. Examples of questions in this test include memory questions such as "What day is today? " , "What prefecture is this? " and "What prefecture are we in?" Behavioral questions include "Please read what is written on the paper and perform the action described" and "Please repeat what I just said.

The experimental procedure is as follows.

1) Description of the experiment. Wearing the equipment (electroencephalograph, heart rate monitor).

2) 5-minute resting EEG with closed eyes (normal condition). Acquisition of HRV.

3) Perform MMSE.

The experiment was conducted for two days, and data were obtained from seven subjects. The EEG and HRV data for the 5-minute resting period only were used for future analysis.

Fig 1. Experiment in progress

4. Calculation of the importance of the indexes

4.1 Analytical method

As an analytical method, feature selection was used with the aim of identifying features that are important in understanding the cognitive function of the elderly. Two methods were used for the analysis: mutual information content and random forest variable importance [9,10]. These two methods have been used in feature selection for physiological indexes such as EEG. Therefore, these two methods were also employed in this research [11,12].

4.2 Calculation of Indexes

The following indexes were created from the acquired EEG and HRV for the construction of a machine learning model for estimating cognitive function in the elderly.

Ten EEG indexes were calculated from the acquired EEG. The details are shown in Table 1.

Table 1. Calculated EEG index

Index	Bandwidth (Hz)	interpretation
δ	1-3	Deep sleep, unconscious
θ	4-7	Intuitive, Creative, Dreaming
Low-α	8-9	Relaxed, Calm
High-α	10-12	Relaxed and focused

Low-β	13-17	Thinking, awareness of self and environment
High-β	18-30	Alert, agitation
Low-γ	31-40	Memory, higher mental activity
Mid-γ	41-50	Visual information processing
total power	1-50	Sum of each frequency band

In addition to the above, moving averages were performed for the EEG indexes with window widths of 5, 10, 15, 20, 25, and 30. The moving averaged indexes are hereinafter referred to as MAx (x=window width). The relative power for each frequency band divided by the total power was also calculated. The relative power for each frequency band is shown below with "relative" in front of the frequency band. A total of 119 EEG indexes were created. Then, 12 HRV indexes were calculated from the acquired HRV. The details are shown in Table 2.

Table 2. Calculated HRV index

Index	Calculation Method	interpretation
IBI	Heart rate interval, in ms	-
IBI_diffrence	Difference from previous IBI	-
HR	Number of oscillations per minute	Nervousness, Calmness
pNNx(x=10,20, 30,40,50)	Percentage of adjacent IBIs with absolute values greater than x ms	Pleasant/ Unsatisfactory Parasympathetic
SDNN	Standard deviation of IBIs	Sympathetic, Parasympathetic
RMSSD	Root-mean-square of the difference between adjacent IBIs	Parasympathetic
SDNN/RMSSD	SDNN/RMSSD	Sympathetic
CVNN	Coefficient of variation of IBI	Sympathetic, Parasympathetic

The number of EEG indexes was 119 and that of HRV indexes was 12, which is about 10 times larger than that of the other indexes. Therefore, we used mutual information content, a feature selection method, to find the optimal window width of the moving average. Specifically, feature selection was performed using all indexes, and the average value was calculated for each window width. As a result, the window width with the highest mean value was adopted for future analysis.

The results are shown in Table 3.

Table 3. Mutual information content of EEG indexes per moving average window width.

	MA0	MA5	MA10	MA15	MA20	MA25	MA30
δ	0.352	0.637	0.687	0.692	0.694	0.694	0.694
θ	0.011	0.044	0.041	0.099	0.108	0.139	0.170
Low-α	0.123	0.291	0.368	0.452	0.536	0.583	0.606
High-α	0.036	0.080	0.100	0.163	0.199	0.238	0.268
Low-β	0.024	0.034	0.055	0.044	0.089	0.061	0.091
High-β	0.004	0.062	0.136	0.189	0.255	0.245	0.266
Low-γ	0.157	0.415	0.520	0.584	0.611	0.612	0.632
Mid-γ	0.062	0.159	0.191	0.277	0.312	0.340	0.376
Average	0.096	0.215	0.262	0.313	0.351	0.364	0.388

Table 3 shows that window width 30 has the highest value of 0.388. Therefore, in future analyses of EEG indexes, 17 indexes will be used, which are moving-average values with a window width of 30.

5. Results

5.1 Results of neuropsychological test

Table 4 shows the distribution of each class based on the results of the neuropsychological test conducted in the experiment.

Table 4. Results of neuropsychological testing

classification	the number of people
Without dementia	1
Mild dementia	3
Dementia	3

In this study, the results in Table 1 are used as the objective variable for feature selection.

5.2 Results of feature selection

The results of feature selection based on the amount of mutual information and the variable importance of random forests are shown, using the 29 created EEG and HRV indices as explanatory variables and the MMSE results as objective variables.

Figure 2 shows the results of the mutual information content. The importance of each index is shown in blue for the EEG index and orange for the HRV index.

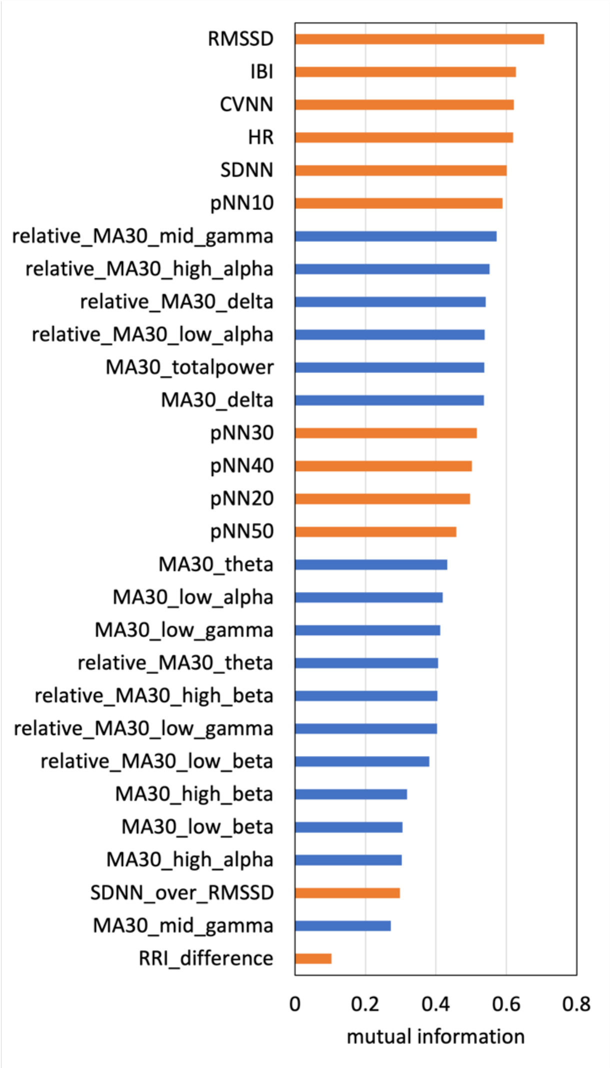

Fig 2. Importance calculation results
using mutual information content

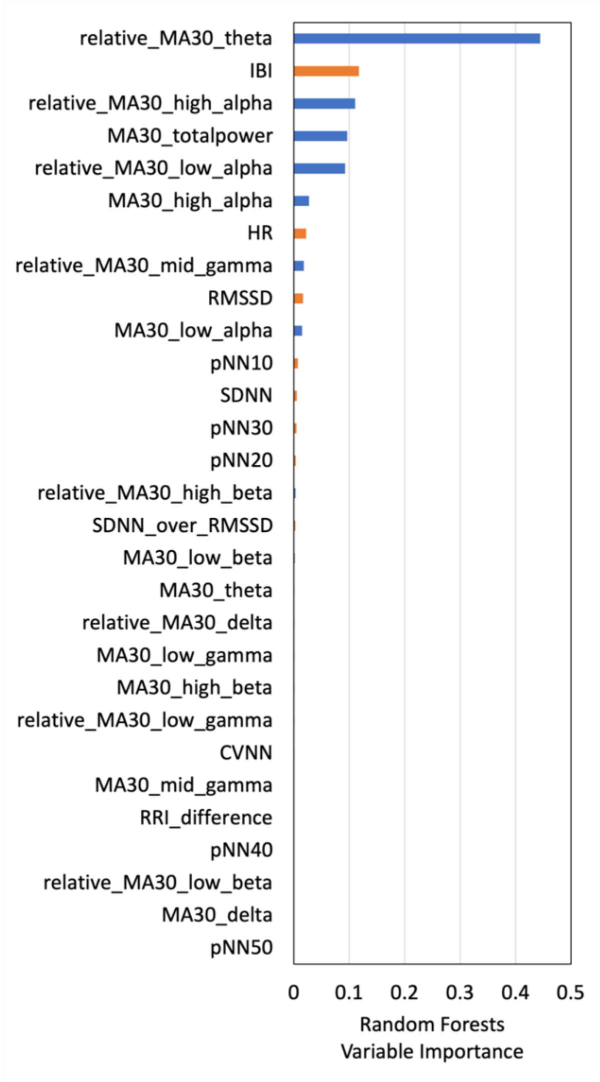

Fig 3. importance calculation results
using random forest variable importance

Figure 2 shows that the importance of the HRV index is higher than the importance of the EEG index in estimating the cognitive function of the elderly when using the mutual information content, because all of the indexes up to 6th place are HRV indexes.

In addition, when looking at the EEG indexes, the importance of the relative power of each frequency band was found to be higher than the importance of the absolute value of the band power for the six EEG indexes: δ, low_α, high_α, low_β, high_β, and mid_γ. This suggests that relative power may be more important than the absolute value of each frequency band in estimating cognitive function in the elderly.

Figure 3 shows the results of the random forest variable importance.

Figure 3 shows that for the important indexes in estimating the cognitive function of the elderly when using the variable importance of Random Forest, the EEG index have 4 indexes, and the HRV index has 1 index when looking at the top 5 indexes. Therefore, it was found that the importance of the EEG index was higher than that of the HRV index.

In addition, when looking at the EEG indexes, the importance of the relative power of each circumferential wavenumber band was found to be higher than the importance of the absolute value of the band power for the six indexes of δ, θ, low_α, high_α, high_β, and mid_γ. Furthermore, the top EEG indexes were θ, low α, high α, and low frequency band indexes.

For the heart rate index, the importance of IBI was higher than that of HR for both feature selection methods. Therefore, it was indicated that the index based on the heart rate interval is more important than the index based on the number of beats, such as HR, in estimating cognitive function in the elderly.

Finally, when looking at the relative power of θ for the mutual information and random forest variable importance, the importance of the random forest variable was more than four times higher than that of the other variables, whereas the importance of the mutual information was not so high. In addition, the importance of the HRV indexes were higher for the mutual information content, and the importance of the EEG index was higher for the random forest variable importance. Thus, the two feature selection methods produced very different results. Therefore, it is necessary to confirm which feature selection method is more useful by constructing a model for each of the indexes of high importance of the mutual information and random forest variables.

6. Conclusion

The results of EEG, HRV, and neuropsychological test were used to calculate the most important indexes for estimating cognitive function in the elderly. The results showed that the HRV index was more important in terms of the amount of mutual information. In the random forest variable importance, the EEG index was more important than the HRV index.

For the EEG indexes, the importance of low frequency band indexes such as θ, Low_α, and High_α was higher.

For the HRV index, the heart rate interval was found to be more important than the heart rate in estimating the cognitive function of the elderly.

Since the important features differ greatly depending on the feature selection method, we plan to construct a model using the results of the mutual information content and the variable importance of random forests respectively, and to construct a model combining the results of the mutual information content and the variable importance of random forests in the future.

Acknowledgment

The authors would like to thank the staff members of the “Care Home for the Elderly Sawayaka Kinu no Sato Shinshu Okaya” for their cooperation in this experiment, the subjects and their families, Mr. Yosuke Uchida and Mr. Eiji Otsuka of TPR Corporation for their support in conducting this experiment.

Reference

[1] Cabinet Office. Health and Welfare of the Elderly.https://www8.cao.go.jp/kourei/whitepaper/w-2016/html/gaiyou/s1_2_3.html.(see February 14, 2022).
[2] R Overshott, A Burns. “Treatment of dementia”. J Neurol Neurosurg Psychiatry, Volume 76, Issue suppl 5, P53-59, 2005.
[3] R. C. Petersen et al. "Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review) ". Neurology, vol 56, pp1133-1142, 2001.
[4] Soraia M. Alacao, Manuel J. Fonseca, Senior Member, IEEE. Emotion Recognition Using EEG Signals: A Survey. IEEE Transactions on Affective Computing, pp 1- 12 June 2017.
[5] Eaman Alharbi,Akram Alomainy,Janelle M. Jones.”Detecting C ognitive Decline in Early Alzheimer’s Patients Using Wearabl e Technologies”. IEEE International Conference on Healthcare Informatics (ICHI),2020.
[6]Je-Eon Lee, Jaeho Park, Sujeong You.”Combining single-chann el EEG measurement and verbal fluency test − A groundwork for ambulatory diagnosis of dementia”. 14th International Conf erence on Ubiquitous Robots and Ambient Intelligence (URAI), 2017.
[7]Marcos Lopes et al.,“Semantic Verbal Fluency test in dementia Preliminary retrospective analysis”. Dement Neuropsychol 2009, pp315-320.
[8] Eaman Alharbi,Akram Alomainy. “Machine Learning approach to Predict Cognitive Performance using HRV”.2022 2nd Interna tional Conference on Computing and Information Technology (ICCIT),pp25-27. 2022.
[9] C.E.Shannon.” A Mathematical Theory of Communication”.The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, 1948.
[10] Leo Breiman.Random Forests.Machine Learning,volume45, pp5–32,2001.
[11] Suzuki Kei, Laohakangvalvit Tipporn, Matsubara Ryota, Sugaya Midori, “Constructing an Emotion Estimation Model Based on EEG/HRV Indexes Using Feature Extraction and Feature Selection Algorithms”, Sensors, 21(9), 2910, 2021.
[12] Ashima Khosla, Padmavati Khandnor, Trilok Chand. “A comp arative analysis of signal processing and classification methods for different applications based on EEG signals”, Biocybernetic s and Biomedical Engineering, Volume40, Isuue 2,Pages 649-690,2020.

