

Article

Early Diagnosis of Dementia from Clinical Data by Machine Learning Techniques

Aram So ¹, Danial Hooshyar ¹, Kun Woo Park ² and Heui Seok Lim ^{1,*}

¹ Department of Computer Science and Engineering, Korea University, Seoul 100-744, Korea; aram@korea.ac.kr (A.S.); danial.hooshyar@gmail.com (D.H.)

² Department of Neurology, College of Medicine, Korea University, Seoul 100-744, Korea; kunu@korea.ac.kr

* Correspondence: limhseok@korea.ac.kr; Tel.: +82-232-902-396

Received: 2 June 2017; Accepted: 20 June 2017; Published: 23 June 2017

Abstract: Dementia is the most prevalent degenerative disease in seniors in which progression can be prevented or delayed by early diagnosis. In this study, we proposed a two-layer model inspired by the method used in dementia support centers for the early diagnosis of dementia and using machine learning techniques. Data were collected from patients who received dementia screening from 2008 to 2013 at the Gangbuk-Gu center for dementia in the Republic of Korea. The data consisted of the patient's gender, age, education, the Mini-Mental State Examination in the Korean version of the CERAD Assessment Packet (MMSE-KC) for dementia screening test, and the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD-K) for the dementia precise test. In the proposed model, MMSE-KC data are initially classified into normal and abnormal. In the second stage, CERAD-K data are used to classify dementia and mild cognitive impairment. The performance of each algorithm is compared with that of Naive Bayes, Bayes Network, Bagging, Logistic Regression, Random Forest, Support Vector Machine (SVM) and Multilayer Perceptron (MLP) using Precision, Recall and F-measure. Comparing the F-measure values of normal, mild cognitive impairment (MCI), and dementia, the MLP was the highest in the F-measure values of normal with 0.97, while the SVM appear to be the highest in MCI and dementia with 0.739. Using the proposed early diagnosis model for dementia reduces the time and economic burden and can help simplify the diagnosis method for dementia.

Keywords: early diagnosis of dementia; machine learning; normal; mild cognitive impairment

1. Introduction

Quality of life has increased with the development of medical technology, and as the average human lifespan increases the senior population is growing. Additionally, the pace of aging is accelerating. Countries today are facing an aging society, which poses many changes and challenges for society [1,2]. In particular, the number of dementia patients is increasing because of the increase of the senior population. Dementia is the most prevalent degenerative disease in seniors. There are 47.5 million people living with dementia around the world, a majority of whom (58%) live in middle- and low-income countries. Each year brings 7.7 million new cases of dementia [1]. The number of dementia patients is expected to more than triple by 2050. As the number of dementia patients increases dramatically, the socioeconomic, psychological, physical, and economic burdens for dependents' families are also increasing [2].

Dementia can be sorted into dementia caused by Alzheimer's disease, cerebrovascular dementia, hypothyroidism, benign brain tumors, etc. Alzheimer's accounts for about 60% to 70% of dementia patients; it is caused by aging, family history, and depression. However, the presence of Alzheimer's disease can serve to delay the progression of dementia when it leads to early diagnosis. Cerebrovascular dementia, which affects about 20% to 30% of demented patients, is caused by diseases such as

hypertension, heart disease, diabetes, arteriosclerosis, cerebral hemorrhage, and cerebral infarction. Cerebrovascular dementia can be prevented through risk factor management, and can be treated by medicine. Other dementia can be treated by surgery, such as removing the thyroid or benign brain tumor.

Previous research on dementia was focused on treatment and care after the onset of the disease. However, as mentioned before, early diagnosis may delay the progression of dementia [3].

Generally, there are three stages to dementia diagnosis. The first stage is a screening test for cognitive ability using the MMSE-KC (Mini-Mental State Examination in the Korean version of the CERAD Assessment Packet). The second stage involves performing neuropsychological assessment using CERAD-K (the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease) for those who are not diagnosed as normal in the screening test. The final stage is to diagnose (R/O) for dementia or mild cognitive impairment (MCI) by doctor consultation and carer interview. After the third stage, the suspected patients are definitively diagnosed using MRI or CT and blood tests in hospital. As a result, patients are classified into categories of normal, MCI, and dementia.

In this paper, we propose a two-layer model for the early diagnosis of dementia, inspired by the diagnosis approach used in dementia support centers and using machine learning methods. The first layer is a screening test to classify subjects as normal or abnormal, while the second layer is close examination, classifying cases as MCI or dementia.

In the first stage, data preprocessing is performed based on the MMSE-KC data. The next step is to select the required features. Once the feature selection is completed, the data are learned by the selected features, and classified into normal and cognitive decline groups. Finally, the first step classifies the normal group. In the second stage, CERAD-K data are learned, using machine learning algorithms, for classifying MCI and dementia.

Therefore, the structure of the model is similar to the existing dementia screening method, and its effect is simplifying the dementia screening process.

The data were collected from patients who visit and are tested at the Gangbuk-Gu center for dementia in Seoul, Republic of Korea. We collected patient information such as age, gender, education, and test results using MMSE-KC and CERAD-K. To these data we applied machine learning techniques, which are useful for data analysis and are used in various domains. We used supervised learning algorithms such as Support Vector Machine (SVM), Naive Bayes, Multilayer Perceptron (MLP), Bayesian Network, Bagging, Logistic Regression, Random Forest evaluation method using F-measure, precision and recall.

We initially examined the influence of each feature through feature selection using chi-squared and information gain algorithms. As a result, MLP, SVM and logistic regression showed the highest F-measure value of normal, MCI, and dementia, respectively.

This paper is organized as follows: Section 2 explains the existing diagnosis methods of dementia; Section 3 explains the architecture of the proposed prediction model for early diagnosis of dementia; and Section 4 explains results and discussions. Finally, Section 5 illustrates the conclusions.

2. Related Works

Dementia, an illness of the brain, attacks cognitive activities such as memory, rationality, and thought. It is caused either by old age or traumatic injury, with approximately 60–70% of cases attributable to Alzheimer's disease [4]. Dementia increases in severity the longer it goes undiagnosed. The process of diagnosis involves three steps: the first involves consulting a physician; the second consists of completing an array of neuropsychological tests; the third involves an MRI scan [5]. This paper addresses the early diagnosis of dementia by means of neuropsychological testing in tandem with demographic information. Commonly used neuropsychological measures include the Mini-Mental State Examination (MMSE), the Consortium to Establish a Registry for Alzheimer's Disease (CERAD), the Blessed Orientation-Memory-Concentration Test (BOMC), the Montreal Cognitive Assessment (MoCA), a brief informant interview to detect dementia (AD8), and

the General Practitioner Assessment of Cognition (GPCOG), with each presenting certain advantages and limitations. MMSE and CERAD are currently most used, since they can be administered regardless of the subject's gender, education, culture or religion [5–8].

MMSE-KC (Mini-Mental State Examination in the Korean version of the CERAD Assessment Packet) is used to screen and measure impairment of cognitive function. The MMSE tests and scores six domains (1) orientation, (2) registration, (3) attention and calculation, (4) recall, (5) language and (6) constructional ability [9,10]. Tables 1 and 2 show the contents of the neuropsychological assessment tests used in the screening test (MMSE-KC) and the precise examination (CERAD-K) used in this paper.

Table 1. Mini-Mental State Examination in the Korean version of the CERAD (Consortium to Establish a Registry for Alzheimer's Disease) Assessment Packet (MMSE-KC).

Category	Description
Orientation to time	From broadest to most narrow. Orientation to time has been correlated with future decline
Orientation to place	From broadest to most narrow. This is sometimes narrowed down to streets, and sometimes to floor
Registration	Repeating named prompts
Recall	Registration recall
Attention and calculation	Serial sevens, or spelling “world” backwards. It has been suggested that serial sevens may be more appropriate in a population where English is not the first language
Language	Naming a pencil and a watch
Repetition	Speaking back a phrase
Complex commands	Draw pentagonal overlap.
Three-stage commands	Give the patient a piece of plain blank paper and ask him to follow your instructions
Understanding and decision	Understanding and decision

Table 2. Neuropsychological Testing (Korean version of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD-K)).

Category	Description
Word fluency	Enumerate in one minute as many instances of “animal” as possible
Boston naming	Respond to the name of the picture shown
MMSE-KC	Table 1
Memory function	Trial 1 Recall after knowing 10 noun words
	Trial 2 Word list (Trial 1) recall
	Trial 3 Present 10 new words in a word list (Trial 1) and make sure that they are distinguished from the old words
Visuospatial function	Trial 1 Transcribe four line drawings of advancing difficulty: circle, trapezoid, overlapping quadrilaterals, cube
	Trial 2 Repeat Trial 1
Mental flexibility	Number Draw a line that sequentially connects randomized numbers
	Number & Letter Link numbers and letters in an alternating sequence

CERAD-K (the Korean version of the CERAD neuropsychological assessment battery) is mainly used for scrutiny. CERAD began with researchers at sixteen major Alzheimer's research centers in the United States and was developed for standard diagnosis and evaluation of Alzheimer's patients [11,12].

CERAD can examine areas that are more intense than MMSE, such as language fluency, verbal memory ability, time span configuration ability, and depression.

Because analysis and decision-making about the results from these sorts of tests depend on the inclinations of the psychologist (and thus human error cannot be avoided), machine-based analysis and data mining approaches have been widely used to alleviate inconsistencies. In this paper, machine learning algorithms are explored to determine if the analysis of neuropsychological and demographic data can be automated for the early diagnosis of dementia. According to Chen and Herskovits [13] in their study of various statistical and machine learning methods, a Bayesian-network classifier and a SVM performed best in assessing participants afflicted by little or no dementia. In a study conducted by Joshi et al. [4], machine learning and neural network methods were used for classifying dementia states to improve accuracy over current dementia screening tools, MMSE and the Functional Activities Questionnaire. The findings showed that the accuracy can be optimized by combining both the tests along with machine learning and neural network.

Trambaiolli and Lorena [3] previously used electroencephalography (EEG) data to classify patients with normal cognition and Alzheimer's or MCI by learning the EEG pattern of Alzheimer's patients using the SVM algorithm. As a result, EEG Epochs showed a high accuracy (79.9%) and the SVM result was about 87%. Williams and Weakley [14] compared the CDR (Clinical Dementia Rating) score and the method of screening dementia using Naive Bayes, Decision Tree, Neural Network, and SVM. The results of the evaluation of the severity of dementia showed that Naive Bayes was the most accurate and SVM had the lowest accuracy. Cho and Chen [15] proposed a hierarchical double layer structure for the early diagnosis of dementia. This is a model that predicts early diagnosis of dementia using a Bayesian network in the top-layer after diagnostic prediction with FCM and PNN algorithm in the base-layer when a cognitive test such as MMSE and CERAD is performed. In this model, the accuracy of FCM and PNN was 74% and 69%, respectively, but MCI and dementia were not well classified when comparing normal, MCI, and dementia. Shanklea and Mani [16] performed CDR prediction using machine learning method and electronic medical records. For Naive Bayes, the accuracy was the highest, while for the other algorithms, it was lower than Bayesian, but it was about 70% accurate.

The diagnosis of dementia consists in large part of assessing different cognitive abilities. As such, physicians frequently interpret test results in conflicting ways: this represents a major impediment to attaining high accuracy with machine learning algorithms in the absence of a specified model. In contrast to the aforementioned studies, we advance a two-tiered hierarchical approach for evaluating and making distinctions between normal, MCI, and early dementia. This approach is derived from the dementia support center's diagnostic method (a combination of cognitive screening, neuropsychological evaluation, and early diagnosis). In this research we aim to use neuropsychological and demographic information in order to predict normal, MCI, and dementia within our proposed model by applying seven frequently used machine learning models: Naive Bayes, Bayes Network, Begging, Logistic Regression, Random Forest, SVM, and MLP. This method offers diagnostics which are at once intuitional and also far-reaching.

3. Prediction Model for Dementia Diagnosis

3.1. Architecture of the Proposed Model

In this paper, we propose a model that learns data using a machine learning algorithm and classifies data into normal, MCI, and dementia. The proposed model is a two-level hierarchical model similar to the dementia diagnosis method used in the dementia support center. The structure of the model is as follows. In the first stage, we classify a normal group and cognitive decline group. In the second stage, we classify a MCI group and a dementia group.

In the first stage, data preprocessing is performed based on the MMSE-KC data. The data preprocessing process removes missing or incorrectly entered data. In addition, due to differences in data range of each attribute (which may affect machine learning algorithms), normalization is

performed to set the range of data to 0~1. The next step is to see how each feature influences the classification result through feature selection and select the required features. Once the feature selection is completed, the data are learned by the selected features, and classified into normal and cognitive decline groups. Finally, the first step classifies the normal group.

In the second stage, CERAD-K data are learned for classifying MCI and dementia. The preprocessing process and feature selection process are the same as in the first stage. After the completion of data preprocessing and normalization and feature selection, machine learning algorithms are used to classify MCI and dementia.

In this paper, performance evaluation was performed by using various algorithms in data learning and classification model generation. The proposed model is shown in Figure 1.

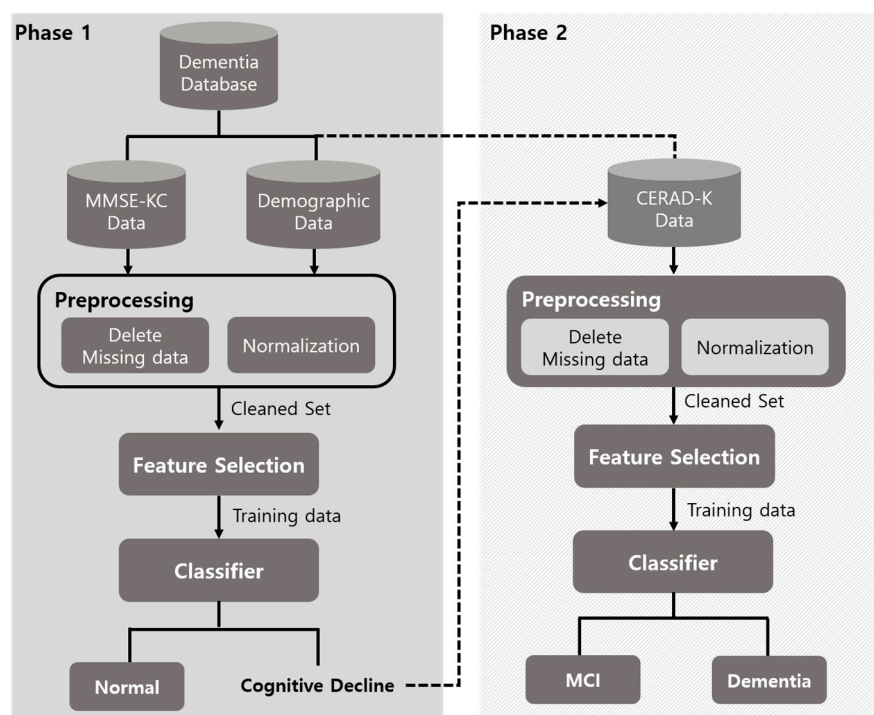


Figure 1. The architecture of the proposed model.

3.2. Data Collection

The data used in the study were collected from people who visited the dementia center in Gangbuk-Gu, Seoul, from 2008 to 2013 and received a screening test.

The data collection method is as follows. First, MMSE-KC examines the cognitive decline of the patient. If the resulting diagnosis indicates cognitive decline, CERAD-K would be further conducted. After the precise examination, it is decided whether or not to take a doctor's examination, consult with the doctor, decide whether to be confirmed at the hospital or participate in the program run by the center.

Two types of data were used in this study. The collected data consists of 14 attributes for Phase 1 data and 31 attributes for Phase 2 data. First, the data used in Phase 1 were gender, age, education, and MMSE-KC scores. MMSE-KC results (normal, cognitive decline) were used as class data for classification. When performing neuropsychiatric treatment, the age, education level, physical condition and basic cognitive ability of the subject should be considered. In addition to MMSE-KC score data, demographic data such as patient's sex, age, and education level were collected. The second data is the data used in Phase 2, which classifies the normal in Phase 1 and adds CERAD-K data to

the data, which is not classified as normal. In Phase 2, we used the data of the patients who were confirmed (dementia or dementia high risk) visiting the hospital after the final examination.

Data from Phase 1 consisted of data from a total of 14,000 patients, 9799 in the normal group and 4201 in the cognitive decline group. The mean age of the patients was 73 years old, 72 years in the normal group, and 74 years in the cognitive impairment group. The MMSE-KC score was 25 points for the normal group and 18 points for the cognitive decline group, which was about 7 points different from the normal group. The overall average was 23 points.

In Phase 2, the average age of all patients was 76 years, the difference average age gap between MCI and dementia patients was 5 years. When measuring cognitive ability, the level of the patient's education also affected the results, but MCI and dementia did not show much difference. The MMSE-KC score was 17 points out of 30 as a whole, with an average of 20 points for MCI patients and a dementia score of 15, slightly lower than the average. Details are shown in Tables 3 and 4.

Table 3. Statistical Information for the Collected MMSE-KC (Phase1).

Variables	All Patients		Patients (Normal)		Patients (Cognitive Decline)	
	Range	Descriptive statistics	Range	Descriptive statistics	Range	Descriptive statistics
Gender	Male/Female	Male: 4808 Female: 9192	Male/Female	Male: 2837 Female: 6962	Male/Female	Male: 1971 Female: 2230
Age	30–106	$\mu = 73.07$ $\sigma = 7.205$	30–101	$\mu = 72.392$ $\sigma = 7.001$	42–106	$\mu = 74.651$ $\sigma = 7.425$
Education	0–25	$\mu = 6.21$ $\sigma = 4.8$	0–25	$\mu = 5.909$ $\sigma = 4.895$	0–22	$\mu = 6.913$ $\sigma = 4.492$
MMSE-KC	0–30	$\mu = 23.348$ $\sigma = 4.8$	14–30	$\mu = 25.214$ $\sigma = 3.098$	0–25	$\mu = 18.995$ $\sigma = 5.226$

Table 4. Statistical Information for the Collected CERAD-K (Phase2). MCI: mild cognitive impairment.

Variables	All Patients		Patients (MCI)		Patients (Dementia)	
	Range	Descriptive statistics	Range	Descriptive statistics	Range	Descriptive statistics
Gender	Male/Female	Male: 501 Female: 735	Male/Female	Male: 287 Female: 376	Male/Female	Male: 214 Female: 359
Age	45–99	$\mu = 76.381$ $\sigma = 7.624$	54–99	$\mu = 74.118$ $\sigma = 6.894$	45–99	$\mu = 79$ $\sigma = 7.596$
Education	0–18	$\mu = 5.979$ $\sigma = 4.71$	0–18	$\mu = 6.596$ $\sigma = 4.592$	0–18	$\mu = 5.265$ $\sigma = 4.747$
MMSE-KC	0–30	$\mu = 17.715$ $\sigma = 6.026$	0–30	$\mu = 20.021$ $\sigma = 4.978$	0–28	$\mu = 15.047$ $\sigma = 6.035$

3.3. Preprocessing

The collected data includes data from patients who have problems with hearing and vision or who are unable to be examined due to anxiety, and data that is lost due to errors or omissions in the data collection process. Because only a few machine learning algorithms ignore missing value (e.g., Bayes, Neural Network) during data training and most algorithms can be affected by such gaps, we deleted missing values and errors using data preprocessing. Data preprocessing is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the data. Data preprocessing involves finding erroneous, incomplete, irrelevant or corrupt data in a record set, table, or database, and then correcting or deleting the data. The missing

values in our data may come from patients who did not properly understand the test. The results for this kind of patient considered cognitive decline based on family interviews and other parts of test results. Thus, if data type is categorical, we changed missing values to abnormal and, if numeric, to 0.

Data normalization changes values in different ranges to values in the same range, preventing an attribute with a larger range of values from having a larger weight than a smaller range of attributes.

For numeric data, there are four approaches to normalizing data: firstly, to convert to a range between 0 and 1; secondly to use a value between -1 and 1 ; thirdly to find the average and standard deviation of the attribute; and, fourthly, to normalize the data using the log value.

In the case of categorical data, algorithms using neural networks and statistical methods among machine learning algorithms cannot process categorical data, so they are converted to binary data or one-hot encoding [17]. For example, in regard to one-hot encoding, if there are three categorical data such as red, blue, and green in the color attribute, they are converted to a format such as (100), (010), (001). In this study, we use the maximum-minimum normalization method for numerical data and one-hot encoding for categorical data among various normalization methods. Equation (1) shows the maximum-minimum normalization formula.

$$\text{Normalization Value } (d) = \frac{\text{original Value} - \text{oldMin}}{\text{oldMax} - \text{oldMin}} \quad (1)$$

3.4. Feature Selection

Feature selection is a method of extracting the most relevant features from data with a certain pattern. It can be used to remove irrelevant data, eliminate redundancy, and identify what features contribute to a model with high accuracy. When creating a model, it is important to write as few features as possible, so that one can reduce the number of features through feature selection. In this paper, feature selection is performed using chi square and information gain.

The chi-square test is used to analyze categories, variables, or relationships, and is useful for studying categorical variables such as regional and political preferences, and the relationship between food and obesity. The chi-square test can be used in two broad contexts: as a fitness test to test whether the observed data follows a predicted distribution, and as an independence test to test whether two random variables are independent of each other. Independence means that there is no cause or effect relationship between them. The following Equation (2) indicates the chi-square feature selection.

$$\text{chi square} = \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \quad (2)$$

Information gain means that when an attribute is selected, the data is well distinguished because of its attribute. Information gain is the value obtained by subtracting the entropy value of the lower node from the entropy of the upper node. Equation (3) is used for calculating the information gain amount when the attribute A is selected, and the entropy of the original node is obtained. This is the result of subtracting the value divided by the smallest m nodes. The larger the value of Gain (A), the greater the information gain and the better the discriminative power; see the study conducted by Garrard et al. [18].

$$\text{Gain } (A) = I(s_1, s_2, \dots, s_m) - E(\text{feature } A) \quad (3)$$

In this study, feature selection was performed in two distinct phases, in which Phase 1 dealt with MMSE-KC data, while Phase 2 selected features among data from both MMSE-KC and CERAD-K.

3.4.1. Phase 1

As a result of feature selection of thirteen attributes in Phase 1, both algorithms (chi-squared, information gain) showed almost identical results. Table 5 shows the results in the order of location, timing, order execution, and memory recall.

Table 5. Feature Selection Using Chi-squared and Information Gain (Phase 1).

Rank	Chi-Squared	Information Gain
1	Orientation to place	Orientation to place
2	Orientation to time	Orientation to time
3	Three-stage commands	Three-stage commands
4	Recall	Recall
5	Attention	Attention
6	Repetition	Repetition
7	Registration	Registration
8	Complex commands	Complex commands
9	Language	Education
10	Education	Language
11	Understanding and decision	Understanding and decision
12	Gender	Gender
13	Age	Age

3.4.2. Phase 2

As a result of feature selection of Phase 2 data, both algorithms showed the same results. The most influential feature appeared to be temporal order, followed by memory function (Trial 1), place order, and a language fluency test. Memory, Word Fluency, Boston Naming, Visuospatial. Regarding MMSE-KC data in Phase 2, only the time and location were ranked and the rest were all at the bottom. Details are given in Table 6. The darker color is the CERAD-K data added in Phase 2.

Table 6. Feature Selection Using Chi-squared and Information Gain (Phase 2).

Rank	Chi-Squared/Information Gain
1	Orientation to time
2	Memory function (Trial 1)
3	Orientation to place
4	Word Fluency
5	Visuospatial function (Trial 2)
6	Boston Naming
7	Memory function (Trial 2)
8	Age
9	Memory function (Trial 3)
10	Visuospatial function (Trial 1)
11	Mental flexibility (number)
12	Mental flexibility (number, letter)
13	Repetition
14	Registration
15	Attention
16	Education
17	Recall
18	Three-stage commands
19	Language
20	Gender

3.5. Classifiers

Of the various uses to which machine learning is put, data mining is the most important. People are liable to err in analyzing data or seeking to discern relationships between various features, and these mistakes interfere with the problem-solving process. Frequently these problems are conducive to the application of machine learning, which can thereby optimize systemic efficiency and design. With machine learning algorithms, each instance within a dataset is represented with a consistent set of features—continuous, categorical, or binary. Supervised learning describes cases wherein instances

are provided with known labels of the corresponding outputs, while unsupervised learning involves no labeling of instances. A great many applications of machine learning necessitate supervised tasks, so we focus here on the requisite techniques for accomplishing this labeling [19].

3.5.1. Support Vector Machine

The machine learning method of Support Vector Machine (SVM) involves a mapping model for analyzing data and recognizing patterns. Classification and regression analysis are its primary uses. When provided with a set of data falling into one of two categories, an SVM algorithm constructs a non-probabilistic binary linear classification model by which it can ascertain from the given data the correct category in which to place new data. The classification model it produces is conveyed as a boundary within the space of the mapped data. An SVM algorithm establishes the widest boundary. It is useful for both linear and nonlinear classification. For non-linear classification, it is necessary to map given data onto a high-dimensional feature space. In order to do this efficiently, a kernel trick may be used [20,21].

3.5.2. Naive Bayes

Naive Bayes is a kind of probability classifier applying the Bayesian theorem, and it is one of the most used classification methods, as in text classification and document classification [22]. Naive Bayes learns using algorithms based on general principles rather than undergoing training through a single algorithm. Naive Bayes is trained very efficiently in a supervised learning environment and estimates parameters using Maximum Likelihood Estimation. The Naive Bayes classification is a combination of the probability model and the decision rule described above, and finds the class with the maximum probability.

3.5.3. Random Forest

The Random forest is a kind of ensemble method that randomly learns decision trees. It consists of a learning step that constructs a large number of decision trees and a test step that classifies and predicts when input vectors come in [23]. Random forests are used in various applications such as detection, classification, and regression. The most important feature of the random forest is that it consists of trees with slightly different characteristics due to randomness, and improves generalization performance by de-correlating the prediction of each tree. In addition, the randomization characteristics can be improved through the ensemble learning method, i.e., the invitation method and the arbitrary node optimization method.

3.5.4. Logistic Regression

Logistic regression analysis is a stochastic model that is used when a dependent variable refers to a binomial problem; it is a statistical technique used to predict the likelihood of an event using a linear combination of independent variables [24]. Therefore, the relationship between the dependent variable and the independent variable is expressed as a concrete function and used in future prediction models. In addition, unlike linear regression analysis, logistic regression analysis is often used as a classification and prediction model in which the results of marine data are divided into specific categories when the input data is given to the categorical data.

3.5.5. Bagging

Bagging is an ensemble learning method designed to improve the safety and accuracy of machine learning algorithms used in statistical classification and regression analysis [25]. Bagging also reduces variance and avoids overfitting, and is applied not only to decision tree learning methods and random forests, but also to other methods.

3.5.6. Bayesian Network

A Bayesian network graphically models probabilistic relationships between pertinent variables. In data analysis, a Bayesian model offers a number of benefits when paired with statistical techniques. First, since the model charts dependencies between all variables, it can easily address an instance in which there are gaps in data entries. Second, one can employ a Bayesian network to discern causal relationships, and thus to more fully grasp a problem domain and anticipate the effects of intervention. Third, a Bayesian model contains both probabilistic and causal semantics, making it particularly well-suited for connecting data with prior knowledge (since the latter frequently takes shape as causal). Lastly, Bayesian networks combined with statistical methods provide a clearly delineated and effective way to prevent data overfitting [26].

3.5.7. Multilayer Perceptron

A Multilayer Perceptron (MLP) is a feedforward neural network, instructed by way of a backpropagation algorithm. Because it is a supervised network, it must have a sought-after response for its training: what MLP learn is how to translate given data into that response. As such, they are frequently employed in pattern classification. They are able, with a hidden layer or two, to match almost any input-output map. In challenging problems, they have proven to be the equal of optimal statistical classifiers. For these reasons, they are at present arguably the most widely used network architecture: nearly all neural network applications make use of MLP. MLP is arranged so that neurons are divided into delineated layers, and each layer's output is conjoined with the nodular input of the subsequent layer. Hence the first (or input) layer represents the inputs to the network, and the last layer's outputs represent those of the network [27].

4. Results and Discussion

In this section, we study the early diagnosis of dementia, according to results of data mining techniques such as multilayer perceptron, random forest, bagging, SVM, logistic regression, Bayesian network, and Naive Bayes that were explained above. We then compare them to discern which is more accurate in the diagnosis of dementia.

As stated earlier, in Section 3.2, we used data from the Gangbuk-Gu center for dementia. Since classification of data includes two classes—namely Phase 1, consisting of 14,000 data including the normal class (9799 data) and cognitive decline class (4201 data), and Phase 2, consisting of 1236 data including the MCI class (663 data) and dementia class (573 data)—we used 10-fold cross-validation.

In cross-validation, data is divided into two segments in order to statistically compare and assess learning algorithms. One data segment (the training set) trains a model, while the other (the validation set) validates it. Usually, these sets are required to be staggered over successive rounds, so that each data point can be validated against the next. K-fold cross-validation is the standard cross-validation form, providing the basis modified in special cases or repeated rounds of cross-validation [28]. For this reason, we show the accuracy of these criteria: precision, recall and F-measure to diagnose dementia. Each of these criteria has been obtained from Equations (4)–(6).

$$precision = \frac{TP}{TP + FP} \quad (4)$$

$$recall = \frac{TP}{TP + FN} \quad (5)$$

$$F - measure = 2 \times \frac{precision \times recall}{precision + recall} \quad (6)$$

According to Equations (4)–(6), *TP* (True Positives) is equivalent to the number of samples that correctly have been identified as positive. Likewise, *FP* (False Positives) is equivalent to the number of

samples that have been wrongly identified as positive, and *FN* (False Negative) is equivalent to the number of samples that have been wrongly identified as negative [29].

4.1. Phase 1

As mentioned in Section 3, Phase 1 classifies subjects into categories of normal and cognitive decline. The results of any use of data mining techniques are based on Table 7. This table shows the achieved accuracy based on normal and cognitive decline class of features in Table 5.

Table 7. The Results of Classification Based on Phase 1 (normal, cognitive decline). TP (True Positives); FP (False Positives); FN (False Negative); Support Vector Machine (SVM).

Algorithms	TP Rate	FP Rate	Precision	Recall	F-Measure
MLP	0.97	0.04	0.97	0.97	0.97
Random Forest	0.95	0.05	0.96	0.95	0.96
Bagging	0.93	0.07	0.94	0.93	0.93
SVM	0.90	0.10	0.91	0.90	0.90
Bayes Network	0.90	0.10	0.91	0.90	0.90
Logistic Regression	0.78	0.22	0.81	0.78	0.79
Naive Bayes	0.74	0.26	0.80	0.74	0.75

Given the results of Table 7, Figure 2 shows the comparison of the accuracy of these criteria: precision, recall and F-measure. According to Figure 2, the highest precision recall and F-measure accuracy in Phase 1 belongs to MLP with 0.97, 0.97, and 0.97, respectively, followed by random forest and bagging.

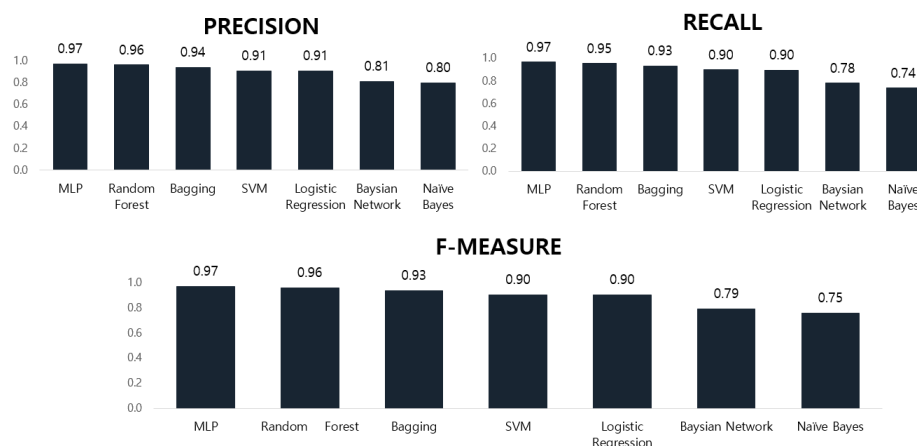


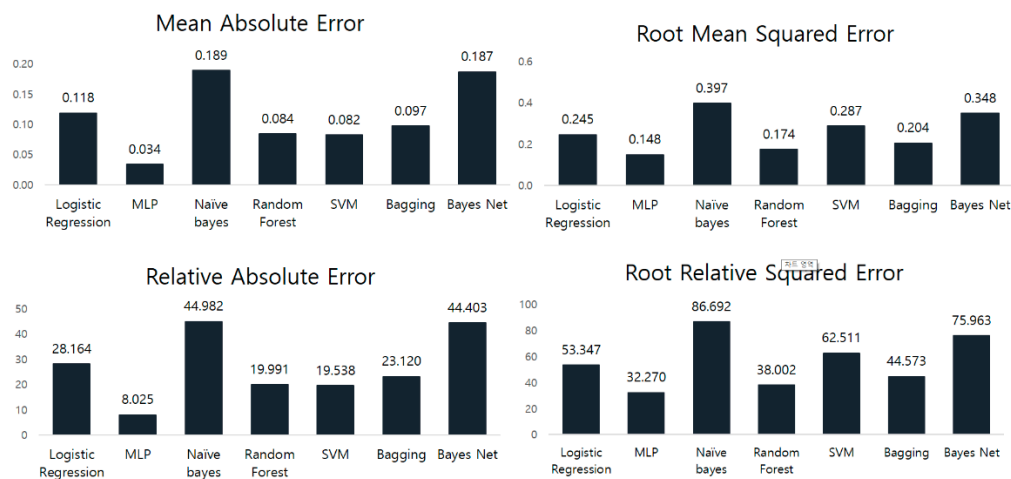
Figure 2. Comparison of Classification Based on the Precision, Recall and F-measure Criteria for Diagnosis of Dementia (Phase 1).

Each algorithm has a level of error in the diagnosis of dementia; by using four criteria—Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE) and Root Relative Squared Error (RRSE) in the Table 8—any technical errors in the diagnosis of dementia is shown.

Table 8. The Results of Errors Obtained from Classification to Diagnosis of Dementia (Phase 1). Multilayer Perceptron (MLP).

Algorithms	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Root Relative Squared Error
Logistic Regression	0.118	0.245	28.164	53.347
MLP	0.034	0.148	8.025	32.270
Naive Bayes	0.189	0.397	44.982	86.692
Random Forest	0.084	0.174	19.991	38.002
SVM	0.082	0.287	19.538	62.511
Bagging	0.097	0.204	23.120	44.573
Bayes Network	0.187	0.348	44.403	75.963

According to the results of Table 8, Figure 3 shows the comparison of the level of error for the diagnosis of dementia by these criteria MAE, RMSE, RAE and RRSE, to decide which algorithms of these criteria have the lowest error. According to Figure 3, concerning MAE, MLP has the lowest error in MAE criterion compared to other algorithms. Also, the level of error with SVM is close to that of MLP and, after MLP, it has less error than other algorithms in diagnosis of dementia. Based on Figure 3, MLP also has the lowest error in RMSE criterion compared to other algorithms; the level of error in random forest, bagging, and logistic regression is almost identical. In regard to RAE, Figure 3 shows that MLP has the lowest error in RAE criterion compared to other algorithms. Furthermore, SVM follows MLP in having less error than other algorithms in diagnosis of dementia. Finally, in regard to RRSE, as Figure 3 shows, MLP has the lowest error in RRSE criterion compared to other algorithms. And after MLP, random forest and bagging are closer together in level of error, with the lowest error in the diagnosis of dementia.

**Figure 3.** Comparison of the Classification Algorithms Based on Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE) and Root Relative Squared Error (RRSE) Criteria for the Diagnosis of Dementia (Phase 1).

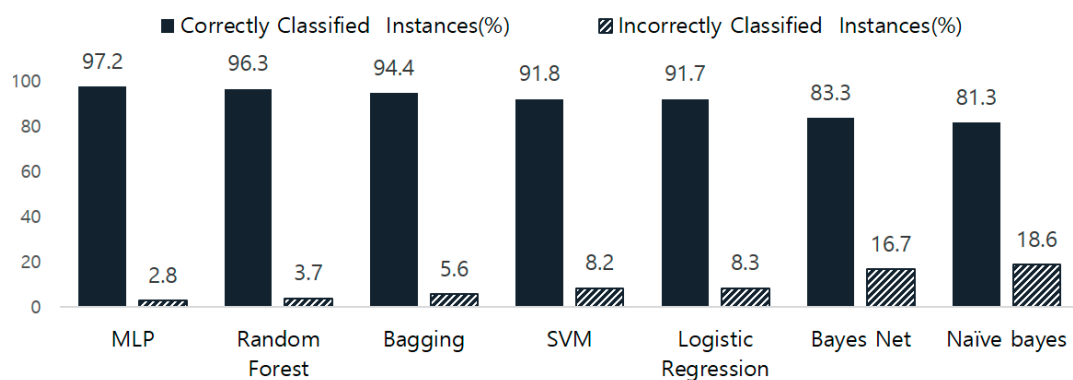
As noted, the Phase 1 data are tested and the data classification accuracy rate for diagnosis of cognitive decline is performed by testing data according to Equation (7).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

According to Table 9, the comparison of data classification accuracy for diagnosis of dementia is shown in Figure 4.

Table 9. Data Classification Accuracy of Dementia by Evaluating Experimental Data.

Algorithms	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)
MLP	97.2	2.8
Random Forest	96.3	3.7
Bagging	94.4	5.6
SVM	91.8	8.2
Logistic Regression	91.7	8.3
Bayes Net	83.3	16.7
Naive Bayes	81.3	18.6

**Figure 4.** The Comparison of Data Classification Accuracy in Phase 1.

As the results in Table 9 and Figure 4 are clear, MLP has the highest accuracy for the diagnosis of dementia (Phase 1) and the value of this algorithm is equal to 97.2%. Additionally, the classification accuracy of random forest and bagging was 96.3% and 94.4%, respectively. According to Figure 4, Bayes Network and Naive Bayes have the lowest classification accuracy in the diagnosis of dementia. Moreover, the accuracy of SVM and logistic regression was almost equal, meaning that both algorithms possessed the classification accuracy of 91.7%.

4.2. Phase 2

As mentioned in Section 3, Phase 2 classifies MCI and dementia. The results of any use of data mining techniques are based on Table 10. This table shows the achieved accuracy based on the MCI and dementia classes of feature in Table 6.

Table 10. The Results of Classification Based on Phase 2 (MCI, dementia).

Algorithms	TP Rate	FP Rate	Precision	Recall	F-Measure
SVM	0.738	0.262	0.739	0.738	0.739
Logistic Regression	0.736	0.265	0.736	0.736	0.736
Random Forest	0.729	0.272	0.729	0.729	0.729
Bagging	0.724	0.276	0.724	0.724	0.724
Naive Bayes	0.712	0.289	0.713	0.712	0.712
Bayes Network	0.709	0.292	0.708	0.709	0.708
MLP	0.679	0.321	0.679	0.679	0.680

Taking the results of Table 10, Figure 5 shows the comparison of the accuracy of these criteria: precision, recall and F-measure. According to Figure 5, the highest precision recall and F-measure accuracy in Phase 1 belongs to SVM with 0.739, 0.738, and 0.739, respectively, followed by logistic regression and random forest.

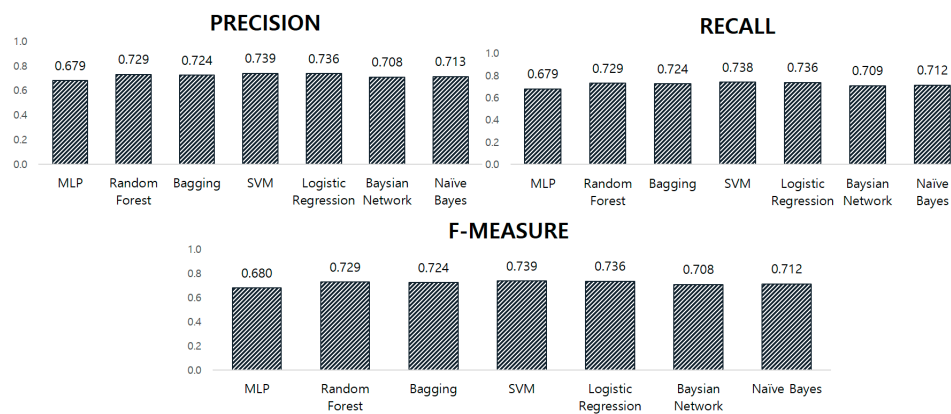


Figure 5. Comparison of the Classification Based on the Precision, Recall, and F-measure criteria for the Diagnosis of Dementia (Phase 2).

Each algorithm has a level of error in the diagnosis of dementia. By using four criteria (MAE, RMSE, RAE and RRSE, in the Table 11) any technical errors in the diagnosis of dementia are shown.

Table 11. The Results of Errors Obtained from Classification to Diagnosis of Dementia (Phase 2).

Algorithms	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Root Relative Squared Error
Logistic Regression	0.35	0.43	70.20	85.27
MLP	0.33	0.54	65.53	107.86
Naive Bayes	0.29	0.51	57.43	101.64
Random Forest	0.37	0.43	73.54	85.29
SVM	0.26	0.51	52.87	102.83
Bagging	0.36	0.43	72.20	86.10
Bayes Net	0.29	0.50	57.83	100.72

Given the results of Table 11, Figure 6 compares the errors in the diagnosis of dementia by these criteria MAE, RMSE, RAE and RRSE, to discern which algorithms of these criteria have the lowest error. According to Figure 6, SVM has the lowest error in the MAE criterion compared to other algorithms. Also, the level of error for Naive Bayes is close to the SVM and after SVM has less error than the other algorithms in the diagnosis of dementia. Regarding RMSE, based on Figure 6, random forest and bagging had the lowest error in RMSE criterion compared to other algorithms; the level of error in bagging, random forest and logistic regression is very close together. Figure 6 also shows that SVM has the lowest error in RRSE criterion compared to other algorithms. After SVM, Naive Bayes has a level of error close to it, and the lowest error to diagnosis of heart disease. Finally, in regard to RRSE, as Figure 6 shows, logistic regression has the lowest error by the RRSE criterion when compared to other algorithms. After logistic regression, Bayes Network had the lowest error for the diagnosis of dementia.

As noted, Phase 2 data have been tested and their data classification accuracy rate for diagnosis of cognitive decline assessed according to Equation (7).

According to Table 12, the comparison of data classification accuracy for diagnosis of dementia is shown in Figure 7.

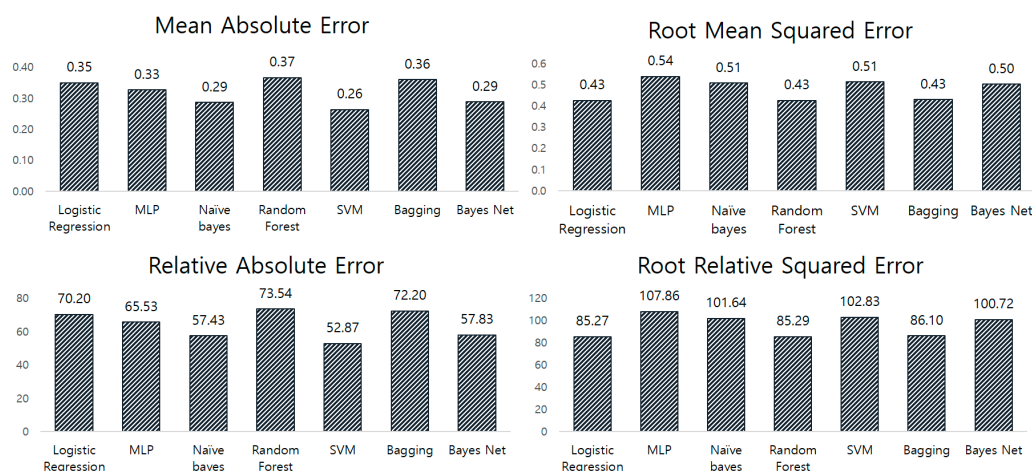


Figure 6. Comparison of the Classification Algorithms Based on MAE, RMSE, RAE and RRSE Criteria for Diagnosis of Dementia (Phase 2).

Table 12. Data Classification Accuracy for Dementia (Phase 2).

Algorithms	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)
SVM	74.03	25.97
Logistic Regression	73.71	26.29
Random Forest	72.98	27.02
Bagging	72.49	27.51
Naive Bayes	71.44	28.56
Bayes Net	70.95	29.05
MLP	68.12	31.88

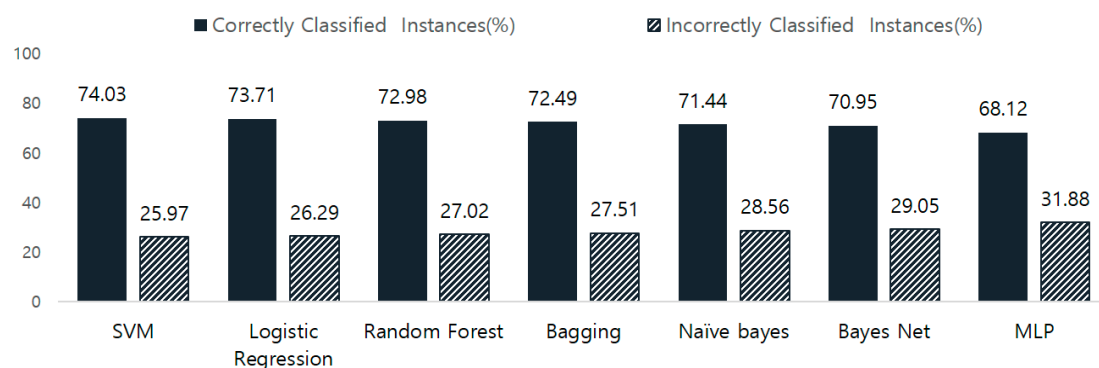


Figure 7. The Comparison of Classification Accuracy by Phase 2.

As the results in Table 12 and Figure 7 clearly demonstrate, SVM has the most accuracy for the diagnosis of dementia (Phase 2) and the value of this algorithm is equal to 74.03%. Moreover, the classification accuracy of logistic regression and random forest were equal to 73.71% and 72.98%, respectively. According to Figure 7, MLP had the lowest correct classification accuracy in the diagnosis of dementia (Phase 2). The accuracy of bagging and Naive Bayes were 72.49% and 71.44%, respectively. The classification accuracy of Bayes Network was 70.95%.

To sum up, in Phase 1, MLP showed the highest accuracy with 97.2%, followed by random forest and bagging. The lowest accuracy was Naive Bayes at 81.3%. In Phase 2, SVM was tops among other classifiers for MCI and dementia cases with 74.03%, followed by logistic regression and random forest. Whereas MLP was the best in Phase 1 for predicting normal, SVM was best in Phase 2 for predicting dementia. The results of this study are consistent with findings from several researchers (e.g., [4,18])

showing that the machine learning approaches can be used to diagnose dementia. Our efforts in the diagnosis of dementia may be similar to those mentioned with respect to the employed machine learning approaches. However, inspired by the method used in dementia support centers for early diagnosis, not only can our proposed model diagnose dementia with a data from simple tests from patients, but also we can achieve higher accuracy in early diagnosis of dementia.

5. Conclusions

As the senior population increases due to social aging, the prevalence of dementia increases, and the number of young dementia patients also increases. In this study, we proposed a two-layer model inspired by the methods used in dementia support centers for the early diagnosis of dementia and using machine learning techniques. MMSE-KC and CERAD-K data have been used in screening and precise screening to reduce time and the economic burden on patients, and increase the accuracy of screening with the employed machine learning algorithms. In the first stage, the patients who need precise screening are classified by MMSE-KC data. In the second stage, MCI and dementia are classified by adding CERAD-K data. In conclusion, we compared various classification models using dementia diagnosis data. In Phase 1, the highest F-measure value belongs to MLP, while in Phase 2 the highest F-measure value belongs to SVM. Our proposed model simplifies the task of interpreting test results by constructing a set of criteria to classify the patient and therefore diagnose dementia at early stages in a fast, inexpensive, and reliable way, which improves the current clinical practice.

In future research, we will study a model that can predict dementia more precisely by using lifestyle or disease information of the patient and plan to improve the accuracy.

Acknowledgments: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (grant number R1610941).

Author Contributions: Kun Woo Park and Heui Seok Lim conceived and designed the experiments; Kun Woo Park and Aram So performed the experiments; Aram So analyzed the data; Heui Seok Lim and Danial Hooshyar contributed reagents/materials/analysis tools; Aram So and Danial Hooshyar wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization. Meeting report. In Proceedings of the First WHO Ministerial Conference on Global Action against Dementia, Geneva, Switzerland, 16–17 March 2015.
2. Prince, M.J. *World Alzheimer Report 2015: The Global Impact of Dementia: An Analysis of Prevalence, Incidence, Cost and Trends*; Alzheimer's Disease International (ADI): London, UK, 2015.
3. Trambaiolli, L.R.; Lorena, A.C.; Fraga, F.J.; Kanda, P.A.; Anghinah, R.; Nitrini, R. Improving Alzheimer's disease diagnosis with machine learning techniques. *Clin. EEG Neurosci.* **2011**, *42*, 160–165. [[CrossRef](#)] [[PubMed](#)]
4. Joshi, S.; Shenoy, P.D.; Venugopal, K.R.; Patnaik, L.M. Evaluation of different stages of dementia employing neuropsychological and machine learning techniques. In Proceedings of the First International Conference on Advanced Computing, Chennai, India, 13–15 December 2009.
5. Lezak, M.D. *Neuropsychological Assessment*; Oxford University Press: Oxford, MS, USA, 2004.
6. Datta, P.; Shankle, W.; Pazzani, M. Applying machine learning to an Alzheimer's database. In Proceedings of the AAAI-96 Spring Symposium AI in Medicine: Applications of Current Technologies, Stanford, CA, USA, 25–27 March 1996.
7. Brodaty, H.; Pond, D.; Kemp, N.M.; Luscombe, G.; Harding, L.; Berman, K.; Huppert, F.A. The GPCOG: A new screening test for dementia designed for general practice. *J. Am. Geriatr. Soc.* **2002**, *50*, 530–534. [[CrossRef](#)] [[PubMed](#)]
8. Nasreddine, Z.S.; Phillips, N.A.; Bédirian, V.; Charbonneau, S.; Whitehead, V.; Collin, I.; Cummings, J.L.; Chertkow, H. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **2005**, *53*, 695–699. [[CrossRef](#)] [[PubMed](#)]

9. Kim, K.W.; Park, J.H.; Kim, M.H.; Kim, M.D.; Kim, B.J.; Kim, S.K.; Kim, J.L.; Moon, S.W.; Bae, J.N.; Woo, J.I.; et al. A nationwide survey on the prevalence of dementia and mild cognitive impairment in South Korea. *J. Alzheimers Dis.* **2011**, *23*, 281–291. [[PubMed](#)]
10. Aarsland, D.; Cummings, J.L.; Larsen, J.P. Neuropsychiatric differences between Parkinson's disease with dementia and Alzheimer's disease. *Int. J. Geriatr. Psychiatr.* **2001**, *16*, 184–191. [[CrossRef](#)]
11. Lee, J.H.; Lee, K.U.; Lee, D.Y.; Kim, K.W.; Jhoo, J.H.; Kim, J.H.; Lee, K.H.; Kim, S.Y.; Han, S.H.; Woo, J.I. Development of the Korean Version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (CERAD-K) clinical and neuropsychological assessment batteries. *J. Gerontol. B Psychol. Sci. Soc. Sci.* **2002**, *57*, P47–P53. [[CrossRef](#)] [[PubMed](#)]
12. Morris, J.C.; Mohs, R.C.; Rogers, H.; Fillenbaum, G.; Heyman, A. Consortium to establish a registry for Alzheimer's disease (CERAD) clinical and neuropsychological assessment of Alzheimer's disease. *Psychopharmacol. Bull.* **1987**, *24*, 641–652.
13. Chen, R.; Herskovits, E.H. Machine-learning techniques for building a diagnostic model for very mild dementia. *Neuroimage* **2010**, *52*, 234–244. [[CrossRef](#)] [[PubMed](#)]
14. Williams, J.A.; Weakley, A.; Cook, D.J.; Schmitter-Edgecombe, M. Machine learning techniques for diagnostic differentiation of mild cognitive impairment and dementia. In Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI-13), Bellevue, WA, USA, 14–18 July 2013.
15. Cho, P.-C.; Chen, W.-H. A double layer dementia diagnosis system using machine learning techniques. In Proceedings of the 13th International Conference, (EANN 2012), London, UK, 20–23 September 2012.
16. Shanklea, W.R.; Mania, S.; Dick, M.B.; Pazzani, M.J. Simple models for estimating dementia severity using machine learning. *Stud. Health Technol. Inform.* **1998**, *52 Pt 1*, 472–476.
17. Harris, D.; Harris, S. *Digital Design and Computer Architecture*; Elsevier: Amsterdam, The Netherlands, 2012.
18. Garrard, P.; Rentoumi, V.; Gesierich, B.; Miller, B.; Gorno-Tempini, M.L. Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse. *Cortex* **2014**, *55*, 122–129. [[CrossRef](#)] [[PubMed](#)]
19. Alberdi, A.; Aztiria, A.; Basarab, A. On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. *Artif. Intell. Med.* **2016**, *71*, 1–29. [[CrossRef](#)] [[PubMed](#)]
20. Buscema, M.; Vernieri, F.; Massini, G.; Scarscia, F.; Breda, M.; Rossini, P.M.; Grossi, E. An improved I-FAST system for the diagnosis of Alzheimer's disease from unprocessed electroencephalograms by using robust invariant features. *Artif. Intell. Med.* **2015**, *64*, 59–74. [[CrossRef](#)] [[PubMed](#)]
21. Kotsiantis, S.B.; Zaharakis, I.; Pintelas, P. Supervised machine learning: A review of classification techniques. *Informatica* **2007**, *31*, 249–268.
22. Rish, I. An empirical study of the naive Bayes classifier. In *Proceedings of IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*; IBM: New York, NY, USA, 2001.
23. Liaw, A.; Wiener, M. Classification and regression by random forest. *R News* **2002**, *2*, 18–22.
24. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013; Volume 398.
25. Quinlan, J.R. Bagging, boosting, and C4. 5. *AAAI/IAAI* **1996**, *1*, 725–730.
26. Friedman, N.; Geiger, D.; Goldszmidt, M. Bayesian network classifiers. *Mach. Learn.* **1997**, *29*, 131–163. [[CrossRef](#)]
27. Gardner, M.W.; Dorling, S. Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmos. Environ.* **1998**, *32*, 2627–2636. [[CrossRef](#)]
28. Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection. *Stat. Surv.* **2010**, *4*, 40–79. [[CrossRef](#)]
29. Glas, A.S.; Lijmer, J.G.; Prins, M.H.; Bonsel, G.J.; Bossuyt, P.M. The diagnostic odds ratio: A single indicator of test performance. *J. Clin. Epidemiol.* **2003**, *56*, 1129–1135. [[CrossRef](#)]

