의학석사 학위논문

# Prediction of Functional Recovery in Patients with Supratentorial Ischemic Stroke by Various Methodology of Machine Learning

기계학습의 다양한 방법론을 통한
천막상 허혈성 뇌졸중 환자의
기능 회복 예측

2018년 8월

서울대학교 대학원
의과대학
협동과정 의료정보학 전공

이 현 행

# Abstract

**Objectives:** The purpose of the present study was to predict the functional recovery of supratentorial ischemic stroke after post-stroke 3 months with the clinical data of patients obtained within 2 weeks from onset of stroke by using various methodology of machine learning (ML) including artificial neural network (ANN).

**Methods:** We extracted a list of patients who had been discharged from the Department of Rehabilitation Medicine, a university hospital from Jan. 2000 to Dec. 2017. Afterward, we collected the clinical data of patients meeting inclusion/exclusion criteria. We selected features for the construction of a prediction model among clinical features which has been known to affect post-stroke recovery and expected to affect it. The selected clinical features are age, sex, initial National Institutes of Health Stroke Scale, presence of internal capsule posterior limb involved, strength of shoulder abduction, wrist extension, hip extension, and knee extension, Mini-Mental State Examination, presence of hemorrhagic transformation, aphasia, visuospatial neglect, and depression. We dichotomized post-stroke 3 months functional status assessed with modified Barthel Index, which was used as an outcome label for the prediction model. We optimized the hyperparameters of ANN model and the other method of machine learning by using the grid search with 2-fold cross validation. We repeated the training and validation session 10 times with the different configuration of training and test dataset generated by randomized sampling. The average of performance of 10 individual models was assigned to represent overall performance of the respective method of machine learning.

**Results**: We screened 5210 patients and eventually enrolled 101 patients with supratentorial ischemic stroke, whose functional recovery was assessed with modified Barthel Index after 3 months post-stroke. The mean age of the enrolled subjects was 62.40 ± 12.67 [19 - 79] years. The patients in group with better functional status after post-stroke 3 months tend to have younger age (59.77 ± 14.18 versus 67.34 ± 7.02) and lower initial NIHSS (7.98 ± 5.26 versus 14.42 ± 5.48), and less likely to have stroke lesion in posterior limb of internal capsule (15.15% versus 57.14%), and have better cognitive function (total MMSE; 24.54 ± 5.97 versus 18.86 ± 7.02). The architectures of ANN with optimized hyperparameters was turned out to have 4 hidden layers with from 64 to 4 nodes. The proposed ANN model used rectified linear unit as activation function, Glorot uniform initializer as the way to set the initial weight, 0.3 dropout rate, Adagrad as optimizer, 0.02 learning rate, 5 batch size, 600 epochs and binary crossentropy as loss function. Sigmoid function as the classifier was placed at the last layer for prediction. The accuracy of model constructed by the method of ANN turned out to be 85.38 ± 6.15 (%), which was superior to those by the other method of machine learning.

**Conclusion**: In the present study, we demonstrated that the prediction of function recovery after supratentorial ischemic stroke can be performed with a high degree of accuracy by the various methodology of machine learning, with the highest in ANN.

**Keywords**: machine learning, artificial neural network, prediction, functional status, stroke

**Student Number**: 2015-22073

# Table of Contents

# List of Tables and Figures

# Introduction

## Study Background

Accurate prediction for post-stroke recovery is important for goal setting and treatment planning of rehabilitation programs. In addition, prediction of recovery after stroke can help patients and caregivers plan their upcoming lives by anticipating how long medical treatment will last and how much financial resources are needed for the disease treatment.

According to the statistics reported by Korean Health Insurance Review and Assessment Service, the number of stroke patients has risen by 11.6% in 2013 compared to 2007 and has not increased since 2011. However, medical costs have increased year by year and have increased by 34.5% in 2013 compared to 2007. In recent years, acute stroke management has been successful with the establishment of aggressive policy on stroke management and improvement of medical delivery system, so that the mortality rate is gradually decreasing. The mortality rate of cerebrovascular disease decreased to 50.3 / 100,000 according to the report of the Korean National Statistical Office in 2013, but the number of patients with impairment after stroke was relatively increased. Taken together, it can be concluded that the increase in the cost of care for patients with chronic stroke is leading to an increase in the cost of medical care for stroke patients, which includes unnecessary hospitalization and treatment. It is expected that unnecessary medical expenditure for post-stroke patients can be drastically reduced through the medical treatment plan by precise prediction on post-stroke recovery. It has been reported that applying an algorithm predicting recovery of upper extremity function to the rehabilitation program reduced the length of hospital stay by 1 week [1].

It is known that post-stroke functional recovery proceeds with a high rate in the first 1-2 months after the stroke, gradually slows down after 3 months and plateau in about 6 months [2]. The pattern of functional recovery is similar in the various functional domains such as cognition, walking, and activities of daily life [3]. The systematic review of methodologically qualified 48 studies about prognostic factors determining the final outcome of ADL revealed that age, initial neurologic status and upper arm paresis were factors affecting outcome of ADL at post-stroke 3 months [4]. Various factors linked directly or indirectly to the patients' prognosis in the post-stroke phase make predicting individualized post-stroke functional recovery complicated. In recent studies, various biomarkers such as neurophysiological tests and neuroimaging techniques having been studied to more accurately predict post-stroke prognosis [5–7].

The method of supervised machine learning is the way of searching for algorithms that reason from input instances to derive general hypotheses, which make predictions for new instances [8]. Because of advances in processing power, memory, storage, and an unprecedented wealth of data, prediction of disease prognosis through machine learning has been made with high accuracy [9].

The artificial neural network (ANN) as one method of machine learning mimics the operation of the human brain using multiple layers of neural networks which can generate automated predictions from input datasets [10]. The method of ANN processes the inputs in a layer-wise nonlinear manner to the pre-trained nodes in subsequent hidden layers to learn 'structures' and representations that are generalizable. Recently, accuracy of the prediction model through ANN including deep neural network has been improved dramatically, which has been widely used for predicting the risk of diseases and diagnosing disease [11–13]. The recent development of high-level modules (e.g. Theano [14], Keras [15], and TensorFlow

[16]) to easily build neural networks allowed physicians to take advantage of methodology of ANN as state-of-the-art solutions for several tasks.

## Purpose of Research

The purpose of the present study was to predict the functional recovery status of supratentorial ischemic stroke after post-stroke 3 months with the clinical information of patients obtained within 2 weeks from onset of stroke, such as demographics, initial neurological status, and brain imaging by using the various method of ML. In addition, we also aim to show the clinical usefulness of ANN by comparing the accuracy of prediction model by ANN with that the other method of ML.

# Methods

## Subjects enrollment

Firstly, we extracted a list of patients who had been discharged from the Department of Rehabilitation Medicine, a university hospital from Jan. 2000 to Dec. 2017. Afterward, we collected the clinical data of patients meeting inclusion/exclusion criteria as follows.

Inclusion criteria were patients 1) who were ≥ 18 and ≤ 80 years old, 2) with ischemic stroke confirmed by brain magnetic resonance imaging within 2 days from onset of stroke, 3) who were transferred to the Department of Rehabilitation in 10 days post-stroke (window period; 3 days) after the intervention for acute stroke in the Department of Neurology, 4) with in-hospital days ≤ 60 days, and 5) with modified Barthel Index (mBI) scored in post-stroke 3 months.

Exclusion criteria were patients 1) with previous stroke, 2) with previous traumatic brain injury, 3) with infratentorial stroke lesion, 4) with cancer affecting general conditions of patients, 5) with other disorders which could lead to sensory or motor deficit, such as Parkinson's disease, multiple sclerosis, lumbosacral radiculopathy, peripheral neuropathy, musculoskeletal problems involving the lower extremity, and visual or vestibular dysfunction, 6) with post-stroke spasticity, 7) with Alzheimer's dementia, 8) with apraxia or visual field defect, 9) with history of post-stroke seizure, and 10) with neurologic event until post-stroke 3 months.

To find out the patients with the diagnosis related with supratentorial ischemic stroke, we included patients with primary diagnosis related with supratentorial ischemic stroke by using the keywords; 'hemorrhage', 'meningitis', 'encephalitis', 'multiple sclerosis', 'hypoxic', 'trauma', 'injury', 'fracture', 'parkinson', 'multiple system atrophy', 'aneurysm', 'cerebral palsy', 'spinal cord injury', 'myelopathy', 'paraplegia', 'tetraplegia', 'cauda equina syndrome', 'spinal stenosis', 'amyotrophic lateral sclerosis', 'guillain barre syndrome', 'myopathy', 'neuropathy', '*oma', 'tumor', 'neoplasm', 'cancer', or 'infection'.

## Feature selection

We selected features for the construction of prediction model among clinical information known until the time of transfer to ward for the department of rehabilitation medicine which has been known to affect post-stroke recovery [17–20] and expected to affect it (Fig. 1). The corresponding clinical variables are as follows: 1) age, 2) sex, 3) initial NIHSS (National Institutes of Health Stroke Scale), 4) presence of internal capsule posterior limb involved, 5) Strength of shoulder abduction, wrist extension, hip extension, and knee extension confirmed by manual muscle test (MMT) in 10 days post-stroke (window period; 3 days), 6) Mini-Mental State Examination (MMSE) in 10 days post-stroke (window period; 3 days), 7) presence of hemorrhagic transformation, 8) aphasia, 9) visuospatial neglect, and 10) depression confirmed by experienced psychiatrist.

The age at the onset of stroke and initial neurologic status including initial NIHSS and the strength of shoulder abduction, finger extension, hip extension, and knee extension are well-identified variables to affect post-stroke recovery [4, 21–24]. The effect of sex difference on post-stroke recovery is still controversial. Studies showed that female patients had achieved lower scores in activities of daily living after stroke [25]. But, some studies argued that the effect of sex difference on post-stroke recovery is inconclusive [26]. The corticospinal tract (CST) is the main pathway that mediates voluntary movements, which originates mainly from the cortex within the precentral gyrus and descends through the corona radiata, posterior limb of internal capsule, and crus of midbrain and continues to the lower end of the brainstem before crossing to the opposite side of the spinal cord. Thus, the involvement of motor-related cortical regions, corona radiata, and internal capsule decrease the probability of upper limb functional recovery [27, 28]. Acute CST damage at the level of the posterior limb of internal capsule turned out to be a significant predictor of unfavorable motor outcome confirmed by diffusion tensor tractography [29]. One study which analyzed the recovery of non-treated depressed patients showed some insights about the effect of depression on post-stroke recovery [30]. However, the effect of depression on post-stroke recovery is still inconclusive. The level of cognition is postulated to affect post-stroke recovery, however, there have been no published studies. The presence of hemorrhagic transformation after ischemic stroke, aphasia, visuospatial neglect, and depression are also suspected to be prognostic factors for post-stroke recovery, which however, do not have reliable evidences.

## Data preprocessing

*Categorization of selected features*

1) We dichotomized the cognitive level of the enrolled subject with total score of MMSE with the criteria for post-stroke dementia [31]. If MMSE < 24, then categorized into post-stroke cognitive dysfunction group. If MMSE ≥ 24, then categorized into non-cognitive dysfunction group.

2) We categorized the level of neurologic deficit with initial NIHSS [32]: score 0, no stroke symptoms; score 1 – 4, minor stroke; score 5 – 15, moderate stroke; score 16 – 20, moderate to severe stroke; and score 21 – 42, severe stroke.

*Normalization of selected features*

We normalized the selected features to reduce the risk of overshoot and model overfitting. Standardization was done for continuous variable such as age at the onset of stroke. We standardized ordinal variables such as muscle strength (shoulder abduction, wrist extension, hip extension, and knee extension), MMSE (score in domain of attention and memory), and categorized initial NIHSS. Binarization was done for nominal scales such as sex, dichotomized total MMSE score, presence of lesion involving posterior limb of internal capsule, presence of hemorrhagic transformation, presence of internal capsule posterior limb involved by hemorrhagic transformation, aphasia, visuospatial neglect, and depression.

*One-hot encoding for dichotomized label*

We dichotomized the functional status after post-stroke 3 months according to Korea's criteria for grading the level of post-stroke disabilities with modified Barthel Index (mBI, 0 – 69; functional status to the extent that continuous assistance of others is partially or entirely necessary, 70 – 99; functional status to the extent that help of others is intermittently necessary or not required at all.). The dichotomized label was transformed to one binary attribute per category, so called one-hot encoding.

## Design of prediction model by ANN and the other method of machine learning

*Artificial neural network*

The framework of ANN can be formulated as follows; vector of input variables : $x=\{x_1, x_2, \cdots, x_{16}\}$, the first hidden layer : $h_{(1)}=f(W \cdot x+b)$ (f as activation function, W as matrix of weight, b as vector of bias), the second hidden layer : $h_{(2)}=f(V \cdot h_{(1)}+c)$ (V as matrix of weight, c as vector of bias), the third hidden layer : $h_{(3)}=f(U \cdot h_{(2)}+d)$ (U as matrix of weight, d as vector of bias), the fourth hidden layer : $h_{(4)}=f(T \cdot h_{(3)}+e)$ (T as matrix of weight, e as vector of bias), output layer : $y=g(S \cdot h_{(4)}+i)$ (g as classifier, S as matrix of weight, i as vector of bias), and output variables : $y=\{y_1, y_2\}$.

We optimized the hyperparameters of ANN model among the following options of variables by using the grid search with 2-fold cross validation; the number of neurons at the first hidden layer: {16, 32, 64, 128}, activation function at hidden layers: {rectified linear unit function, hyperbolic tangent function, sigmoid function}, initializer: {uniform distribution, normal distribution, Glorot normal initializer Glorot uniform initializer}, dropout rate at hidden layers: {0, 0.1, 0.3, 0.6}, optimizer: {stochastic gradient descent optimizer, RMSProp optimizer, Adagrad optimizer}, learning rate if needed for the chosen optimizer: {0.001, 0.01, 0.02, 0.1, 0.2, 0.3} / momentum if needed for the chosen optimizer: {0.0, 0.2, 0.4, 0.6, 0.8} / batch size: {5, 10, 20, 30} / the number of epochs: {10, 50, 100, 500, 600, 1000}.

We split the entire dataset into the training and test dataset at 1:1 ratio. With ANN architecture with optimized hyperparameters, we went through training session with the training dataset and validated the model with test dataset.

All analyses were conducted by using Keras API (version 2.0.8) [15] based on Tensorflow [16], one of the neural network frameworks. The training process of ANN was visualized by taking advantage of ‘ggplot2’ package of the R software (version 3.3.1; http://www.r-project.org).

*Other ML methods*

In order to compare the performance of ANN method with the other method of machine learning, we used logistic regression, k-nearest neighbors (kNN), Bayes with Bernoulli method, support vector machine, decision tree, and decision-tree based ensemble method; random forest and gradient boosting. All analyses were conducted by using Scikit-Learn packages, which contains the interface for machine learning method based on the Python programming language [33].

We optimized hyperparameters of the respective method of machine learning among the following options of variables by using the grid search with 2-fold cross validation; for logistic regression; penalty to specify the norm used in the penalization: {‘l1’, ‘l2’}, C as inverse of regularization strength: {0.001, 0.01, 0.1, 1, 10, 100}; for K-nearest neighbors; the number of neighbors: {1, 3, 5, 7, 9, 10}, weight function used in prediction: {‘uniform’, ‘distance’}, algorithm used to compute the nearest neighbors: {‘auto’, ‘ball_tree’, ‘kd_tree’, ‘brute’}; for support vector machine; type of kernel to be used in the algorithm: {‘poly’, ‘rbf’, ‘sigmoid’}, C as penalty parameter of the error term: {0.01, 0.1, 1, 10, 100}, class weigh: {‘balanced’, none}; for decision tree; function to measure the quality of a split: {‘gini’, ‘entropy’}, strategy to choose the split at each node: {‘best’, ‘random’}, maximum depth of tree: {1, 2, 3, 4, 5}; for random forest; the number of trees in forest: {1,5, 10, 20, 30, 40, 50}, function to measure the quality of a split: {‘gini’, ‘entropy’}, the number of features to consider: {2, 4, 6, 8, 10, 12, 14, 16}, maximum depth of tree: {1, 2, 3, 4, 5}; for gradient boosting; loss function to be optimized: {‘deviance’, ‘exponential’}, learning rate: {0.001, 0.005, 0.01, 0.05, 0.1, 0.5}, the number of boosting stages to perform: {1, 5, 10, 15, 20, 30, 40, 50}, maximum depth of the individual regression estimators: {1, 2, 3, 4, 5}, function to measure the quality of a split: {‘friedman_mse’, ‘mse’}.

We split the entire dataset into the training and test dataset at 1:1 ratio. With each machine learning architecture with optimized hyperparameters, we went through training session with the training dataset and validated the model with test dataset.

## Comparison of model performance

We repeated the training and validation sessions 10 times with the different configuration of training and test datasets selected by randomized sampling. The average of performance of 10 individual model was assigned to represent overall performance of the respective method of machine learning. The performance of each model was analyzed in terms of accuracy, precision and recall. Accuracy is a ratio of correctly predicted observation to the total observations. Accuracy is an intuitive and general performance measure only when applied to the balanced dataset. Since the present study used the unbalanced data, we also performed model evaluation in terms of precision and recall. Precision is the fraction of relevant instances among the retrieved instances, while recall is the fraction of relevant instances that have been retrieved over the total amount of relevant instances. In the present study, we define case of good functional recovery as the relevant instance.

We also repeated the training and test sessions 100 times with the prediction model based on the respective configuration of training dataset by randomized sampling to compare each methodology of machine learning through the distribution and frequency of misclassification for each dataset.

# Results

We screened 5210 patients who were discharged from the department of rehabilitation medicine from Jan. 2000 to Dec. 2017. Among them, 3179 patients with diagnosis related with brain disease were selected. As flowchart shown in Fig. 2, we excluded the patients not meeting inclusion/exclusion criteria. We eventually enrolled 101 patients with supratentorial ischemic stroke, whose functional recovery was assessed with mBI after 3 months post-stroke.

The mean age of the enrolled subjects was 62.40 ± 12.67 [19 - 79] years. Fifty-three were men accounting for 52.5% of all the enrolled subjects (Table 1). The mean of initial NIHSS was 10.07 ± 6.10 [1 - 24]. Interestingly, patients in group with better functional status after post-stroke 3 months tend to have younger age (59.77 ± 14.18 versus 67.34 ± 7.02) and lower initial NIHSS (7.98 ± 5.26 versus 14.42 ± 5.48), and less likely to have stroke lesion in posterior limb of internal capsule (15.15% versus 57.14%), and have better cognitive function (total MMSE; 24.54 ± 5.97 versus 18.86 ± 7.02). In addition, hemorrhagic transformation, aphasia, visuospatial neglect, and depression were more frequent in patients with poor functional status after 3 months post-stroke.

The architecture of ANN with optimized hyperparameters was represented as shown in Fig. 3, which had 4 hidden layers with from 64 to 4 nodes. The proposed ANN model used rectified linear unit as activation function, Glorot uniform initializer as the way to set the initial weight, 0.3 dropout rate, Adagrad as optimizer [34], 0.02 learning rate, 5 batch size, 600 epochs and binary crossentropy as loss function. Sigmoid function as the classifier was placed at the last layer for prediction.

We traced the values of loss function using the test dataset during the entire learning process to validate the constructed model (Fig. 4.). The values of loss function continued to decline and plateaued during the learning process, which meant that overfitting of model did not occur.

The accuracy of the ANN model turned out to be 85.38 ± 6.15 (%), which was superior to that of models constructed by the other methods of machine learning (support vector machine; 83.46 ± 5.72 (%), K-nearest neighbors; 83.46 ± 7.10 (%)) (Table 2.). The recall of the ANN model was 93.11 ± 7.16 (%), which was the first rank among all the methodology of machine learning. The model by logistic regression showed the lowest prediction accuracy (77.69 ± 6.39 (%)). The prediction accuracy of the model constructed by tree-based methods (decision tree, random forest, and gradient boosting) was found to be 74.23 ± 6.89, 77.69 ± 7.65, and 81.92 ± 6.21 (%). In terms of precision of model, ANN showed the third highest performance following Bayesian with Bernoulli method, and K-nearest.

As show in Fig. 5, classification through the model derived through the method of bayes with Bernoulli led to misclassification of 28 individual dataset, and 16 (57.14%) of 28 individual dataset had over 90% rate of misclassification. This tendency also occurred in the classification model by logistic regression, but the number of misclassified data was increased in logistic regression classification model, which meant that the misclassified individual dataset appeared to be scattered. A classification model based on decision tree did not have the individual dataset with the probability of misclassification exceeding 25%. We also found that no data showed the probability of misclassification exceeding 75% in the classification model through kNN or ANN. ANN also showed 77 (76.24%) misclassified individual dataset, which had the most scattered distribution of misclassified individual dataset comparing with the other methods of machine learning.

Table 1. Characteristics of enrolled patients

| Characteristics | Total (n=101) | mBI* < 70 (n=35) | mBI* ≥ 70 (n=66) |
|---|---|---|---|
| Age at stroke onset (years) [range] | 62.40 ± 12.67 [ 19 - 79 ] | 67.34 ± 7.02 [ 48 - 77 ] | 59.77 ± 14.18 [ 19 - 79 ] |
| Sex (M / F) [n (%)] | 53 / 48 (52.48% / 47.52%) | 16 / 19 (45.71% / 54.29%) | 37 / 29 (56.06% / 43.94%) |
| Initial NIHSS | 10.07 ± 6.10 [ 1 - 24 ] | 14.42 ± 5.48 [ 4 - 24 ] | 7.98 ± 5.26 [ 1 - 23 ] |
| Involvement of IC posterior limb (+ / -) | 30 / 71 (29.70% / 70.30%) | 20 / 15 (57.14% / 42.86%) | 10 / 56 (15.15% / 84.85%) |
| MMSE | | | |
| Total | 22.63 ± 6.85 [ 0 - 30 ] | 18.86 ± 7.02 [ 0 - 28 ] | 24.54 ± 5.97 [ 7 - 30 ] |
| Attention | 2.55 ± 2.00 [ 0 - 5 ] | 1.38 ± 1.74 [ 0 - 5 ] | 3.14 ± 1.87 [ 0 - 5 ] |
| Memory | 4.36 ± 1.57 [ 0 - 6 ] | 3.93 ± 1.53 [ 0 - 6 ] | 4.58 ± 1.56 [ 0 - 6 ] |
| Presence of hemorrhagic transformation (+ / -) | 28 / 73 (27.72% / 72.28%) | 15 / 20 (42.86% / 57.14%) | 13 / 53 (19.70% / 80.30%) |
| Aphasia (+ / -) | 27 / 74 (26.73% / 73.27%) | 14 / 21 (40.00% / 60.00%) | 13 / 53 (19.70% / 80.30%) |
| Visuospatial neglect (+ / -) | 11 / 90 (10.89% / 89.11%) | 9 / 26 (25.71% / 74.29%) | 2 / 64 (3.03% / 96.97%) |
| Depression (+ / -) | 22 / 79 (21.78% / 78.22%) | 12 / 23 (34.29% / 65.71%) | 10 / 56 (15.15% / 84.85%) |

mBI; modified Barthel Index (measured after 3 months post-stroke), NIHSS; National Institutes of Health Stroke Scale, IC; Internal capsule, MMSE; Mini-Mental State Examination

Table 2. Comparison of performance between artificial neural network and the other method of machine learning

| Methods | Optimized hyperparameters | Accuracy (%) | Precision (%) | Recall (%) |
|---|---|---|---|---|
| Logistic regression | penalty, inverse of regularization strength | 77.69 ± 6.39 | 84.77 ± 8.59 | 81.68 ± 8.99 |
| K-nearest neighbors | the number of neighbors, weight function, algorithm used to compute the nearest neighbors | 83.46 ± 7.10 | 87.18 ± 7.49 | 88.69 ± 7.58 |
| Bayes with Bernoulli method | (-) | 80.77 ± 6.20 | **89.10 ± 7.26** | 82.31 ± 8.59 |
| Support vector machine | type of kernel, penalty parameter, class weight | 83.46 ± 5.72 | 85.18 ± 6.84 | 90.73 ± 5.35 |
| Decision tree | function to measure the quality of a split, strategy to choose the split at each node, maximum depth of tree | 74.23 ± 6.89 | 78.72 ± 7.28 | 84.90 ± 12.47 |
| Random forest | the number of trees in forest, function to measure the quality of a split, the number of features to consider, maximum depth of tree | 77.69 ± 7.65 | 80.02 ± 6.73 | 86.76 ± 7.00 |
| Gradient boosting | loss function, learning rate, the number of boosting stages, maximum depth of the individual regression estimators, function to measure the quality of a split | 81.92 ± 6.21 | 86.15 ± 10.09 | 87.85 ± 8.65 |
| Artificial neural network | the number of neurons at the fist layer, activation function, kernel initializer, dropout rate, optimizer, learning rate, batch size, epochs | **85.38 ± 6.15** | 86.24 ± 8.47 | **93.11 ± 7.16** |

Presented as mean ± standard error

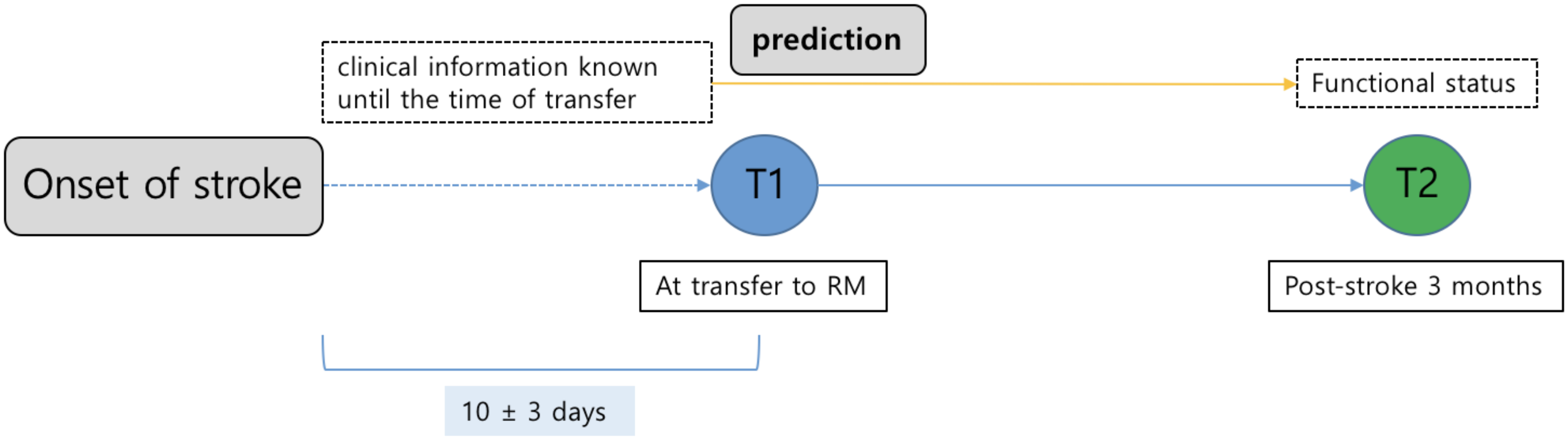


Figure 1. Scheme for predicting functional status after post-stroke 3 months

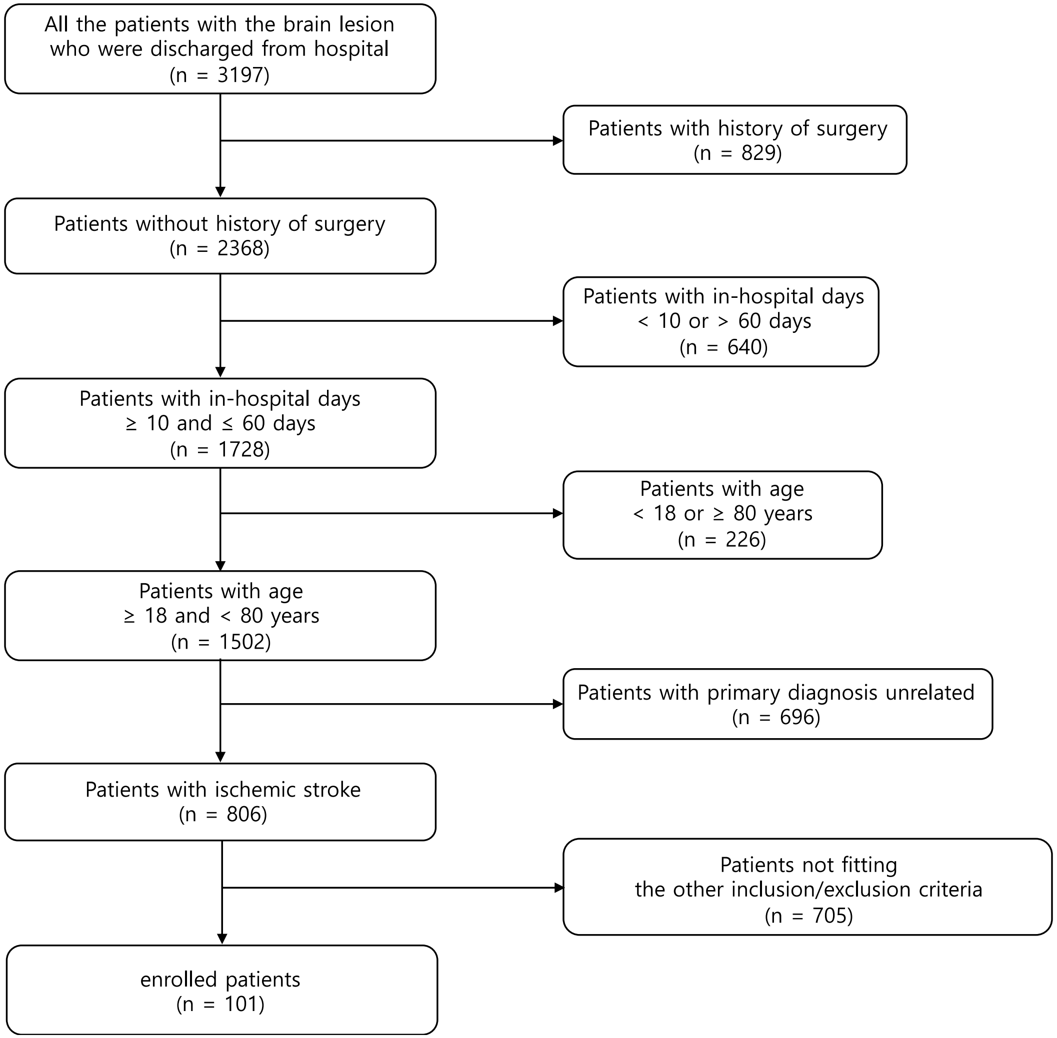


Figure 2. Flow of patient enrollment

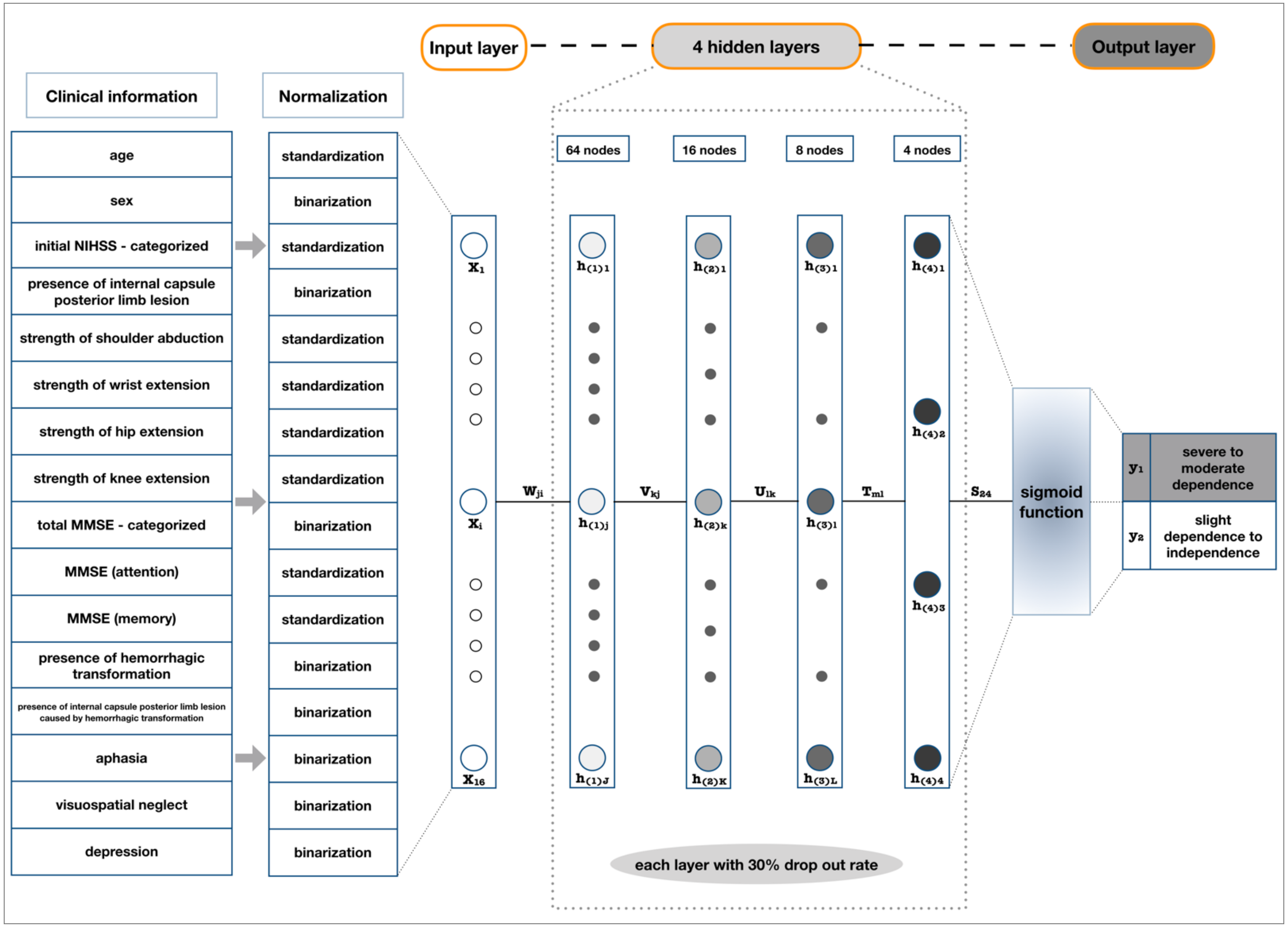


Figure 3. Framework of artificial neural network

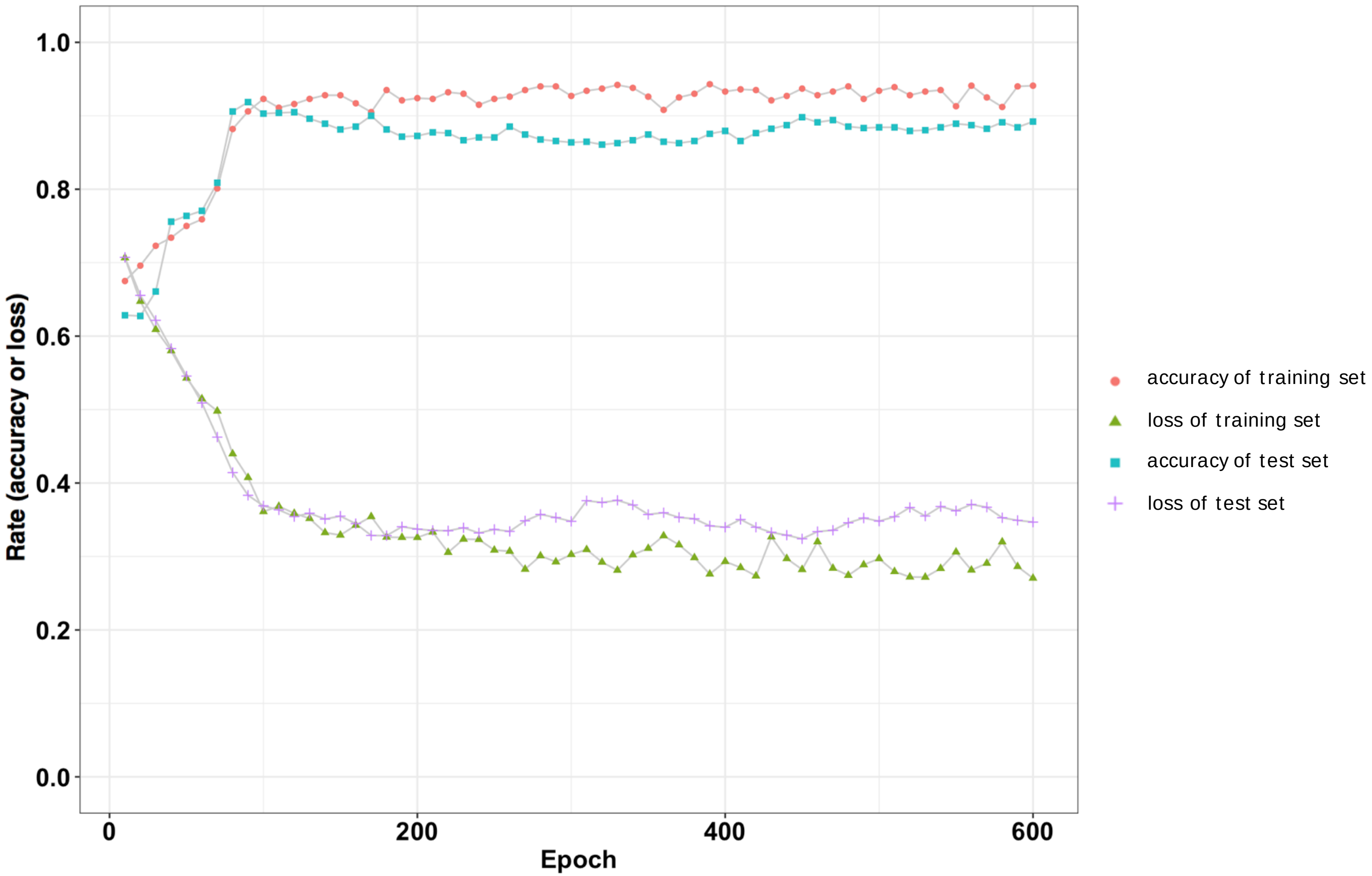

Figure 4. Training curve of artificial neural network

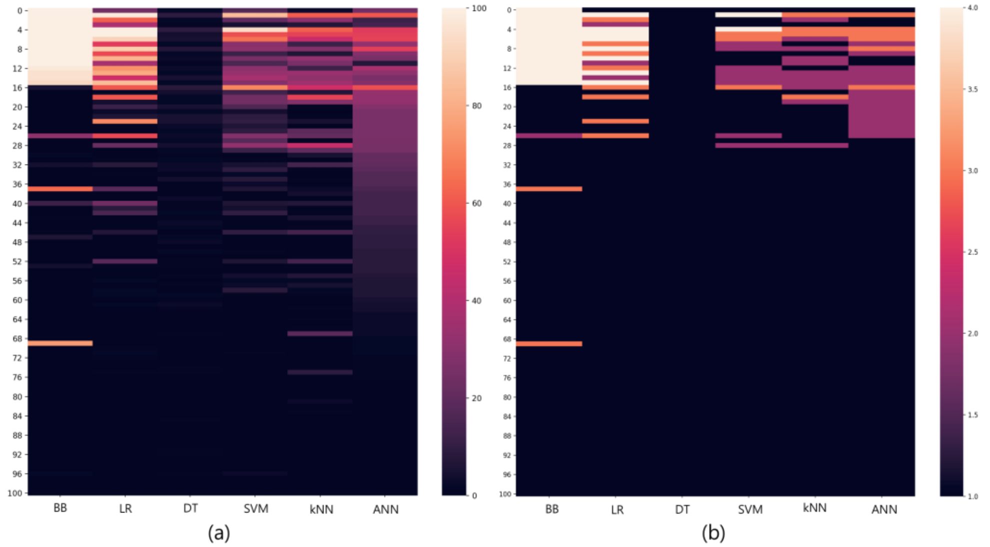


Figure 5. Comparison of misclassification for each data between the prediction model generated through the respective method of machine learning. BB; bayes with Bernoulli method, LR; logistic regression, DT; decision tree, SVM; support vector machine, kNN; k-nearest neighbors, ANN; artificial neural network. The respective enrolled subject was presented as the number of y-axis. (a) Number of misclassification for the dataset of the respective patient (The lighter color means the larger number of misclassification.); (b) Categorized number of misclassification for the dataset of the respective patient (≤ 25%; > 25% and ≤ 50%; > 50 and ≤ 75%; > 75%; The lighter color means the larger number of misclassification.).

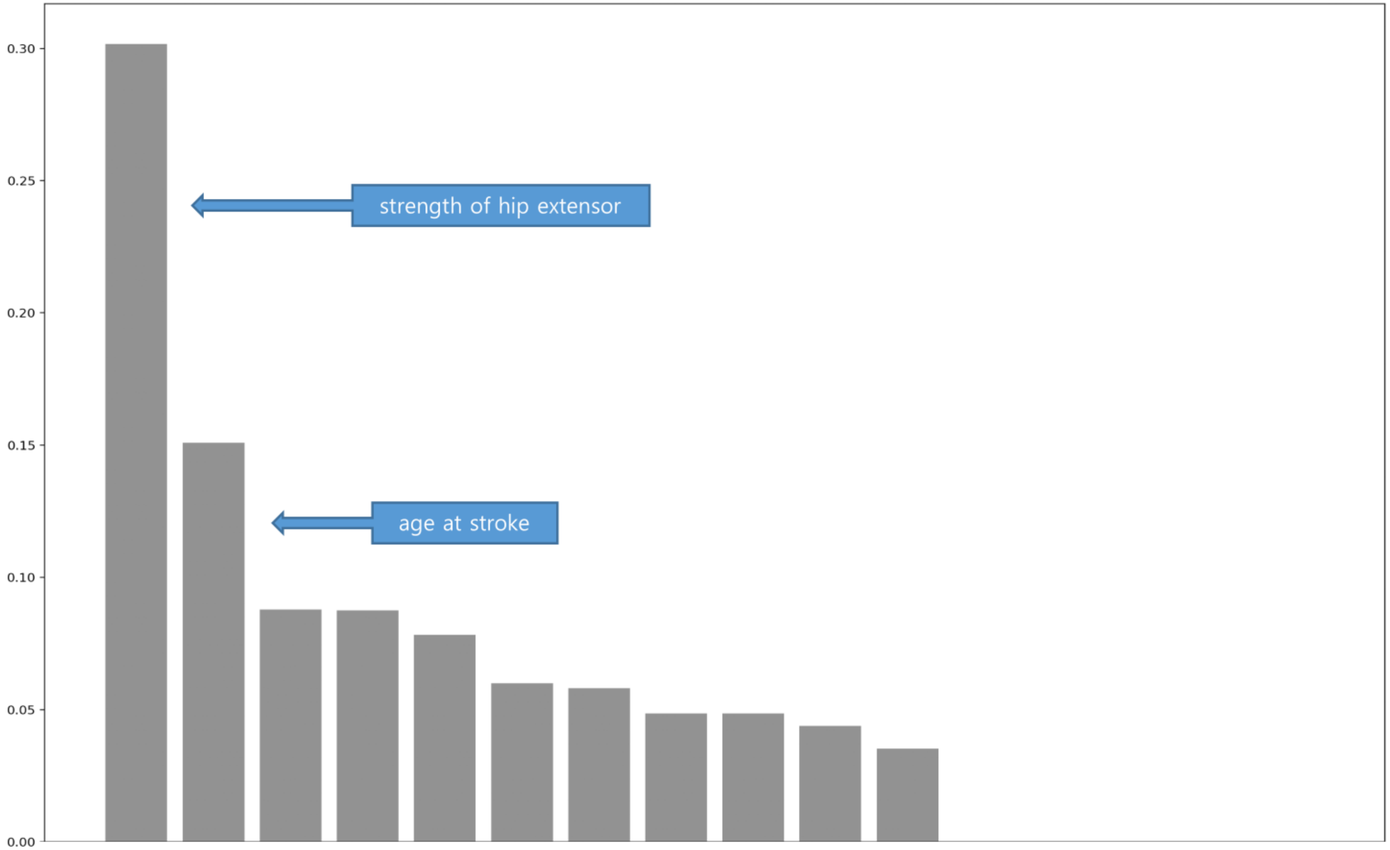

Figure 6. Importance of features in the prediction model generated through the algorithm of decision tree

# Discussion

The results showed that the model generated by ML can predict post-stroke recovery with considerable accuracy. It means that through the prediction model by ML, physicians can explain patients (or their caregivers) his/her post-stroke 3 months functional level with a high level of accuracy at the time of transfer to rehabilitation ward. This study also showed that ANN works better than other machine learning methods even with small dataset which has various type of clinical information.

The key question when dealing with ML classification is not whether one learning algorithm is superior to others, because ML explores the algorithms that can learn from and make predictions or diagnosis based on data. So that, when comparing the efficiency of the ML methodology, the characteristics of the data must be considered. The data used in the study is characterized by small sample size and high dimensioned feature compared with the sample size. Each feature also contains continuous variables, nominal variables, and variables with ordinal scale, which is characteristic of data that is often available in EMR. In the present study, we have demonstrated that the neural network works considerably well with dataset with small sample sizes, and heterogeneous and high dimensioned features. Various features of neural networks such as nonlinear, data-driven, universal function approximating, noise-tolerance, and parallel processing of large number of variables are especially desirable for data mining applications [8]. Therefore, it is expected that various applications of neural network will be possible based on EMR based data. We also showed the characteristics of the respective method of machine learning through misclassification frequency of individual dataset and distribution of the misclassified individual dataset. As shown in Fig. 5, the classification model of bayes with Bernoulli method and logistic regression showed little difference in the distribution of misclassified data according to the respective configuration of the training dataset and the test dataset, which meant that the frequency of misclassification was concentrated in a specific data set. On the other hand, in the classification model through decision tree, kNN, and ANN, there was difference in the distribution of misclassified data according to the configuration of training dataset and test dataset, which meant that the frequency of misclassification was scattered in each individual dataset. Through the results above, it can be said that the bayes with Bernoulli method and logistic regression do not allow flexible modeling based on the clinical data with small sample size and heterogenous features, while the decision tree, kNN and ANN do.

In the present study, kNN also had considerable accuracy. The kNN is based on the principle that the instances within a dataset will generally exist in close proximity to other instances that have similar properties. It is known to be robust to noise in training data and effective irrespective of size of dataset [8]. So, if we have enough computation power, we can use kNN potentially in clinical field. SVMs are particularly well suited for classification of complex but small- or medium-sized datasets [35]. It is also sensitive to the scale of the data. Therefore, it seems that SVM can produce a model with considerable accuracy if it is subjected to proper preprocessing for small-sized datasets, which was demonstrated in the present study. In the present study, the performance of the prediction model based on decision tree turned out to be relatively low. A decision tree is a classifier expressed as a recursive partition of the instance space, which is considered to be one of the most popular approaches for representing classifiers. The method of decision tree is considered to have advantages in the point that it is capable of handling both nominal and numeric input attributes and datasets which have errors or missing values [35]. In addition, the model generated through the method of decision tree by using clinical data has another advantage to compare the importance between the features of clinical data. In the present study, 'strength for hip extensor' and 'age at stroke' were identified as the most important clinical features, which is similar to the results of previous studies (Fig. 6) [4, 21, 24]. However, as decision trees use the “divide and conquer” method, they tend to perform less if many complex interactions are present [36], which explains the low accuracy and precision of the prediction model constructed by decision tree.

There were several limitations of the present study. First, we used the relatively small number of data for constructing the model via ANN, which could lead to the poor accuracy and over-fitting of the proposed model. The strict inclusion/exclusion criteria however, increased the homogeneity of the enrolled subjects with supratentorial ischemic stroke, which resulted in the improvement of the accuracy of the predictive model. As way to overcome the risk of model over-fitting, we used the normalization of the input variables, dropout rates of each layer in ANN, and initialization of weight value.

Second, patients with very good functional status tend not to be transferred to the rehabilitation department after acute treatment for stroke. On the other hand, patients with very poor functional status tend not to be evaluated with mBI after post-stroke 3 months. These could lead to biased enrollment of the patients, which could affect the accuracy of proposed model.

In addition, although volume of brain lesion has apparently known to be a factor associated with post-stroke recovery [37, 38], there was no consideration of brain lesion volume in the present study. However, quantitative consideration of brain lesions was achieved by using the presence of posterior limb of IC involvement as a feature constructing the prediction model. Finally, there were a few studies that showed functional recovery even after post-stroke 3 months, albeit with slight rise [39–41]. In the present study, we set the time of evaluation of mBI to 3 months or later to secure larger number of data, which is also limitation of the present study.

The degree of corticomotor pathway integrity assessed by neurophysiologic test such as transcranial magnetic stimulation had also been known to be a reliable prognostic factor for post-stroke recovery [42, 43], which was not considered in the present study. Genetic factors also can affect neural plasticity which leads to the different pattern of post-stroke recovery. Variation of genotype could have important roles in post-stroke recovery [44]. In further study, we need to consider taking account of genomic data as one of reliable prognostic factor on post-stroke recovery.

We expect the methodology of ML including ANN to give physicians the insight on the post-stroke functional recovery, as well as prognosis of diseases. In addition, we anticipate that the prediction model generated in the present study will be incorporated into future health information technology and used clinically, which would make additional data collection much easier. The accuracy of the prediction model is expected to be improved with new larger dataset including genomic data and signals obtained from biosensors.

# Conclusion

In the present study, we demonstrated that the prediction of function recovery after supratentorial ischemic stroke can be performed with a high degree of accuracy by various methodology of ML, with the highest accuracy in ANN.

# References


1. Stinear CM, Byblow WD, Ackerley SJ, Barber PA, Smith MC. Predicting Recovery Potential for Individual Stroke Patients Increases Rehabilitation Efficiency. Stroke 2017;48:1011-1019.

2. Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. Lancet 2011;377:1693-1702.

3. Lim JH, Han AR, Ryu BJ, Pyun SB. Long-term Functional Outcome and Related Factors in Stroke Patients. Brain Neurorehabil 2013;6:26-32.

4. Veerbeek JM, Kwakkel G, Van Wegen EEH, Ket JCF, Heymans MW. Early prediction of outcome of activities of daily living after stroke: A systematic review. Stroke 2011;42:1482-1488.

5. Boyd LA, Hayward KS, Ward NS, Stinear CM, Rosso C, Fisher RJ, Carter AR, Leff AP, Copland DA, Carey LM, Cohen LG, Basso DM, Maguire JM, Cramer SC. Biomarkers of stroke recovery: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. Int. J. Stroke 2017;12:480-493.

6. Stinear CM, Ward NS. How useful is imaging in predicting outcomes in stroke rehabilitation? Int. J. Stroke 2013;8:33-37.

7. Kim B, Winstein C. Can Neurological Biomarkers of Brain Impairment Be Used to Predict Poststroke Motor Recovery? A Systematic Review. Neurorehabil. Neural Repair 2017;31:3–24.

8. Kotsiantis SB. Supervised Machine Learning: A Review of Classification Techniques. Proceeding Conf. Emerg. Artif. Intell. Appl. Comput. Eng. Real Word AI Syst. with Appl. eHealth, HCI, Inf. Retr. Pervasive Technol. 2007;31:3–24.

9. Deo RC. Machine learning in medicine. Circulation 2015;132:1920–1930.

10. Krittanawong C, Zhang HJ, Wang Z, Aydar M, Kitai T. Artificial Intelligence in Precision Cardiovascular Medicine. J. Am. Coll. Cardiol. 2017;69:2657–2664.

11. Cheng Y, Wang F, Zhang P, Hu J. Risk Prediction with Electronic Health Records: A Deep Learning Approach. SIAM Int. Conf. Data Min. 2016;432–440.

12. Nguyen P, Tran T, Wickramasinghe N, Venkatesh S. Deepr: A Convolutional Net for Medical Records. IEEE J. Biomed. Heal. Informatics 2017;21:22–30.

13. Liang Z, Zhang G, Huang JX, Hu QV. Deep learning for healthcare decision making with EMRs. Proc. - 2014 IEEE Int. Conf. Bioinforma. Biomed. IEEE BIBM 2014 2014;556–559.

14. Bastien F, Lamblin P, Pascanu R, Bergstra J, Goodfellow I, Bergeron A, Bouchard N, Warde-Farley D, Bengio Y. Theano: new features and speed improvements. 2012;1–10.
15. Tokui S, Oono K, Hido S, Clayton J. Chainer: a Next-Generation Open Source Framework for Deep Learning. Proc. Work. Mach. Learn. Syst. Twenty-ninth Annu. Conf. Neural Inf. Process. Syst. 2015;1–6.
16. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mane D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016;
17. Stinear C. Prediction of recovery of motor function after stroke. Lancet. Neurol. 2010;9:1228–1232.
18. Kong K-H, Lee J. Temporal recovery of activities of daily living in the first year after ischemic stroke: a prospective study of patients admitted to a rehabilitation unit. NeuroRehabilitation 2014;35:221–226.
19. Stinear CM, Barber PA, Petoe M, Anwar S, Byblow WD. The

PREP algorithm predicts potential for upper limb recovery after stroke. Brain 2012;135:2527–2535.

20. Rondina JM, Park C, Ward NS. Brain regions important for recovery after severe post-stroke upper limb paresis. J. Neurol. Neurosurg. Psychiatry 2017;jnnp-2016-315030.

21. Kwah LK, Harvey LA, Diong J, Herbert RD. Models containing age and NIHSS predict recovery of ambulation and upper limb function six months after stroke: An observational study. J. Physiother. 2013;59:189–197.

22. Nijland RHM, Van Wegen EEH, Harmeling-Van Der Wel BC, Kwakkel G. Presence of finger extension and shoulder abduction within 72 hours after stroke predicts functional recovery: Early prediction of functional outcome after stroke: The EPOS cohort study. Stroke 2010;41:745–750.

23. Stinear CM, Barber PA, Petoe M, Anwar S, Byblow WD. The PREP algorithm predicts potential for upper limb recovery after stroke. Brain 2012;135:2527–2535.

24. Smith MC, Barber PA, Stinear CM. The TWIST Algorithm Predicts Time to Walking Independently After Stroke. Neurorehabil. Neural Repair 2017;31:955–964.

25. Gargano JW, Reeves MJ. Sex differences in stroke recovery and

stroke-specific quality of life: Results from a statewide stroke registry. Stroke 2007;38:2541-2548.

26. Lai S-M, Duncan PW, Dew P, Keighley J. Sex differences in stroke recovery. Prev. Chronic Dis. 2005;2:A13.

27. Schiemanck SK, Kwakkel G, Post MWM, Kappelle LJ, Prevo AJH. Impact of internal capsule lesions on outcome of motor hand function at one year post-stroke. J. Rehabil. Med. 2008;40:96-101.

28. Shelton F d. NAP, Reding MJ. Effect of Lesion Location on Upper Limb Motor Recovery After Stroke. Stroke 2001;32:107-112.

29. Puig J, Pedraza S, Blasco G, Daunis-i-Estadella J, Prados F, Remollo S, Prats-Galino A, Soria G, Boada I, Castellanos M, Serena J. Acute damage to the posterior limb of the internal capsule on diffusion tensor tractography as an early imaging predictor of motor outcome after stroke. Am. J. Neuroradiol. 2011;32:857-863.

30. Gainotti G. Relation between depression after stroke, antidepressant therapy, and functional recovery. J. Neurol. Neurosurg. Psychiatry 2001;71:258-261.

31. Pendlebury ST, Rothwell PM. Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: a systematic review and meta-analysis. Lancet Neurol. 2009;8:1006-1018.

32. Hage V. The NIH stroke scale: a window into neurological status.

Nurs. Spectr. 2011;24:44–49.

33. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Louppe G, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: Machine Learning in Python. 2012;12:2825–2830.

34. Zeiler MD. ADADELTA: An Adaptive Learning Rate Method. 2012;

35. Cios KJ, Pedrycz W, Swiniarski RW. Data Mining and Knowledge Discovery. 1998;

36. Pagallo G, Haussler D. Boolean Feature Discovery in Empirical Learning. Mach. Learn. 1990;5:71–99.

37. Turner RC, DiPasquale K, Logsdon AF, Tan Z, Naser ZJ, Huber JD, Rosen CL, Lucke-Wold BP. The role for infarct volume as a surrogate measure of functional outcome following ischemic stroke. J. Syst. Integr. Neurosci. 2016;2 VN-re:

38. Rehme AK, Volz LJ, Feis D-LL, Eickhoff SB, Fink GR, Grefkes C. Individual prediction of chronic motor outcome in the acute post-stroke stage: Behavioral parameters versus functional imaging. Hum. Brain Mapp. 2015;36 VN-r:4553–4565.

39. Langhorne P, Coupar F, Pollock A. Motor recovery after stroke: a

systematic review. Lancet Neurol. 2009;8:741-754.

40. Hofgren C, Björkdahl A, Esbjörnsson E, Stibrant-Sunnerhagen K. Motor Recovery after stroke: Cognition, ADL function and return to work. Acta Neurol. Scand. 2007;115:73-80.

41. Stinear C. Prediction of recovery of motor function after stroke. Lancet Neurol. 2010;9:1228-1232.

42. van Kuijk AA, Pasman JW, Hendricks HT, Zwarts MJ, Geurts ACH. Predicting Hand Motor Recovery in Severe Stroke: The Role of Motor Evoked Potentials in Relation to Early Clinical Assessment. Neurorehabil Neural Repair 2008;1545968308317578.

43. Di Lazzaro V, Profice P, Pilato F, Capone F, Ranieri F, Pasqualetti P, Colosimo C, Pravatã E, Cianfoni A, Dileone M. Motor cortex plasticity predicts recovery in acute stroke. Cereb. Cortex 2010;20:1523-1528.

44. Murphy TH, Corbett D. Plasticity during stroke recovery: From synapse to behaviour. Nat. Rev. Neurosci. 2009;10:861-872.

# Abstract in Korean

**목적:** 본 연구의 목적은 기계학습의 다양한 방법론 활용하여 뇌졸중 발생 후 2주 이내의 임상 정보를 바탕으로 뇌졸중 후 3개월째의 기능 상태를 높은 정확도로 예측하는 것이다.

**방법:** 2000년 1월부터 2017년 12월까지 대학 병원 재활의학과에서 퇴원한 환자들의 명단을 추출한 후, 포함/제외 기준에 맞는 환자를 추렸다. 재활의학과로 전과되는 시점까지 알려진 임상정보 중에서 뇌졸중 이후의 회복에 영향을 미치는 것으로 알려져 있고, 영향을 줄 것으로 예상되는 변수를 활용하여 예측 모델을 구성하였다. 예측 모델의 결과값은 3개월 이후 환자의 기능상태를 수정바델지수 70점을 기준으로 양분화한 값을 활용하였다. DNN 모델 및 여타의 기계학습을 활용한 모델의 하이퍼파라미터는 2배 교차검증을 통한 그리드 검색을 통해 최적화 하였다. 무작위 추출에 의해 구성된 훈련세트과 시험세트로 훈련 및 검증 과정을 10번 반복하였고, 각 반복시의 정확도를 평균화한 수치를 해당 모델의 정확도를 나타내는 값으로 정하였다.

**결과:** 2000년 1월부터 2017년 12월까지 대학 병원 재활의학과로부터 퇴원한 환자 5210명을 검토하였고, 그중 101명의 자료를 활용하여 예측 모델을 만들었다. 대상 환자들의 나이는 62.40 ± 12.67 [19 - 79] 였다. 뇌졸중 3개월 이후 더 좋은 기능을 가진 환자들은 나쁜 기능을 보였던

환자들에 비해 더 젊은 연령대에 분포했고 (59.77 ± 14.18 대 67.34 ± 7.02), 더 낮은 초기 NIHSS 점수 (7.98 ± 5.26 대 14.42 ± 5.48), 더 좋은 인지 기능상태 (MMSE 점수; 24.54 ± 5.97 대 18.86 ± 7.02) 를 보였다. 하이퍼 파라미터 최적화 과정을 통해 제시된 심층신경망 예측모델은 4개의 은닉층을 가지는 것으로 나타났고, 해당 모델을 통한 예측 정확도는 85.38 ± 6.15 (%) 로 여타의 기계학습을 통한 예측 모델의 성능보다 높은 것으로 나타났다.

**결론:** 본 연구에서, 환자의 임상데이터를 기반으로 예측모델을 생성할 때, 심층신경망을 포함한 다양한 기계학습 방법론을 활용할 경우, 상당한 수준의 정확도를 가짐을 알 수 있었다.

**주요어:** 기계학습, 인공신경망, 예측, 기능상태, 뇌졸중

**학번:** 2015-22073