Machine Learning Models for Predicting Hearing Prognosis in Unilateral Idiopathic Sudden Sensorineural Hearing Loss

Keon Vin Park[1],* ⓘD · Kyoung Ho Oh[2],* ⓘD · Yong Jun Jeong[2] ⓘD · Jihye Rhee[2,3] ⓘD
Mun Soo Han[2] ⓘD · Sung Won Han[1] ⓘD · June Choi[2] ⓘD

[1]School of Industrial Management Engineering, Korea University, Seoul; [2]Department of Otorhinolaryngology-Head and Neck Surgery,
Korea University Ansan Hospital, Korea University College of Medicine, Ansan; [3]Department of Otorhinolaryngology-Head and Neck Surgery,
Veterans Health Service Medical Center, Seoul, Korea

Objectives. Prognosticating idiopathic sudden sensorineural hearing loss (ISSNHL) is an important challenge. In our study, a dataset was split into training and test sets and cross-validation was implemented on the training set, thereby determining the hyperparameters for machine learning models with high test accuracy and low bias. The effectiveness of the following five machine learning models for predicting the hearing prognosis in patients with ISSNHL after 1 month of treatment was assessed: adaptive boosting, K-nearest neighbor, multilayer perceptron, random forest (RF), and support vector machine (SVM).

Methods. The medical records of 523 patients with ISSNHL admitted to Korea University Ansan Hospital between January 2010 and October 2017 were retrospectively reviewed. In this study, we analyzed data from 227 patients (recovery, 106; no recovery, 121) after excluding those with missing data. To determine risk factors, statistical hypothesis tests (e.g., the two-sample *t*-test for continuous variables and the chi-square test for categorical variables) were conducted to compare patients who did or did not recover. Variables were selected using an RF model depending on two criteria (mean decreases in the Gini index and accuracy).

Results. The SVM model using selected predictors achieved both the highest accuracy (75.36%) and the highest F-score (0.74) on the test set. The RF model with selected variables demonstrated the second-highest accuracy (73.91%) and F-score (0.74). The RF model with the original variables showed the same accuracy (73.91%) as that of the RF model with selected variables, but a lower F-score (0.73). All the tested models, except RF, demonstrated better performance after variable selection based on RF.

Conclusion. The SVM model with selected predictors was the best-performing of the tested prediction models. The RF model with selected predictors was the second-best model. Therefore, machine learning models can be used to predict hearing recovery in patients with ISSNHL.

Keywords. *Sudden Hearing Loss; Machine Learning; Prognosis; Prediction*

INTRODUCTION

Idiopathic sudden sensorineural hearing loss (ISSNHL) refers to rapid-onset hearing loss of more than 30 dB at least three consecutive frequencies within 72 hours [1-3]. This disease is a common otologic emergency, with an estimated incidence and prevalence of 10–20 per 100,000 person years and 5–160 per 100,000 person years, respectively [1,2]. Although the exact etiology and pathogenesis of ISSNHL remain ambiguous, viral disease and vascular compromise have been strongly suggested as causes. In addition, retrocochlear lesions such as vestibular schwannoma

and strokes should be considered [3,4]. Because ISSNHL is diagnosed only by describing the symptoms of hearing loss, regardless of the cause, treatment methods are inevitably diverse. Despite the lack of standardized treatment guidelines for ISSNHL, steroid therapy—including systemic, intratympanic, or both in combination—is widely accepted as the most effective option [3-5]. Numerous prognostic factors have also been identified in recent years, including age, duration from onset to treatment, initial hearing level, vertigo/dizziness, smoking, alcohol use, and blood test parameters [4,6]. Based on these diverse prognostic factors, the ability of various models to prognosticate ISSNHL has been studied. In addition to traditional statistical techniques such as multiple-regression and logistic-regression (LR) models, researchers have attempted to predict the prognosis of ISSNHL using Bayesian cure-rate models [7-10]. Recently, some researchers have applied various machine learning techniques [1,11].

In a previous report, several machine learning models, including LR, support vector machines (SVMs), multilayer perceptrons (MLPs), and deep-belief networks (DBN), were applied to predict the hearing outcomes of sudden sensorineural hearing loss [1]. The sample size of that study was 1,220, and 149 variables were analyzed. For variable selection, the study utilized the results of univariate statistics ($P<0.05$) and expert knowledge [1]. In the present study, the effectiveness of five potential machine learning models (K-nearest neighbor [KNN], random forest [RF], SVM, adaptive boosting [AdaBoost], and MLP) for predicting hearing recovery in ISSNHL patients was investigated. The sample size of our study was 227, and 32 variables were. Furthermore, we conducted variable selection based on an RF algorithm and medical domain knowledge. This study aimed to assess the effectiveness of predicting hearing recovery in patients with ISSNHL after 1 month of treatment using the five aforementioned machine learning models.

MATERIALS AND METHODS

Study population
We performed a retrospective study of in-patients admitted to Korea University Ansan Hospital for treatment of ISSNHL. We inspected 523 unilateral ISSNHL patients enrolled from January 2010 to October 2017. In this study, we analyzed the hearing improvement after 1 month of treatment for 227 patients with all variables analyzed. The diagnostic criteria for ISSNHL included sudden hearing loss (30 dB or more) for at least three contiguous frequencies over 3 days. We evaluated their entire medical history, laboratory examinations, and audiologic results. Exclusion criteria were: (1) bilateral ISSNHL, (2) conductive hearing loss >10 dB, (3) vestibular schwannoma, and (4) data with missing values.

All included patients were hospitalized and treated using systemic steroids (e.g., methylprednisolone 64 mg, tapering for 14 days or dexamethasone 5 mg intravenously three times/day for 5 days and then tapered over 7 days) with/without intratympanic dexamethasone injection (ITDI; 1–4 times). During the initial hospital day, each patient underwent a pure-tone audiogram, which was then repeated posttreatment 1 month later. We measured hearing thresholds at 0.125, 0.25, 0.5, 1, 2, 3, 4, and 8 kHz. Siegel's criteria were classified as follows: (1) complete recovery included final hearing levels better than 25 dB, (2) partial recovery was >15-dB gain and final hearing levels between 25 and 45 dB, (3) slight recovery was >15-dB gain and final hearing poorer than 45 dB, and (4) no improvement was <15-dB gain, or final hearing poorer than 75 dB [12].

In this study, patients with hearing gains of at least "partial recovery" based on Siegel's criteria, i.e., complete recovery and partial recovery were considered to be the recovery group. Patients under the "slight recovery" classification according to Siegel's criteria were regarded as the no-recovery group. This study obtained the approval of the Institutional Review Board at Korea University Ansan Hospital (IRB. No. 2018AS0044). The overall procedure of this study is shown in Fig. 1.

Original variables and variable selection
The original number of variables was 32 (31 predictors and 1 response), as extracted from demographic data, medical records, inner-ear symptoms, pure-tone audiometry, and laboratory data. Hearing improvements after 1 month of treatment were based on Siegel's criteria, which considered patients under complete recovery and partial recovery as the recovery group (positive in classification). The others were the no-recovery group (negative in classification). In detail, there were 11 independent categorical variables: sex, smoking status, hypertension, diabetes mellitus, hyperlipidemia, stroke, chronic kidney disease, myocardial infarction/angina, dizziness, tinnitus, and timing of ITDI. The categorical variables were mostly binary variables with "yes/no" values. The timing of ITDI was classified into three conditions: never (0), 0–12 days from onset (1), and more than 13 days from onset (2). The smoking status was divided into three conditions: never (0), current (1), and ex-smoker (2). Then, the timing of ITDI and smoking statuses were changed to binary variables having the same number of classes for KNN and MLP. Additionally, there were 20 independent continuous variables, including age

at time of diagnosis, pack-year for smoking, systolic blood pressure, diastolic blood pressure, duration from onset to treatment (day), hemoglobulin, erythrocyte sedimentation rate, activated partial thromboplastin time (aPTT), blood urea nitrogen (BUN), creatinine (Cr), initial hearing by frequency (0.125, 0.25, 0.5, 1, 2, 3, 4, and 8 KHz), prothrombin time (PT) result (second), and PT result (%). All continuous variables were scaled into a range between 0 and 1 via min–max scaling to fairly compare the five machine learning algorithms. Because KNN and MLP are very sensitive to scaling, it was necessary to scale the data. In this study, we used variable selection to improve the predictive performance for hearing recovery.

For variable selection using a RF, mean decrease accuracy and Gini were used [13]. The greater the decrease in the accuracy of RF after a permutation of a specific variable, the more important that variable [13]. Therefore, variables with a large mean decrease in accuracy were more important when predicting hearing outcome. However, RF is based on Gini, which is a splitting criterion.
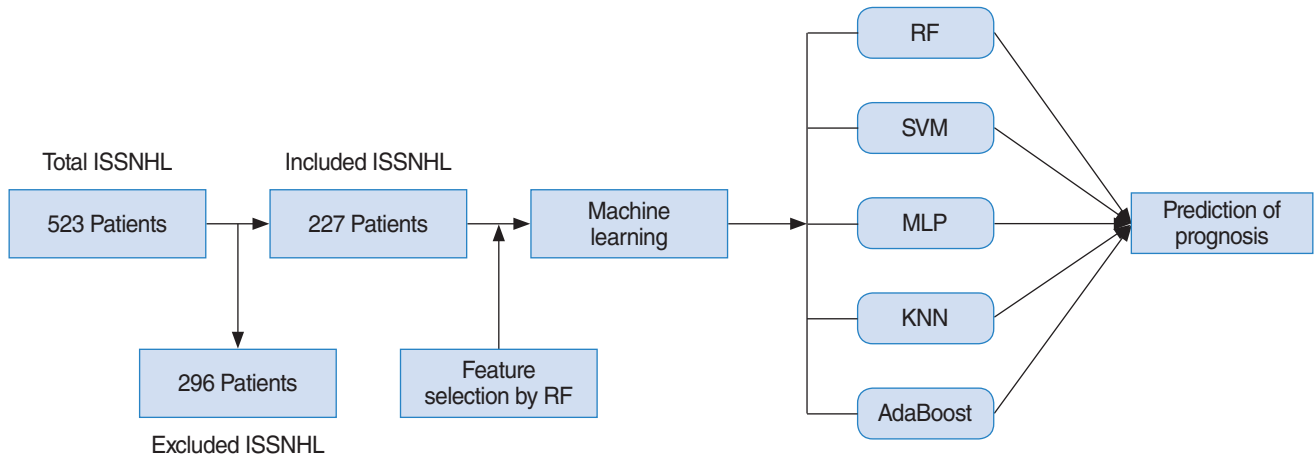

Fig. 1. Flowchart detailing patient inclusion and exclusion. ISSNHL, idiopathic sudden sensorineural hearing loss; RF, random forest; SVM, support vector machine; MLP, multilayer perceptron; KNN, K-nearest neighbor; AdaBoost, adaptive boosting.
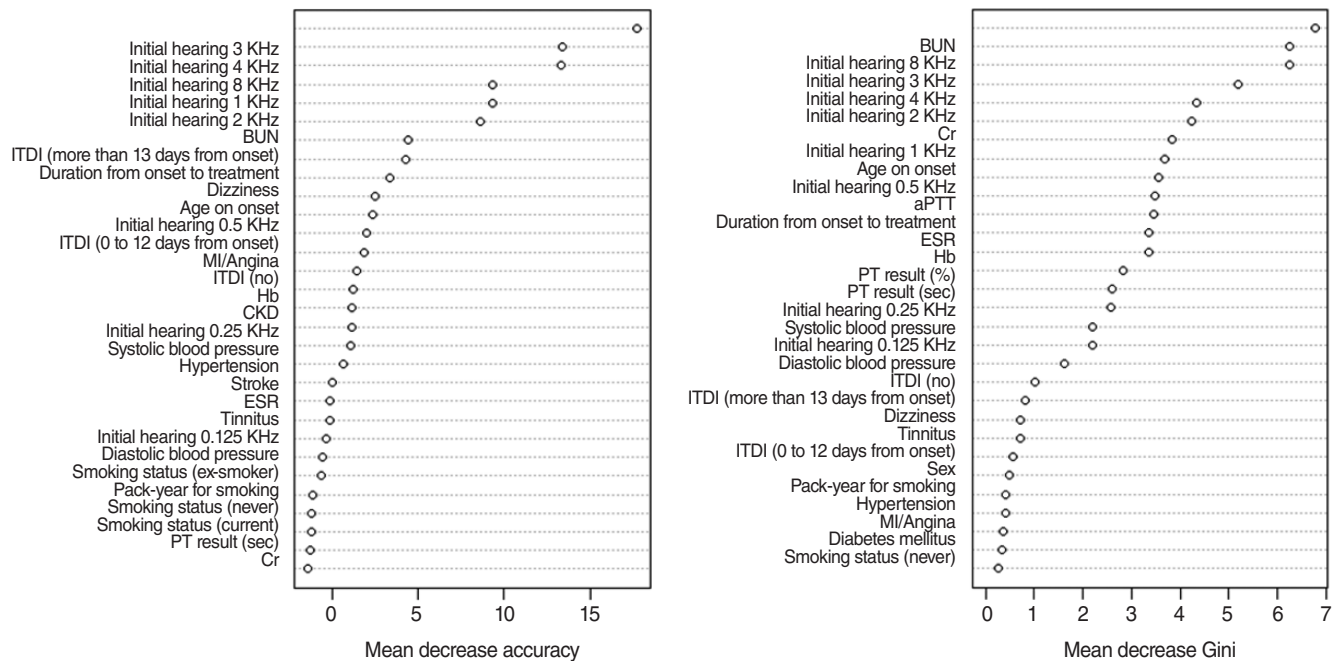
Fig. 2. Variable selection by a random forest (RF) using mean decreases in accuracy and the Gini index, according to which the importance score of each variable was calculated. The top five variables for each criterion, excluding initial hearing by frequency (0.125, 0.25, 0.5, 1, 2, 3, 4, and 8 KHz), were included and applied in the five machine learning models. BUN, blood urea nitrogen; ITDI, intratympanic dexamethasone injection; MI, myocardial infarction; Hb, hemoglobulin; CKD, chronic kidney disease; ESR, erythrocyte sedimentation rate; PT, prothrombin time; Cr, creatinine; aPTT, activated partial thromboplastin time.

When a decrease in Gini is significant, the parent node is split into two child nodes. Therefore, to evaluate the importance of a specific variable, the mean decrease of Gini can be calculated for all decision trees in an RF. The greater the mean decrease of Gini, the more important the variable when predicting a hearing outcome [13]. When training an RF with 1,000 trees and six variables at random, the variable importance score is calculated according to the mean decrease accuracy and mean decrease Gini, as shown in Fig. 2.

The top five variables, excluding initial hearing by frequency (0.125, 0.25, 0.5, 1, 2, 3, 4, and 8 KHz) in each criterion, are included, and all values for initial hearing by frequency are included. As a result, the 15 variables, including initial hearing by frequency, BUN result, timing of ITDI, duration from onset to treatment (days), dizziness, age of onset, Cr result, and aPTT result are shown in Fig. 2. Additionally, original variables and selected variables were applied to the five machine learning models to compare the effect on the performance of the two sets of variables in predicting hearing outcome after 1 month of treatment.

Statistical analysis methods and model development

A total of 523 enrolled ISSNHL patients were admitted between January 2010 to October 2017. The number of missing data points for each variable varied, especially for recovery status, which had the most missing data because of follow-up losses. Therefore, we implemented a complete deletion method using a rule in which patients having at least one missing value in variables were excluded. As a result, we obtained 227 patients in total, 106 for the recovery group and 121 for the no-recovery group. All statistical hypothesis tests were two-sided. Statistical tests were implemented using R v.3.5.0 (The R Foundation, Vienna, Austria). In this study, to inspect the differences among groups with recovery and those with no recovery, two-sample *t*-tests were used for the continuous random variables and chi-square or Fisher's exact tests for the categorical random variables. The difference was considered statistically significant at a two-tailed *P*-value of less than 0.05. According to our results, initial hearing by frequency (0.125, 0.25, 0.5, 1, 2, 3, 4, and 8 KHz), age on onset, BUN result, hypertension, dizziness, and the timing of ITDI, turned out to be statistically significantly different between the recovery and no-recovery groups under a significance level of 0.05. The statistical test results are listed in Table 1.

We split the data (n=227) into training and test sets. The ratio of the sizes of the training set and test set was 7 to 3, respectively, which resulted in a training set with a sample size of 158 and a test set with a sample size of 69. We then evaluated five different machine learning models and implemented three-fold cross validation on the training set to determine the hyperparameters for each model to optimize the validation accuracy by calculating the accuracy of the validation set. Therefore, we selected the optimal hyperparameters for each model that maximized the accuracy of the validation set. Then, we evaluated our model using

Table 1. Clinical characteristics of patients according to their recovery status

Variable	No recovery (n=121)	Recovery (n=106)	*P*-value
Pack-years of smoking	3.35±10.33	4.11±23.73	0.761
Age at the time of diagnosis (yr)	51.87±11.16	47.58±11.87	0.006
Systolic BP (mmHg)	121.56±13.01	120.33±12.87	0.475
Diastolic BP (mmHg)	76.27±7.85	75.73±8.57	0.619
Duration from onset to treatment (day)	5.13±6.54	4.90±9.61	0.813
Hb	13.57±1.50	13.48±1.25	0.644
ESR	27.09±21.06	24.24±18.20	0.274
PT (sec)	11.06±2.68	10.80±0.80	0.985
PT (%)	107.91±15.46	108.43±12.24	0.780
aPTT	28.23±4.13	28.20±2.68	0.059
BUN	16.96±8.45	13.95±5.33	0.001
Cr	1.14±1.65	0.90±0.19	0.114
Initial hearing (kHZ)			
0.125	74.73±38.47	59.92±36.62	0.003
0.25	74.92±34.41	59.95±31.13	<0.001
0.5	77.56±32.83	61.37±26.37	5.387e-05
1	79.55±34.30	59.86±27.52	2.988e-06
2	81.45±32.92	57.45±28.03	1.133e-08
3	83.51±31.83	56.46±28.58	1.27e-10
4	86.32±30.67	59.34±29.46	1.232e-10
8	92.23±29.98	65.05±32.04	3.655e-10
Male:female	43:78 (47.78:52.22)	27:79 (25.47:74.53)	0.135
Smoking status			0.883
0 (Never)	103 (85.12)	89 (83.96)	
1 (Current)	16 (13.23)	16 (15.09)	
2 (Ex-smoker)	2 (1.65)	1 (0.95)	
Hypertension	34 (28.10)	15 (14.15)	0.017
Diabetes mellitus	27 (22.31)	15 (14.15)	0.159
Stroke	1 (0.83)	1 (0.94)	1
Chronic kidney disease	3 (2.48)	0	0.25
Myocardial infarction/angina	6 (4.96)	2 (1.89)	0.289
Dizziness	44 (36.36)	16 (15.09)	<0.001
Tinnitus	77 (63.64)	79 (74.53)	0.105
The timing of ITDI			<0.001
0 (No treatment)	25 (20.66)	42 (39.62)	
1 (0–12 days from onset)	74 (61.16)	56 (52.83)	
2 (After 13 days from onset)	22 (18.18)	8 (7.55)	
Hyperlipidemia	3 (2.48)	5 (4.72)	0.478

Values are presented as mean± standard deviation or number (%).
BP, blood pressure; Hb, hemoglobin; ESR, erythrocyte sedimentation rate; PT, prothrombin time; APTT, activated partial thromboplastin time; BUN, blood urea nitrogen; Cr, creatinine; ITDI, intratympanic dexamethasone injection.

the selected hyperparameters on the test set. The accuracy of the test set for each model is listed in Table 2. The training set was randomly divided into three subsets that were similar in size, among which, two were used to train each model, and one was used as the validation subset to verify model efficiency.

In this study, several model evaluation criteria such as accuracy,

Table 2. Performance of KNN, SVM, RF, AdaBoost, and MLP on the test set

Model	Variable set	Accuracy (%)	F-score	ROC-AUC	95% CI for the ROC-AUC	Precision (%)	Recall (%)
KNN	31 Predictors	60.87	0.61	0.62	0.50–0.73	55.26	67.74
	15 Predictors	65.22	0.64	0.65	0.54–0.77	60	67.74
SVM	31 Predictors	60.87	0.49	0.59	0.48–0.70	59.09	41.94
	15 Predictors	75.36	0.74	0.76	0.65–0.86	70.59	77.42
RF	31 Predictors	73.91	0.73	0.74	0.64–0.85	68.57	77.42
	15 Predictors	73.91	0.74	0.75	0.64–0.85	67.57	80.65
AdaBoost	31 Predictors	69.57	0.71	0.71	0.61–0.81	61.90	83.87
	15 Predictors	72.46	0.73	0.74	0.63–0.84	65	83.87
MLP	31 Predictors	66.67	0.58	0.65	0.54–0.76	66.67	51.61
	15 Predictors	72.46	0.71	0.73	0.62–0.83	67.65	74.19

KNN, K-nearest neighbor; SVM, support vector machine; RF, random forest; AdaBoost, adaptive boosting; MLP, multilayer perceptron; ROC-AUC, area under the receiver operating characteristic curve; CI, confidence interval.

recall (sensitivity), and F-score were utilized for machine learning techniques and to compare the prediction performance of the five models. The machine learning models are summarized in detail below.

K-nearest neighbor

KNN is based on the assumption that the observations within a dataset generally exist closely to other observations having similar attributes. If the observations have a classification label, the label of an unclassified observation can be identified by considering the class of its nearest neighbors. KNN finds the k-nearest observations to the test observation and determines its class by majority voting (Fig. 3). While implementing three-fold cross validation, the hyperparameter, K, was tuned until the training subset produced the highest accuracy. The optimal parameter was then applied to the test set, which did not affect the training or cross-validation procedures [14].

Random forest

RF uses bootstrap sampling to create new training samples. The new training samples are then used to train m different decision trees. As a result, this algorithm aggregates prediction results from m different decision trees, usually via majority voting [13]. The overall model is shown in Fig. 4. Additionally, we determined two parameters using three-fold cross validation, which are the number of decision trees to train and the number of predictors to sample from the original variables.

Support vector machine

SVM is used to establish a separating hyperplane having a maximum margin. This model leads to a good generalization ability and can classify nonlinear data more easily than a linear classifier such as logistic regression. The concept of SVM is shown in Fig. 5. To classify nonlinear data, a kernel function such as a Gaussian kernel is required [15]. In this study, a Gaussian kernel function was used to classify the nonlinear data. Furthermore, we determined two hyperparameters for SVM, which are the regu-
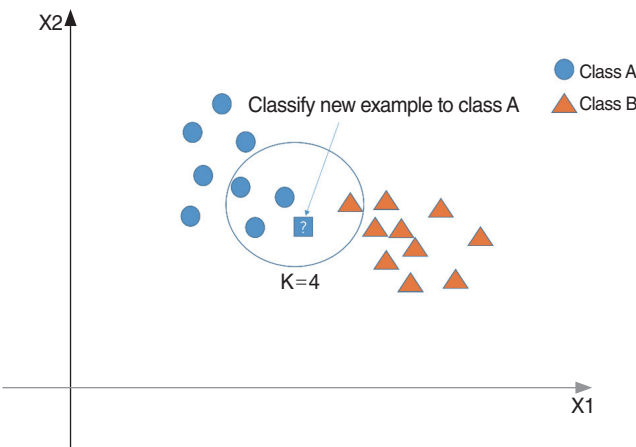

Fig. 3. K-nearest neighbor.

larization parameter (C) and spread-parameter sigma (σ) to optimize three-fold cross-validation accuracy.

AdaBoost

Boosting is a machine learning approach based on the concept of creating high-accuracy models by combining many low-accuracy ones, called weak learners. The AdaBoost algorithm by Freund is the most widely used boosting algorithm [16]. It is a very accurate classifier that uses a decision tree as the base model, i.e., the weak learner. It is used to train a multiple decision trees based on weight-updated and aggregated results from classifier models. The overall AdaBoost algorithm is shown in Fig. 6. In this study, the AdaBoost model was trained using the "fast Adaboost" package in R ver. 3.5.0 (R Foundation for Statistical Computing, Vienna, Austria). Additionally, we determined two parameters, the number of decision trees and maximum depth, using three-fold cross validation accuracy.

Multilayer perceptron

The MLP is a feed-forward neural network that uses the back-propagation method for training. This model can classify nonlin-
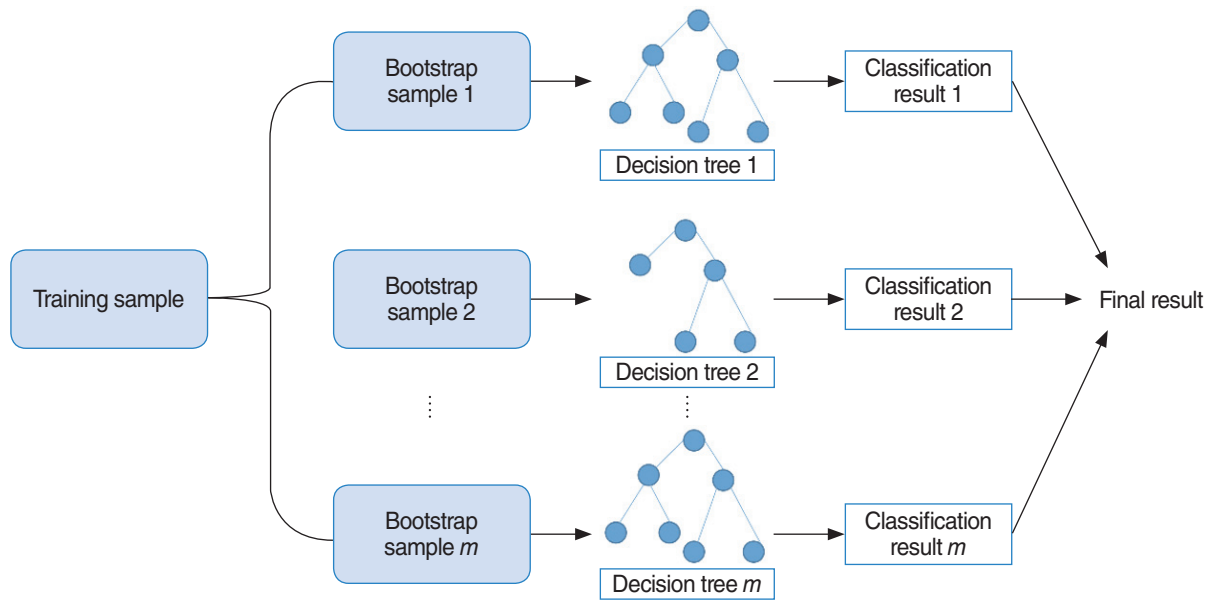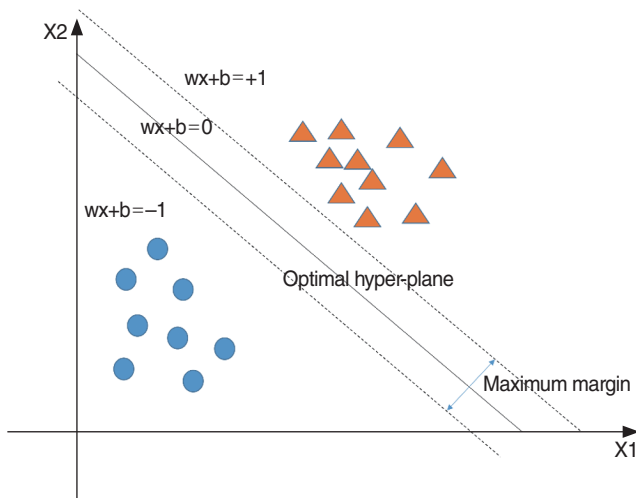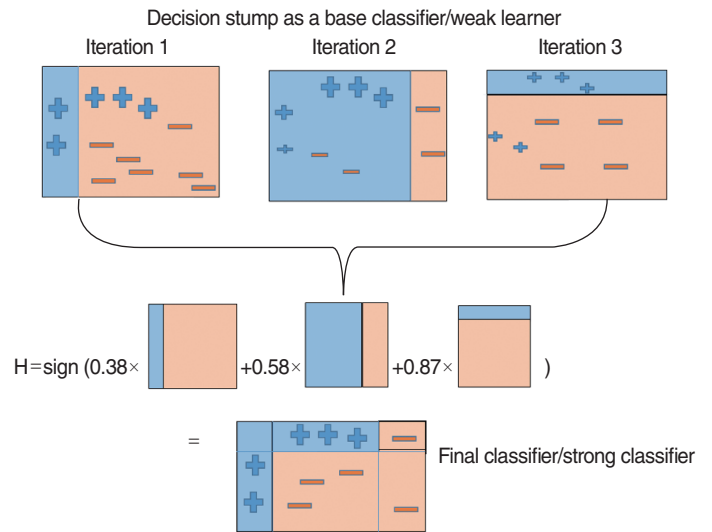
Fig. 4. Random forest algorithm.

Fig. 5. Support vector machine. The optimal hyperplane consists of w and b which are the the optimal parameters estimated by data.

Fig. 6. Adaptive boosting algorithm. H means the predicted class.

ear data. Generally, MLP comprises one input layer, one or more hidden layers, and one output layer. The overall architecture of this model is shown in Fig. 7. It can be used both for classification and for regression problems. In this study, we used it as a binary classifier to classify in-patients after 1 month of treatment. The original and selected variables were utilized as input variables. Thus, they determined the number of input nodes. All nodes of each hidden layer were fully connected to the nodes of the previous layer. Because this study regarded a binary classification problem, the number of nodes of the output layer was one. This node represents the probability of hearing recovery after 1 month of treatment. We determined two hyperparameters, which are

the number of hidden layers and its activation function. In addition, the learning rate was 0.1, number of epochs was 10,000, and batch size was 20. In this study, MLP was developed by adopting the "deepnet" package in R ver. 3.5.0.

RESULTS

Clinical characteristics of patients according to their recovery status

The study population consisted of 227 unilateral ISSNHL patients, including 121 who did not recover and 106 who recovered. There

were 70 men and 157 women, with a mean and standard deviation (SD) age of 46.43±14.77 years and 51.3±99.65 years, respectively. The clinical characteristics of patients according to
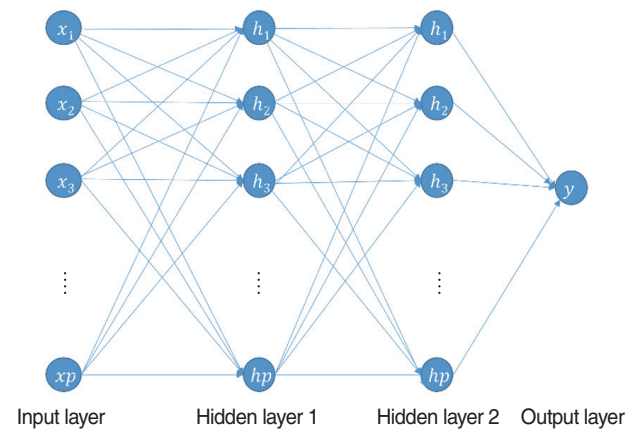

Fig. 7. Multilayer perceptron. χ, input variable; h, hidden variable; y, predicted probability of hearing recovery after 1 month of treatment.

their recovery status are listed in Table 1.

Continuous and categorical random variables were presented as mean±SD and percentages, respectively. The clinical and demographic characteristics of the patients who did or did not recover were compared using the two-sample t-test for continuous random variables and the chi-square test or Fisher exact test for categorical random variables.

All statistical hypothesis tests were two-sided. Statistical tests were implemented using R ver. 3.5.0. A mean difference was considered to be statistically significant when the P-value was less than 0.05. Statistically significant differences were found between the recovery and no-recovery groups according to initial hearing by frequency (0.125, 0.25, 0.5, 1, 2, 3, 4, and 8 KHz), age at onset, BUN, hypertension, dizziness, and the timing of ITDI. The statistical test results are presented in Table 1.

Model performance results

The SVM model with selected variables was determined to be the best-performing prediction model (accuracy, 75.36%; F-score,
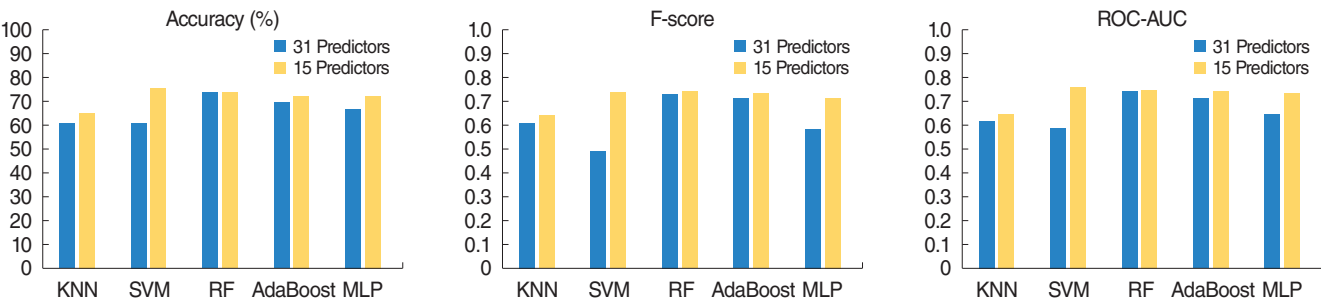

Fig. 8. Test-set accuracy, F-score, and area under the receiver operating characteristic curve (ROC-AUC) of five machine learning models. The support vector machine (SVM) model with selected predictors achieved both the highest accuracy (75.36%) and highest F-score (0.74) on the test set. The random forest (RF) model with selected variables demonstrated the second-highest accuracy (73.91%) and F-score (0.74). The RF model with the original variables showed the same accuracy (73.91%) as that of the RF model with selected variables, but with a lower F-score (0.73). All the tested models, except RF, demonstrated better performance after variable selection based on RF. KNN, K-nearest neighbor; AdaBoost, adaptive boosting; MLP, multilayer perceptron.
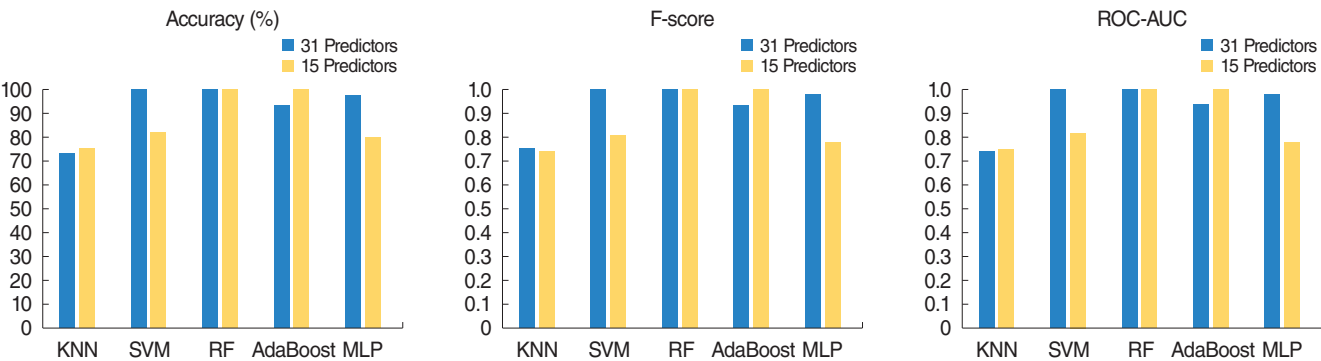

Fig. 9. Training set accuracy, F-score, and area under the receiver operating characteristic curve (ROC-AUC) of five machine learning models. The support vector machine (SVM) model with the original predictors, the random forest (RF) model with the original predictors and selected predictors, and the adaptive boosting (AdaBoost) model with selected predictors achieved the highest accuracy (100%) and highest F-score (1.0) on the training set. The K-nearest neighbors (KNN) and AdaBoost models demonstrated better performance after variable selection based on RF. MLP, multilayer perceptron.

Table 3. Performance of KNN, SVM, RF, AdaBoost, and MLP on the training set

Model	Variable set	Accuracy (%)	F-score	ROC-AUC	95% CI for the ROC-AUC	Precision (%)	Recall (%)
KNN	31 Predictors	73.42	0.75	0.74	0.67–0.81	68.13	82.67
	15 Predictors	75.32	0.74	0.75	0.69–0.82	73.68	74.67
SVM	31 Predictors	100	1	1	1–1	100	100
	15 Predictors	82.28	0.81	0.82	0.76–0.88	81.33	81.33
RF	31 Predictors	100	1	1	1–1	100	100
	15 Predictors	100	1	1	1–1	100	100
AdaBoost	31 Predictors	93.67	0.94	0.94	0.90–0.98	89.16	98.67
	15 Predictors	100	1	1	1–1	100	100
MLP	31 Predictors	98.1	0.98	0.98	0.96–1	98.65	97.33
	15 Predictors	79.75	0.78	0.80	0.73–0.86	81.16	74.67

KNN, K-nearest neighbor; SVM, support vector machine; RF, random forest; AdaBoost, adaptive boosting; MLP, multilayer perceptron; ROC-AUC, area under the receiver operating characteristic curve; CI, confidence interval.

0.74), and the RF model with selected variables was the second-best model (accuracy, 73.91%; F-score, 0.74). The accuracy of the RF model with the original variables was the same as that of the RF model with selected variables (accuracy, 73.91; F-score, 0.73). However, the prediction time for the test set (n=69) of the SVM model with selected variables was nearly zero (R.3.5.0, RAM 8 GB), while that of RF with selected variables was 0.01 s (R.3.5.0, RAM 8 GB). Thus, the SVM model with selected variables performed better than the RF model with selected variables, considering both accuracy and the computation time required for the classification under real clinical circumstances. To summarize, the SVM model with 31 predictors was the best model, whereas the second-best model was RF with 15 predictors (Table 2, Fig. 8). Additionally, the results of the training set are provided in Table 3 and Fig. 9.

DISCUSSION

In our analysis, the RF model with the original variables predicted the prognosis of ISSNHL. In a recent study, several machine learning models [1] were used to predict the hearing outcomes of ISSNHL patients. In that study [1], the accuracy of prediction of hearing outcomes was 67.45%–73.32% for LR, 68.14%–73.41% for SVM, 63.90%–74.03% for MLP, and 60.67%–77.58% for DBN. A total of 1,220 inpatients with ISSNHL were admitted between June 2008 and December 2015, and four-fold cross-validation was applied to the dataset. In contrast, our study included a total of 227 inpatients with ISSNHL between January 2010 and October 2017. Three-fold cross-validation was applied to the dataset. In our study, the accuracy of the prediction of hearing outcomes was 75.36% for SVM, 73.91% for RF, 72.46% for AdaBoost, 72.46% for MLP, and 65.22% for KNN. Thus, compared with the previous study, we trained a model with a similar generalization performance, but using a much smaller sample size. Furthermore, in another study, the accuracy of prediction of hearing outcomes was 76.6% for AdaBoost, 76.9% for RF, 81.9% for

MLP, and 83.0% for SVM. In that study, 10-fold cross-validation was implemented, and the study population was 1,113 subjects from 17 factories [11]. To the best of our knowledge, our study is the first in Korea to compare the effectiveness of multiple machine learning models for predicting the prognosis of ISSNHL.

A limitation of this study is its small sample size (227 patients), compared to previous studies that included 1,220 patients [1] and 1,113 patients [11]. Furthermore, the number of predictors may not have been sufficient compared with that of a former study that used 149 potential predictors [1]. Machine learning performance depends on the sample size and number of predictors. As the sample size and number of predictors increase, the model performance is expected to improve. Therefore, further research with a larger sample size is needed.

In conclusion, the accuracy of predicting hearing outcomes was 75.36% for SVM, 73.91% for RF, 72.46% for MLP, 72.46% for AdaBoost, and 65.22% for KNN, which are comparable to those of a previous study with a sample size of 1,220 [1]. Therefore, given a higher sample size and more predictors, we can expect better prediction performance with our model. Moreover, the SVM model with selected variables (accuracy, 75.36%) was found to be the most helpful for clinicians to predict the hearing prognosis of patients diagnosed with ISSNHL.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

ORCID

Keon Vin Park	https://orcid.org/0000-0002-4372-1321
Kyoung Ho Oh	https://orcid.org/0000-0003-2075-5905
Yong Jun Jeong	https://orcid.org/0000-0001-9664-9549
Jihye Rhee	https://orcid.org/0000-0001-8129-3973
Mun Soo Han	https://orcid.org/0000-0003-1359-6248
Sung Won Han	https://orcid.org/0000-0002-0040-3542
June Choi	https://orcid.org/0000-0002-6330-279X

AUTHOR CONTRIBUTIONS

Conceptualization: KVP, KHO, JR, SWH, JC. Data curation: KVP, KHO, YJJ, JR, MSH, JC. Formal analysis: KVP, KHO, JR, SWH, JC. Funding acquisition: JC, SWH. Methodology: KVP, SWH, JC. Visualization: KVP, YJJ, SWH, JC. Writing - original draft, review & editing: KVP, KHO, SWH, JC.

REFERENCES

1. Bing D, Ying J, Miao J, Lan L, Wang D, Zhao L, et al. Predicting the hearing outcome in sudden sensorineural hearing loss via machine learning models. Clin Otolaryngol. 2018 Jun;43(3):868-74.
2. Weiss D, Bocker AJ, Koopmann M, Savvas E, Borowski M, Rudack C. Predictors of hearing recovery in patients with severe sudden sensorineural hearing loss. J Otolaryngol Head Neck Surg. 2017 Apr; 46(1):27.
3. Stachler RJ, Chandrasekhar SS, Archer SM, Rosenfeld RM, Schwartz SR, Barrs DM, et al. Clinical practice guideline: sudden hearing loss. Otolaryngol Head Neck Surg. 2012 Mar;146(3 Suppl):S1-35.
4. Lin RJ, Krall R, Westerberg BD, Chadha NK, Chau JK. Systematic review and meta-analysis of the risk factors for sudden sensorineural hearing loss in adults. Laryngoscope. 2012 Mar;122(3):624-35.
5. Han X, Yin X, Du X, Sun C. Combined intratympanic and systemic use of steroids as a first-line treatment for sudden sensorineural hearing loss: a meta-analysis of randomized, controlled trials. Otol Neurotol. 2017 Apr;38(4):487-95.
6. Cao Z, Li Z, Xiang H, Huang S, Gao J, Zhan X, et al. Prognostic role of haematological indices in sudden sensorineural hearing loss: review and meta-analysis. Clin Chim Acta. 2018 Aug;483:104-11.
7. Suzuki H, Tabata T, Koizumi H, Hohchi N, Takeuchi S, Kitamura T, et al. Prediction of hearing outcomes by multiple regression analysis in patients with idiopathic sudden sensorineural hearing loss. Ann Otol Rhinol Laryngol. 2014 Dec;123(12):821-5.
8. Suzuki H, Mori T, Hashida K, Shibata M, Nguyen KH, Wakasugi T, et al. Prediction model for hearing outcome in patients with idiopathic sudden sensorineural hearing loss. Eur Arch Otorhinolaryngol. 2011 Apr;268(4):497-500.
9. Chao TK, Hsiu-Hsi Chen T. Predictive model for improvement of idiopathic sudden sensorineural hearing loss. Otol Neurotol. 2010 Apr; 31(3):385-93.
10. Cvorovic L, Deric D, Probst R, Hegemann S. Prognostic model for predicting hearing recovery in idiopathic sudden sensorineural hearing loss. Otol Neurotol. 2008 Jun;29(4):464-9.
11. Zhao Y, Li J, Zhang M, Lu Y, Xie H, Tian Y, et al. Machine learning models for the hearing impairment prediction in workers exposed to complex industrial noise: a pilot study. Ear Hear. 2019 May/Jun; 40(3):690-9.
12. Jun HJ, Chang J, Im GJ, Kwon SY, Jung H, Choi J. Analysis of frequency loss as a prognostic factor in idiopathic sensorineural hearing loss. Acta Otolaryngol. 2012 Jun;132(6):590-6.
13. Breiman L. Random forests. Mach Learn. 2001;45:5-32.
14. Elgamal M. Automatic skin cancer images classification. Int J Adv Comput Sci Appl. 2013;4(3):287-94.
15. Burges CJ. A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov. 1998;2:121-67.
16. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci. 1997; 55(1):119-39.