

Model Development for Predicting the Occurrence of Benign Laryngeal Lesions using Support Vector Machine: Focusing on South Korean Adults Living in Local Communities

Haewon Byeon

Department of Speech Language Pathology
Honam University
Gwangju, Republic of Korea

Abstract—The disease is a consequence of interactions between many complex risk factors, rather than a single cause. Therefore, it is necessary to develop a disease prediction model by using multiple risk factors instead of using a single risk factor. The objective of this study was to develop a model for predicting the occurrence of benign laryngeal lesions based on support vector machine (SVM) using ear, nose and throat (ENT) data from a national-level survey and to provide a basis for selecting high-risk groups and preventing a voice disorder. This study targeted 16,938 adults (≥ 19 years) who participated in the ENT examination among the people who completed the Korea National Health and Nutrition Examination Survey from 2010 to 2012. This study compared the prediction power of the Gauss function, which was used for this study, with that of a linear algorithm, that of a polynomial algorithm, and that of a sigmoid algorithm. Moreover, four kernels were divided into C-SVM and Nu-SVM to compare the prediction accuracy of C-SVM with that of Nu-SVM. The ‘benign laryngeal lesion prediction model’ based on SVM could derive preventive factors and risk factors. The final prediction rate of this SVM using 479 support vectors was 97.306. The fitness results indicated that the difference between C-SVM and Nu-SVM was not large in the benign laryngeal lesion prediction model. In terms of kernel type, the prediction accuracy of Gauss kernel was the highest and the prediction accuracy of the sigmoid kernel was the lowest. The results of this study will provide an important basis for preventing and managing benign laryngeal lesions.

Keywords—Support vector machine; SVM; dysphonia; voice disorder; prediction model; risk factor; data mining

I. INTRODUCTION

Benign laryngeal lesions have a different meaning from laryngeal cancer and it is used to describe voice disease due to the changes in the laryngeal structure such as vocal fold nodules, vocal polyp, intracordal cyst, Reinke’s edema, granuloma, sulcus vocalis, and larynx keratosis [1, 2]. Since benign laryngeal lesions lead to structural changes in the larynx including the vocal cords, it ultimately becomes a direct cause of voice disorder.

Roy et al. [3] reported that the prevalence of the laryngeal diseases was 6.6% in the United States as of 2005 and 10% of all Americans experienced a voice problem at least once in

their lifetime. Voice is a very important function in maintaining daily life, and maintaining a healthy voice can greatly affect the quality of life particularly for those who use voice for occupations (e.g., teacher). For example, the voice problem of the teachers can cause unnecessary socio-economic costs such as job loss or unemployment due to the loss of their ability to teach [4]. Therefore, it is very important to accurately understand the risk factors and diagnose the earlier for providing appropriate rehabilitation according to the assessment.

It has been reported that voice abuse and the physical stimulation to the vocal cord due to inappropriate speech habits are the most common causes of benign laryngeal lesions [5-13]. Moreover, various other factors such as smoking, drinking, virus, upper airway infection, and laryngopharyngeal reflux are also reported as risk factors [5-13]. The disease is a consequence of interactions between many complex risk factors, rather than a single cause [7, 14]. Therefore, it is necessary to develop a disease prediction model by using multiple risk factors instead of using a single risk factor [12]. Nevertheless, studies evaluating the risk factors for the laryngeal disease have mostly focused on single risk factors [15].

Lifestyle heavily influences the occurrence and rehabilitation of benign laryngeal lesions [15]. Moreover, even if surgical treatment is successfully conducted, it is likely to reoccur if the factor negatively affects the voice is not removed [15]. Additionally, although the shapes of lesions on the vocal fold are similar, different treatments must be performed depending on the etiology. Therefore, it is important to identify the complex risk factors in order to fully understand the cause of the disease and to make accurate diagnosis and treatment.

Machine learning based on the supervised learning such as support vector machine (SVM) has been used as a method to identify complex risk factors of diseases [16-18]. SVM is known to have the better predictive power in classifying binary data such as the presence of disease than the decision tree method or the artificial neural network (ANN) method [19-24]. Nevertheless, the voice disorder prediction models based on SVM, which mostly uses acoustic analysis indices, have been

mainly utilized to evaluate laryngeal diseases [25, 26]. There are not enough studies examining prediction models reflecting the health behaviors and sociodemographic characteristics.

The objective of this study was to develop a model for predicting the occurrence of benign laryngeal lesions based on SVM using ear, nose and throat (ENT) data from a national-level survey and to provide a basis for selecting high-risk groups and preventing a voice disorder. Construction of our study is as follows. chapter II explains study sample and analyzed variables and chapter III defines SVM and explains the procedure of model development. Lastly, chapter IV presents discussion and direction for future research.

II. METHODS

A. Data Source

This study targeted adults (≥ 19 years) who participated in the ENT examination among the people who completed the Korea National Health and Nutrition Examination Survey (KNHANES) from 2010 to 2012. This study was approved by the Research Ethics Committee of Honam University (IRB Number: 1041223-201801-HR-40). The KNHANES extract samples from sampling plots by using the proportional allocation systematic sampling method, which stratifies the administrative districts and dwelling types of the national regional classes. Therefore, the samples were proportional to the population. This study selected 16,938 adults (7,703 males, 9,235 females), who were 19 years or older and completed the health questionnaire, ENT examination questionnaire, and the laryngeal endoscopy.

B. Measurements

Benign laryngeal disease in this study were defined as vocal nodules, laryngeal polyps, intracordal cysts, reinke's edema, laryngeal granuloma, glottic sulcus and laryngeal keratosis [10]. The explanatory variables were age(19-39, 40-59, 60+), gender, occupation, educational level, Income, smoking, high-risk drinking, and self-reported voice problems. Occupations were classified into economically-inactive, non-manual and manual. Levels of education were classified as elementary school graduates and lower, junior high school graduates, high school graduates and college graduates and over. Levels of income for households were classified into four quartiles.

III. ANALYSIS METHODS

A. Development of Prediction Model using SVM

The difference between the groups by the prevalence of benign laryngeal lesions was tested by chi-square test. The prediction model for benign laryngeal lesions was developed using SVM. SVM is a machine learning algorithm that finds the most optimal decision boundary by linearly separating the hyperplane after converting the learning data to a higher dimension through nonlinear mapping [27]. The concept of SVM is shown in Figure 1.

For example, $A=[a, d]$ and $B=[b, c]$ are non-linearly separable in the two-dimensional space. If they are mapped in the three-dimensional space, they have a linearly separable characteristic. Thus, data containing two classes can always be

separated in hyperplanes when the appropriate nonlinear mapping is used with sufficiently large dimensions [28]. Consequently, SVM is very accurate because it can model the complex nonlinear decision-making areas and it tends to have less over-fitting than other models, which is an advantage [29].

The objective of SVM is to find the interface that maximizes the margin (Fig. 2). For the convenience of calculation, the algorithm will be explained by minimizing the half of the reciprocal of squared half margin.

$$\max_{2\|w\|^2} \rightarrow \min_{1/2\|w\|^2} \max_{2\|w\|^2} \rightarrow \min_{1/2\|w\|^2}$$

The meaning of the above equation is as follows. The observations above the plus-plane satisfy $y=1$ and $wTx+b$ is larger than 1. On the other hand, the observations below the minus-plane are $y=-1$ and $wTx+b$ is smaller than -1. When these two conditions are combined, it can be converted to the following constraint.

$$y_i(wTx_i+b) \geq 1$$

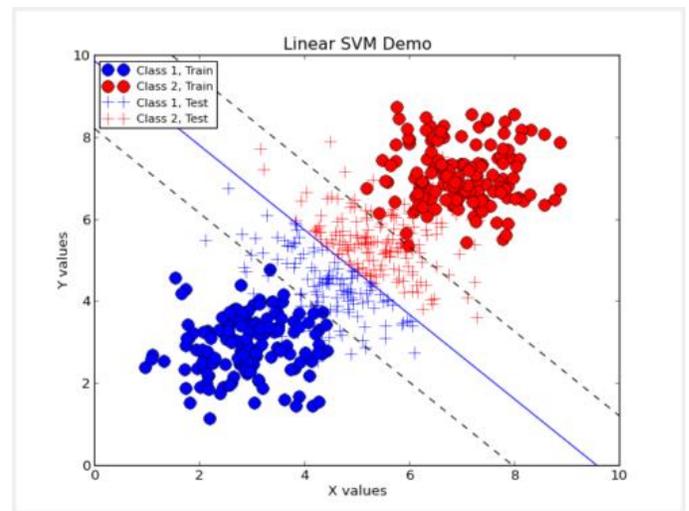


Fig. 1. Concepts of SVM, Source: MTECH Project [30].

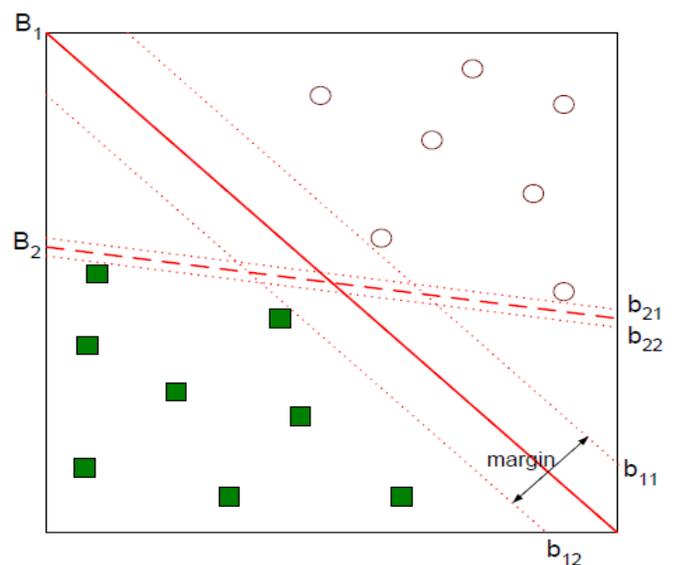


Fig. 2. Maximum Margin Classifiers.

```
> model <- svm(Species~., data=LD)
> model
Call:
svm(formula = Species ~ ., data = LD)
Parameters:
SVM-Type: C-classification
SVM-Kernel: radial
cost: 1
gamma: 0.25
Number of Support Vectors:
> table(pred, LD$Species)
predict(object, newdata, decision.values = FALSE,
probability = FALSE, ..., na.action = na.omit)
#object
Object of class "svm", created by svm.
#newdata
An object containing the new input data
# p <- predict(m, test, type = "response")
>plot(cmdscale(dist(LD[,5])), col = as.integer(LD[,5]),
pch = c("o", "+")[1:150 %in% model$index + 1])
library(kernlab)
LDmodel <- ksvm(Species ~ ., data = LD,
type = "C-bsvc", kernel = "rbfdot",
kpar = list(sigma = 0.1), C = 10,
prob.model = TRUE)
svm_problem prob = new svm_problem();
ArrayList pointsList = iv.getPointsList();
svm_node[][] nodes = new svm_node[2][pointsList.size()];
for (int i = 0; i < pointsList.size(); i++) {
float point = pointsList.get(i);
nodes[0][i] = new svm_node();
nodes[0][i].index = i;
nodes[0][i].value = point;
}
for (int i = 0; i < pointsList.size(); i++) {
float point = pointsList.get(i);
nodes[1][i] = new svm_node();
```

```
nodes[1][i].index = i;
nodes[1][i].value = point+5;
}
prob.l = 2;
prob.y = new double[] { 1, 2 };
prob.x = nodes;
svm_parameter param = new svm_parameter();
libsvm.svm.svm_check_parameter(prob, param);
svm_model model = libsvm.svm.svm_train(prob, param);
System.out.println(libsvm.svm.svm_predict(model, nodes[1]));
```

Fig. 3. SVM Algorithm Source of R Program.

When new data is entered, the observations are substituted into $y_i(wTx_i+b-1)y_i(wTx_i+b-1)$. If the answer is larger than 0, it is predicted as 1 category. If it is smaller than 0, it is predicted as -1 category. The analysis is performed by using R version 3.4.2. Figure 3 presents the SVM algorithm source of R program. This study chose a radial basis function (Gauss function) that uses the parameter C (unit cost) in the SVM algorithm. This study compared the prediction power of the Gauss function, which was used for this study, with that of a linear algorithm, that of a polynomial algorithm, and that of a sigmoid algorithm. Moreover, four kernels were divided into C-SVM and Nu-SVM to compare the prediction accuracy of C-SVM with that of Nu-SVM.

IV. RESULTS

A. The General Characteristics of Subjects

The general characteristics of 16,938 subjects were analyzed by using frequency analysis (Table 1). The ages of the subjects were distributed as 31.4% for 19-39 years old, 36.9% for 40-59 years old, and 31.7% for the 60 years and older. The 54.5% of subjects were women and 45.5% of them were men. In terms of occupation, the not economically active population was 44.1%, non-manual workers were 31.9%, and manual workers were 23.9%. The education level of subjects was elementary school graduate and below (39.2%), middle school graduate (11.5%), high school graduate (26.2%), and college graduate and above (23.1%). The 68.7% of subjects were nonsmokers, the 15.5% of them were former smokers, and 15.8% of them were current smokers. The 32.4% of subjects experienced the high-risk drinking during the past one year. The prevalence of subjective voice disorder was 7% and the prevalence of benign laryngeal lesions was 1.8%.

B. The Potential Causes of Benign Laryngeal Lesions

The general characteristics of subjects by the prevalence of benign laryngeal lesions and the potential causes of benign laryngeal lesions are presented in Table 2. The results of chi-square test showed that the subjects with benign laryngeal lesions and those without benign laryngeal lesions were not significantly different in all variables (i.e., age, gender, income level, occupation, education level, smoking, and high-risk

drinking) except voice disorder. The prevalence of benign laryngeal lesions was significantly higher in subjects with subjective voice disorder (9.8%) than in those without subjective voice disorder (2.1%, $p < 0.05$).

C. Major Predictors based on SVM

Table 3 shows the 'function weight' of SVM based on Gaussian Kernel algorithm. It is impossible to simply compare the magnitude (priority) of influence on the function weight of SVM by a variable. However, it is possible to identify if a major predictor variable has a positive relationship with a disease or a negative relationship with a disease. The 'benign laryngeal lesion prediction model' based on SVM could derive preventive factors and risk factors.

Risk factors were people perceiving voice problem subjectively, age between 40 and 59, 60 years old or older, female, occupation (non-manual worker and manual worker), high school graduate (the highest level of education), and high-risk drinking.

TABLE I. GENERAL CHARACTERISTICS OF SUBJECTS, %

Variables	Subcategory	Total (n=16,938)
Age	19-39	31.4
	40-59	36.9
	60+	31.7
Gender	Male	45.5
	Female	54.5
Income (quartile)	First quartile	17.3
	Second quartile	26.9
	Third quartile	28.8
	Fourth quartile	27.0
Job	Economically inactive population	44.1
	Non-manual worker	31.9
	Manual worker	23.9
highest level of education	Elementary school graduation	39.2
	Middle school graduate	11.5
	High school graduation	26.2
	College graduate	23.1
Smoking	Non smoker	68.7
	Past smoker	15.5
	Current smoker	15.8
High-risk drinking	No	67.6
	Yes	32.4
Subjective voice disorder	No	93.7
	Yes	7.0
Benign laryngeal lesions	No	98.2
	Yes	1.8

Preventive factors were people not-perceiving subjective voice problem, the income level second quartile (medium-low), third quartile (medium-high), and fourth quartile (high) than the income level first quartile (low), middle school graduate and college graduate (the highest level of education), and former smokers and nonsmokers than smokers. The final prediction rate of this SVM using 479 support vectors was 97.306.

TABLE II. THE GENERAL CHARACTERISTICS OF SUBJECTS BY THE PREVALENCE OF BENIGN LARYNGEAL LESIONS AND THE POTENTIAL CAUSES OF BENIGN LARYNGEAL LESIONS, N (%)

Variables	Subcategory	Benign laryngeal lesions		P
		No (n=16,636)	Yes (n=302)	
Age	19-39	3,974 (98.1)	79 (1.9)	0.564
	40-59	4,688 (98.2)	84 (1.8)	
	60+	4,023 (98.4)	67 (1.6)	
Gender	Male	7,564 (98.2)	139 (1.8)	0.862
	Female	9,072 (98.2)	163 (1.8)	
Income (quartile)	First quartile	2,840 (98.4)	47 (1.6)	0.605
	Second quartile	4,421 (98.4)	73 (1.6)	
	Third quartile	4,712 (98.2)	88 (1.8)	
	Fourth quartile	4,421 (98.0)	88 (2.0)	
Job	Economically inactive population	5,426 (98.4)	87 (1.6)	0.178
	Non-manual worker	3,907 (97.9)	83 (2.1)	
	Manual worker	2,939 (98.3)	51 (1.7)	
Highest level of education	Elementary school graduation	6,022 (98.2)	108 (1.8)	0.142
	Middle school graduate	1,768 (98.8)	22 (1.2)	
	High school graduation	4,029 (98.3)	70 (1.7)	
	College graduate	3,535 (97.9)	76 (2.1)	
Smoking	Nonsmoker	10,566 (98.4)	176 (1.6)	0.158
	Past smoker	2,372 (98.0)	49 (2.0)	
	Current smoker	2,423 (97.9)	52 (2.1)	
High-risk drinking	No	6,752 (98.2)	125 (1.8)	0.985
	Yes	3,231 (98.2)	60 (1.8)	
Subjective voice disorder	No	10,357 (97.9)	224 (2.1)	<0.001
	Yes	721 (90.2)	78 (9.8)	

D. Major Predictors based on SVM

Table 3 shows the 'function weight' of SVM based on Gaussian Kernel algorithm. It is impossible to simply compare the magnitude (priority) of influence on the function weight of SVM by a variable. However, it is possible to identify if a major predictor variable has a positive relationship with a disease or a negative relationship with a disease. The 'benign laryngeal lesion prediction model' based on SVM could derive preventive factors and risk factors. Risk factors were people perceiving voice problem subjectively, age between 40 and 59, 60 years old or older, female, occupation (non-manual worker and manual worker), high school graduate (the highest level of education), and high-risk drinking. Preventive factors were people not-perceiving subjective voice problem, the income level second quartile (medium-low), third quartile (medium-high), and fourth quartile (high) than the income level first quartile (low), middle school graduate and college graduate (the highest level of education), and former smokers and nonsmokers than smokers. The final prediction rate of this SVM using 479 support vectors was 97.306.

TABLE III. FUNCTION WEIGHT OF SVM BASED ON GAUSSIAN KERNEL ALGORITHM

Perceiving voice problem subjectively	.007
Not-perceiving subjective voice problem	-.007
Age between 40 and 59	.029
60 years old or older	.034
Female	.019
Income level second quartile (medium-low)	-.017
Income level third quartile (medium-high)	-.023
Income level fourth quartile (high)	-.045
Non-manual worker	.008
Manual worker	.005
Middle school graduate	-.003
High school graduate	.011
College graduate	-.011
Former smokers	-.029
Nonsmokers	-.001
High-risk drinking	.004
Number of Support Vector: 679	

E. The Accuracy of Benign Laryngeal Lesion Prediction According to the SVM Classification Algorithm

The accuracy of benign laryngeal lesion prediction according to the SVM classification algorithm is shown in Table 2. One of the SVM's disadvantages is that the fitness of the model varies by the type of kernel. Therefore, this study compared the Gaussian kernel (used for this study), linear, polynomial, and sigmoid algorithms in order to confirm the prediction accuracy of the models according to various kernel

types. Additionally, these four algorithms were divided into C-SVM and Nu-SVM and the prediction accuracy of each algorithm was compared with each other. The fitness results indicated that the difference between C-SVM and Nu-SVM was not large in the benign laryngeal lesion prediction model. In terms of kernel type, the prediction accuracy of Gauss kernel was the highest and the prediction accuracy of the sigmoid kernel was the lowest.

TABLE IV. THE ACCURACY OF BENIGN LARYNGEAL LESION PREDICTION ACCORDING TO THE SVM CLASSIFICATION ALGORITHM, %

Type of SVM	Type of kernel			
	Linear	Polynomial	Gaussian	Sigmoid
C-SVM	96.582	95.825	97.306	94.858
Nu-SVM	96.431	95.767	97.120	93.158

V. DISCUSSION

This study developed a benign laryngeal lesion prediction model based on SVM by using the highly reliable ENT examination data. The results of this study revealed that subjective voice problem perception, 40 years and older, female, occupation, high school graduate, and high-risk drinking were risk factors of benign laryngeal lesions. The preventive factors of benign laryngeal lesions were people not perceiving subjective voice problem, above the second quartile, middle school graduate and college graduate (highest level of education), former smokers, and nonsmokers. Numerous previous studies that evaluated the risk factors of voice disorder reported that sociodemographic characteristics such as gender [13], age [6], and occupation [3,4,31,32], drinking [2], and smoking [31,15] were the risk factors of voice disorder. These results support the results of this study. One interesting finding of this study was that middle school graduate and college graduate were the preventive factors of benign laryngeal lesions. It was speculated that the middle school graduate had a low risk of benign laryngeal lesions because they were mostly manual workers engaging in simple labor. Byeon & Lee (2010) [33] analyzed the prevalence of benign laryngeal lesions, similar to this study, and reported that simple manual workers (e.g., maids, cleaners, and construction workers) had a low voice disorder prevalence. Byeon & Lee (2010) [33] argued that it was because they had a low probability to abuse their voice because they worked along and did repetitive jobs. It will be necessary to conduct a longitudinal study that will demonstrate the causality among the highest level of education, occupation, and voice disorder.

Another finding of this study was that the prediction accuracy of C-SVM's Gaussian kernel was the highest when the prediction accuracies of the eight SVM classification algorithms were compared after dividing Gaussian kernel, linear kernel, polynomial kernel, and sigmoid algorithms into C-SVM and Nu-SVM. The performance of the nonlinear SVM is largely determined by a kernel function and the parameters consisting of the kernel [28]. Kuo et al. [34] also reported that the Gaussian kernel algorithm had high prediction accuracy. The Gaussian kernel is to map in the specific space with

infinite dimensions. It is believed that using Gaussian kernel-based C-SVM will be effective in predicting binomial variables.

VI. CONCLUSION

The results of this study will provide an important basis for preventing and managing benign laryngeal lesions. It will be required to systematically manage the high-risk groups in order to prevent benign laryngeal lesions.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07041091).

REFERENCES

- [1] H. D. Soni, S. Gandhi, M. Goyal, and U. Shah, Study of clinical profile of benign laryngeal lesions. *International Journal of Medical Science and Public Health*, vol. 5, no. 4, pp. 656-660, 2016.
- [2] H. Byeon, A population-based cross-sectional study of alcohol consumption and risk of benign laryngeal disease in Korean adults. *Journal of Voice*, vol. 30, no. 4, pp. 443-447, 2016.
- [3] N. Roy, R. M. Merrill, S. D. Gray, and E. M. Smith, Voice disorders in the general population: prevalence, risk factors, and occupational impact. *The Laryngoscope*, vol. 115, no. 11, pp. 1988-1995, 2005.
- [4] A. A. Assunção, I. B. Bassi, A. M. de Medeiros, C. de Souza Rodrigues, and A. C. C. Gama, Occupational and individual risk factors for dysphonia in teachers. *Occupational Medicine*, vol. 62, no. 7, pp. 553-559, 2012.
- [5] H. Byeon, The factors that affects the experience of discrimination in children in multi-cultural families using QUEST algorithm: focusing on Korean language education. *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, no. 4, vol. 2, pp. 303-312, 2014.
- [6] H. Byeon, The prediction model for self-reported voice problem using a decision tree model. *Journal of the Korea Academia-Industrial cooperation Society*, vol. 14, no. 7, pp. 3368-3373, 2013.
- [7] P. N. Carding, S. Roulstone, K. Northstone, and ALSPAC Study Team. (2006). The prevalence of childhood dysphonia: a cross-sectional study. *Journal of Voice*, 20(4), 623-630, 2006.
- [8] N. R. Williams, Occupational groups at risk of voice disorders: a review of the literature. *Occupational Medicine*, vol. 53, no. 7, pp. 456-460, 2003.
- [9] M Behlau, F. Zambon, A. C. Guerrieri, and N. Roy, Epidemiology of voice disorders in teachers and nonteachers in Brazil: prevalence and adverse effects. *Journal of Voice*, vol. 26, no. 5, pp. 665.e9-18, 2012.
- [10] R. M. B. De Alvear, F. J. Barón, and A. G. Martínez-Arquero, School teachers' vocal use, risk factors, and voice disorder prevalence: guidelines to detect teachers with current voice problems. *Folia Phoniatrica et Logopaedica*, vol. 63, no. 4, pp. 209-215, 2011.
- [11] H. Byeon, Relationships among smoking, organic, and functional voice disorders in Korean general population. *Journal of Voice*, vol. 29, No. 3, pp. 312-316, 2015.
- [12] H. Byeon, and Y. Lee, Laryngeal pathologies in older Korean adults and their association with smoking and alcohol consumption. *The Laryngoscope*, vol. 123, no. 2, pp. 429-433, 2013.
- [13] H. Byeon, Gender differences in risk factors of benign vocal fold disease in Korea: the fifth Korea National Health and Nutritional Examination Survey. *Logopedics Phoniatrics Vocology*, vol. 41, no. 2, pp. 85-91, 2016.
- [14] V. Kontis, C. D. Mathers, J. Rehm, G. A. Stevens, K. D. Shield, P. R. Bonita, L. M. Riley, V. Poznyak, and M. Ezzati, Contribution of six risk factors to achieving the 25x 25 non-communicable disease mortality reduction target: a modelling study. *The Lancet*, vol. 384, no. 9941, pp. 427-437, 2014.
- [15] H. Byeon, The risk factors of laryngeal pathology in Korean adults using a decision tree model. *Journal of Voice*, vol. 29, no. 1, pp. 59-64, 2015.
- [16] B. Zheng, S. W. Yoon, and S. S. Lam, Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Systems with Applications*, vol. 41, no. 4, pp. 1476-1482, 2014.
- [17] H. Byeon, A prediction model for mild cognitive impairment using random forests. *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 12, pp. 8-12, 2015.
- [18] H. Byeon, H. Jin and S. Cho, Development of Parkinson's disease dementia prediction model based on verbal memory, visuospatial memory, and executive function. *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 7, pp. 1517-1521, 2017.
- [19] A. Begue, V. Kowlessur, U. Singh, F. Mahomoodally, and S. Pudaruth, Automatic Recognition of medicinal plants using machine learning techniques. *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 4, pp. 166-175, 2017.
- [20] A. Amri, I. A'Fifah, A. R. Ismail, and A. A. Zarir, Comparative performance of deep learning and machine learning algorithms on imbalanced handwritten data. *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 2, pp. 258-264, 2018.
- [21] H. Byeon. Developing a model to predict the occurrence of the cardio-cerebrovascular disease for the Korean elderly using the random forests algorithm. *International Journal of Advanced Computer Science and Applications*, no. 9, vol. 9.
- [22] H. Byeon, H. Development of prediction model for endocrine disorders in the Korean elderly using CART algorithm. *Development*, vol. 6, no. 9, 2015.
- [23] H. Byeon, and H. W. Koh, A feature analysis of risk factors for stroke in the middle-aged adults. *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 12, pp. 95-99, 2015.
- [24] H. Byeon, Chi-square automatic interaction detection modeling for predicting depression in multicultural female students. *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 12, pp. 179-183, 2017.
- [25] P. Saidi, and F. Almasganj, Voice disorder signal classification using m-band wavelets and support vector machine. *Circuits, Systems, and Signal Processing*, vol. 34, no. 8, pp. 2727-2738, 2015.
- [26] M. Shahbakhi, D. T. Far, and E. Tahami, Speech analysis for diagnosis of Parkinson's disease using genetic algorithm and support vector machine. *Journal of Biomedical Science and Engineering*, vol. 7, no. 4, pp. 147-156, 2014.
- [27] G. Raho, R. Al-Shalabi, G. Kanaan, and A. Nassar, A Different classification algorithms based on Arabic text classification: feature selection comparative study. *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 2, pp. 23-28, 2015.
- [28] I. Steinwart, and A. Christmann, Support vector machines, Springer Science & Business Media, New York, 2008.
- [29] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, vol. 16, no. 10, pp. 906-914, 2000.
- [30] MTECH Project, <http://www.mtechprojects.org/linear-svm-projects.html>
- [31] H. Byeon, The association between lifetime cigarette smoking and dysphonia in the Korean general population: findings from a national survey. *Peer J*, vol. 3, pp. e912, 2015.
- [32] H. Byeon, H. Occupational risks for voice disorders: Evidence from a Korea national cross-sectional survey. *Logopedics Phoniatrics Vocology*, vol. 42, no. 1, pp. 39-43, 2017.
- [33] H Byeon, and Y. Lee, Prevalence and risk factors of benign laryngeal lesions in the adult population. *Communication Sciences & Disorders*, vol. 15, no. 4, pp. 648-656, 2010.
- [34] B. C. Kuo, H. H. Ho, C. H., Li, C. C. Hung, and J. S. Taur, A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 1, pp. 317-326, 2014.