# Developing a Model for Predicting the Speech Intelligibility of South Korean Children with Cochlear Implantation using a Random Forest Algorithm

Haewon Byeon

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

*Abstract*—**The random forest technique, a tree-based study model, predicts the results by using random decision trees based on the bootstrap technique. Therefore, it has a high prediction power and fewer errors, which are advantages of this method. This study aimed to provide baseline data regarding the language therapy after cochlear implantation by identifying the factors associated with the speech intelligibility of children with cochlear implantation. This study evaluates the factors associated with the articulation accuracy of children with cochlear implantation. This study targeted 82 hearing-impaired children, who lived in Seoul, Incheon, and Suwon areas, were between 4 and 8 years old, and had been worn cochlear implant at least one year and less than five years. Explanatory variables used in this study included gender, age, household income, the wear time of a cochlear implant, vocabulary index, and corrected hearing. Speech intelligibility was analyzed using the 'speech intelligibility test tool' composed of nine sentences. The predictive model for speech intelligibility of children with cochlear implants was developed using random forest. The major predictors of the articulation accuracy of children with cochlear implantation were the wear time of a cochlear implant, the time since cochlear implantation, vocabulary, household income, age, and gender, in the order of the magnitude. The final error rate of the random forest model developed by generating 500 bootstrap samples was 0.22, and the prediction rate was 78.8%. The results of this study on a prediction model suggested that it would be necessary to implement cochlear implantation and to develop a customized aural rehabilitation program considering the linguistic ability of a subject for enhancing the speech intelligibility of a child with cochlear implantation.**

*Keywords—Random forest; hearing impairment; vocabulary index; speech intelligibility; risk factor; data mining*

## I. INTRODUCTION

The number of patients with hearing loss in South Korea is continuously increasing. Health Insurance Review & Assessment Service (2017)[1] reported that the number of patients who visited a doctor's office due to hearing loss increased by 18.7% in five years (from 746,499 in 2012 to 886,091 in 2016). Considering that a large number of people with hearing problems do not visit a doctor's office, the number of potential patients with hearing loss is expected to be higher than the estimate. National Survey of Ministry of Health and Welfare (2016)[2] reported that one out of two people with hearing loss did not visit a doctor until it prohibits them from

living normal lives and only 12.6% of people who were diagnosed with hearing loss wore a hearing aid. It is expected that the number of patients with hearing loss will increase steadily because the noise environment is expanding to the advancing industrialization. The more population with hearing loss is a problem affecting various domains of a community and a nation, such as the loss of workforce and the support of special education.

As the effectiveness of cochlear implantation has been proven, a wider range of patients has become a target of this surgery. Since the Food and Drug Administration (FDA) approved cochlear implantation in 1990, the number of cochlear implant recipients has been increasing worldwide [3]. The cochlear implantation has been covered by the national insurance system in Korea since 2005, so patients are only responsible for 20% of testing and surgical cost. Consequently, the size of the market is growing rapidly.

Many studies have proven the effectiveness of cochlear implant. The cochlear implant is expected to increase the speech perceptivity through biaural gain, noise signal discrimination under a noise environment, sound localization identification, head shadow effect, biaural squelch, and biaural summation [4-7]. The cochlear implantation has been introduced worldwide more and more as more advantages (e.g., the increase of speech discrimination in noise, improvement of sound quality, improvement of signal-to-noise ration, enhancement of equilibrium sensation, suppression of tinnitus, and cost-effectiveness of cochlear implants) of it are known [8,9]. Generally, the target patients of cochlear implantation are when patients are from 12 months to 17 years old, sensorineural hearing loss for both ears above 90dBHL and, when patients are equal to or older than 18 years, hearings loss for both years above 70dB, who do not show advancement in hearing ability and linguistic ability even after wearing a hearing aid long than 3 months [9].

Numerous studies have already reported that hearing-impaired children with cochlear implantation had better voice perceptivity than hearing-impaired children with hearing aids [10,11]. These studies suggested that children with cochlear implantation had better speech perceptivity than those with hearing aids because the cochlear implant is connected directly to the cochlear hair cell in the labyrinth to compensate the

severe hearing loss effectively, unlike the hearing aid that only amplifies sound [12].

Numerous studies have already reported that hearing-impaired children with cochlear implantation had better voice perceptivity than hearing-impaired children with hearing aids [10,11]. These studies suggested that children with cochlear implantation had better speech perceptivity than those with hearing aids because the cochlear implant is connected directly to the cochlear hair cell in the labyrinth to compensate the severe hearing loss effectively, unlike the hearing aid that only amplifies sound [12].

The speech intelligibility of hearing-impairment children can be effectively improved if auditory compensation is made in the early stage [13]. However, many children with hearing impairment receive distorted auditory feedback due to the acoustic limitations of hearing aids, which cannot amplify sounds above 4000 Hz, resulting in the production of misarticulation [14]. On the other hand, the cochlear implant can deliver a wide frequency band from 100 Hz to 8000 Hz without distortion, unlike the hearing aid, so it is more beneficial in recognizing consonants belonging to the high-frequency band.

It has been reported that the onset of hearing loss, the level of language comprehension, the wear time of a cochlear implant, the time of cochlear implant operation, and the number of cochlear implant electrodes affect the voice perceptivity of children with cochlear implantation [7,9,13,15]. It has been also reported that the time of cochlear implantation, speech perceptivity, speech production ability, hearing ability, and language ability are related to the speech intelligibility [16]. Although it is expected that the factors influencing voice perceptivity are also associated with speech intelligibility, there are not enough studies evaluating the factors related to the speech intelligibility of children with cochlear implantation in South Korea. Furthermore, the previous studies evaluating the predictive variables of the speech intelligibility of children with cochlear implantation were conducted using linear regression models [17-19]. The linear regression model is a statistical technique that estimates the effects of independent variables on a dependent variable. Since this technique assumes the linear relationship between independent variables and a dependent variable, it does not have a good predictive power when the data have non-linear attributes.

Recently, the data mining technology has been advanced greatly and it is very useful to extract information from big data. Owing to the advancement, this technique has been used in various fields such as marketing, education, and healthcare [20, 21, 22, 23]. Among them, the random forest technique, a tree-based study model, predicts the results by using random decision trees based on the bootstrap technique. Therefore, it has a high prediction power and fewer errors, which are advantages of this method. This study aimed to provide baseline data regarding the language therapy after cochlear implantation by identifying the factors associated with the speech intelligibility of children with cochlear implantation.

## II. MATERIALS AND METHODS

### A. Subjects

It is a descriptive study evaluating the factors associated with the articulation accuracy of children with cochlear implantation. This study targeted 82 hearing-impaired children, who lived in Seoul, Incheon, and Suwon areas, were between 4 and 8 years old, and had been worn cochlear implant at least one year and less than five years. The subject selection criteria were as follows. The subjects of this study were hearing-impaired children (1) who had been wearing cochlear implant at least one year, (2) who had received aural rehabilitation regularly after receiving the surgery, and (3) who used oral speech in conversation. Children accompanying other disorders such as visual impairment, cognitive impairment, emotional impairment, and intellectual impairment were excluded from the study. The appropriate sample size was estimated as 80 people by using G-Power 3.1 program when there are seven predictive variables at alpha=0.05, power (1-B) = 0.8, and effect size (f2) of 0.2. Therefore, the sample size of this study met the appropriate sample size for conducting statistical tests (Figure 1) (Figure 2).
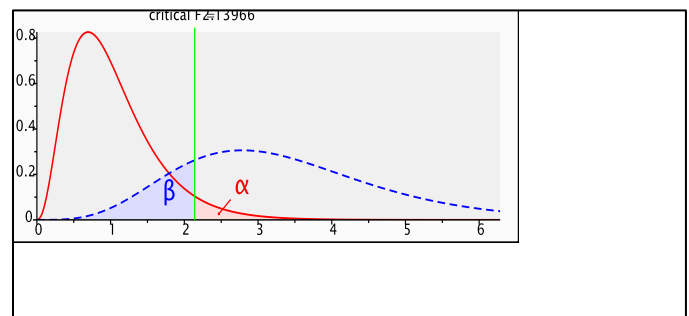


Fig. 1. Power and sample size calculations

### B. Explanatory Variable

Explanatory variables used in this study included gender, age, household income, the wear time of a cochlear implant, vocabulary index, and corrected hearing. The corrected hearing was defined as the mean threshold decibel (dB) of hearing tests measured at 250 Hz, 500 Hz, 1 kHz, 2 kHz, and 4 kHz ranges after wearing a cochlear implant. When both ears had cochlear implants, a mean threshold value was used. On the other hand, when only one ear wore a cochlear implant and the other ear used a hearing aid, only the hearing of the cochlear implant side was used.

### C. Speech Intelligibility

Speech intelligibility was analyzed using the 'speech intelligibility test tool' composed of nine sentences and developed by Yoon et al. (2005) [24]. This tool includes eight consonant oriented sentences (i.e., a nasal sound, a liquid consonant, a bilabial plosive, an alveolar plosive, a velar plosive, an alveolar fricative, a glottal fricative, a palatal affricate) and one vowel oriented sentence and all sentences are composed of two phrases. The test items are shown in Table 1.

The number of collected words and the number of correctly pronounced words were scored as a percentage (%). All test words produced by subjects were recorded with ESONIC MQ-

94 (4GB) voice recorder. In order to evaluate the consistency between examiners, two graduate students majoring in speech pathology evaluated the speech intelligibility after listening to all recordings of the speech intelligibility tests. The consistency between the two examiners was 91%. The estimated speech intelligibility was transformed into quartiles: the 1st, 2nd, and 3rd quartiles were defined as low-rank groups of speech intelligibility and the 4th quartile was defined as a high-rank group.

**F tests** – Linear multiple regression: Fixed model, R² deviation

from zero

**Analysis:** A priori: Compute required sample size

| **Input:** | Effect size f² | = | 0.2 |
| | α err prob | = | 0.05 |
| | Power (1−β err prob) | = | 0.8 |
| | Number of predictors | = | 7 |
| **Output:** | Noncentrality parameter ë | = | 16.0000000 |
| | Critical F = | 2.1396555 | |
| | Numerator df | = | 7 |
| | Denominator df | = | 72 |
| | Total sample size = | 80 | |
| | Actual power | = | 0.8061255 |

Fig. 2.   Estimation of Appropriate Sample Size

### D. Vocabulary Index

Receptive vocabulary and expressive vocabulary were measured using the receptive and expressive vocabulary test (REVT)[25]. The age equivalent-month of receptive vocabulary capability and expressive vocabulary capability was calculated from the test results and then it was converted to the vocabulary index. Equation (1) is the vocabulary index formula.

Vocabulary Index = Receptive Vocabulary Age Equivalent-Month + Expressive Vocabulary Age Equivalent-Month / 2 (1)

TABLE I.        THE TEST ITEMS OF SPEECH INTELLIGIBILITY TEST TOOL

| No | Target phoneme | Test sentence |
|---|---|---|
| 1 | Nasal | noran ɲyaŋmarine. |
| 2 | Liquids | norɛrɯl pullʌyo. |
| 3 | Bilabial plosive | pʰirirɯl punda. |
| 4 | Alveolar plosive | tʰadzoga taʎʎinda. |
| 5 | Velar plosive | kʰokìrinɯn kʰɯda. |
| 6 | Palatal affricate | tsadoŋtsʰa undzʌnhɛyo. |
| 7 | Alveolar fricative | sʰasʰɯmi sʰumʌyo. |
| 8 | Glottal fricative | horaŋiga ʌhɯŋhɛyo. |
| 9 | Vowel | agiga urʌyo. |

## III.   ANALYSIS METHODS

### A. Bagging Algorithm

When the variability of the classifier is large while the data is changed even a little, a classifier can be acquired by using a bootstrap method to reduce the variability of a predictor. This method is called a bagging (Bootstrap Aggregating), which is an ensemble algorithm using the bootstrap method [26].
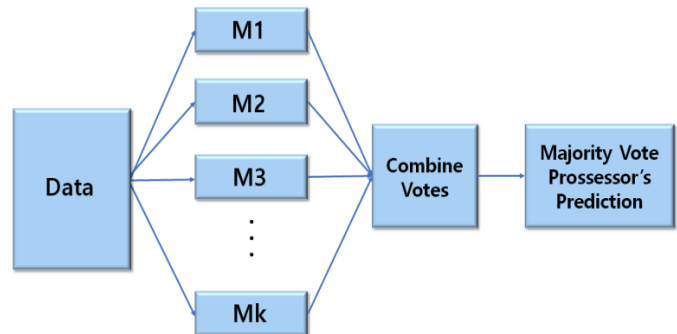


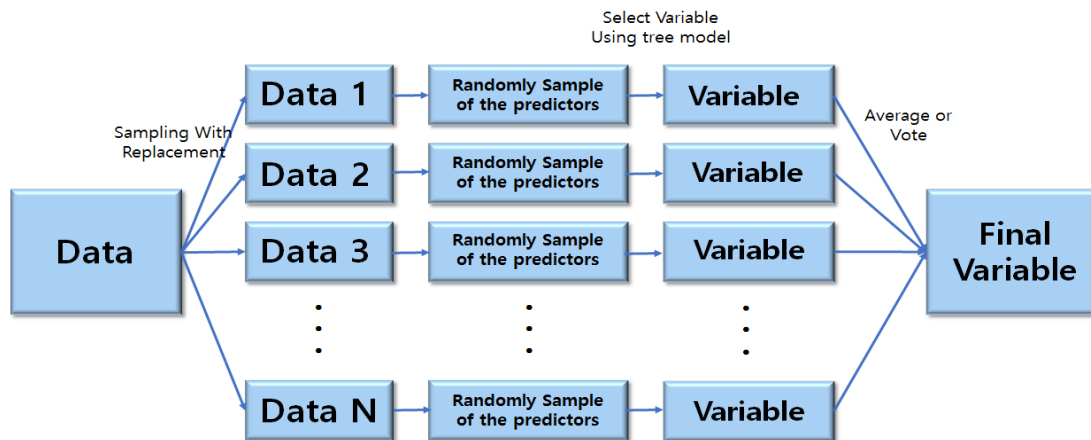Fig. 3.   Bootstrap Aggregating algorithm



Fig. 4.   Random Forest algorithm

The bagging algorithm process generates the bootstrap dataset by sampling with replacement from the training dataset, which is extracted from the population (Figure 3). N bootstrap datasets are generated by repeating this step N times and N classifier groups were obtained after forming each single decision tree by applying an appropriate classification algorithm (e.g., decision tree) to each bootstrap dataset [26]. There are several ways to combine these single classifiers: the mean is used when the response variable is continuous, while a vote is used when it is categorical [27-29].

If a training dataset is unstable, the classification performance is improved by combining bagging classifiers [30, 31]. However, if a training dataset is stable, the bagging-based decision tree obtained from the bagging process is similar to the single decision tree obtained from the training dataset [32, 33]. The bagging is effective in increasing the classification accuracy and improving the stability. Moreover, it has the advantages of reducing the variance and preventing the over fitting [34].

### B. Random Forest Model

The core of the bagging method is to predict the classification algorithm results of the bootstrap dataset by averaging or voting the bootstrap dataset. The random forest is an algorithm that adds a random additional layer to this bagging algorithm [32].

Unlike the decision tree expressing each node as a partition providing the most optimal results by using all variables, the random forest uses a method providing the most optimal results among the selected explanatory variable groups by randomly selecting explanatory variables for expressing each node (Fig. 4).

The random forest algorithm process generates bootstrap datasets by sampling with replacement from the training dataset extracted from the population. N bootstrap datasets were generated by repeating this process N times. Moreover, when applying the decision tree algorithm, m explanatory variables are randomly selected from each node. By using these selected variables, this study obtained a group composed of N single classifiers after identifying the most optimal split combination. If a response variable is continuous, a single classifier is combined using a mean. If a response variable is categorical, a single classifier is combined using a vote, just the same as the bagging algorithm.

One of the advantages of the random forest method is that it has less variance than the bagging algorithm because it reduces the correlation between trees. Moreover, it provides more accurate results than other algorithms and is useful to identify an important variable from a large data because it can utilize thousands of independent variables without removing a variable. Especially, it is known that it provides a similar or better predictive power than bagging or boosting when there are many input variables [35]. The analysis is performed by using R version 3.4.3. Figure 5 presents the Random Forest algorithm source of R program.

## IV. RESULTS

### A. General Characteristics of Subjects

The general characteristics of the whole subjects showed that the mean age was 6.3 years (standard deviation (SD)=3.1), and 32% of them were females (n=26) and 68% of them were males (n=56). The mean vocabulary index was 41.2 (SD=23.6). The mean wear time of cochlear implant was 38 months (SD=8.7) and the mean corrected hearing was 33.8 dB (SD=4.9).

```
Usage
## S3 method for class 'formula'
randomForest(formula, data=NULL, ..., subset, na.action=na.fail)


## Default S3 method:
randomForest(x, y=NULL, xtest=NULL, ytest=NULL, ntree=500,
mtry=if (!is.null(y) && !is.factor(y))
max(floor(ncol(x)/3), 1) else floor(sqrt(ncol(x))),
replace=TRUE, classwt=NULL, cutoff, strata,
sampsize = if (replace) nrow(x) else ceiling(.632*nrow(x)),
nodesize = if (!is.null(y) && !is.factor(y)) 5 else 1,
maxnodes = NULL,
importance=FALSE, localImp=FALSE, nPerm=1,
proximity, oob.prox=proximity,
norm.votes=TRUE, do.trace=FALSE,
keep.forest=!is.null(y) && is.null(xtest), corr.bias=FALSE,
keep.inbag=FALSE, ...)


## S3 method for class 'randomForest'
print(x, ...)
```

Fig. 5. Random Forest algorithm source of R program

### B. The Potential Factors related to the Articulation Accuracy of Children with Cochlear Implantation

The general characteristics and potential factors of the subjects according to the articulation accuracy are shown in Table 1. The articulation accuracy of the top group was 19 people out of 82 people (23.1%). The results of the chi-square test showed that the high-rank group and the low-rank group were significantly different in gender, age, the wear time of a cochlear implant, and vocabulary index (p<0.05).

### C. Major Predictors in the Random Forest Model

Table 3 shows the 'significance of variables' for the articulation accuracy of the children with cochlear implantation based on the random forest. The major predictors of the articulation accuracy of children with cochlear implantation were the wear time of a cochlear implant, the time since cochlear implantation, vocabulary index, household income, age, and gender, in the order of the magnitude. In other words, the wear time of a cochlear implant was the most important

predictor of the articulation accuracy. The final error rate of the random forest model developed by generating 500 bootstrap samples was 0.22, and the prediction rate was 78.8%.

TABLE II. POTENTIAL FACTORS RELATED TO THE ARTICULATION ACCURACY OF CHILDREN WITH COCHLEAR IMPLANTATION, %

| Characteristics | Articulation accuracy | | p |
|---|---|---|---|
| | High-rank group (n=19) | Low-rank groups (n=63) | |
| Sex, n (%) | | | |
| Male | 11 (19.6) | 45 (80.4) | <0.001 |
| Female | 8 (30.7) | 18 (69.3) | |
| Household income (1000 KRW), mean±SD | 3,850±1,230 | 3,370±1,750 | <0.001 |
| Age, mean±SD | 5.7±2.8 | 4.6±1.3 | 0.002 |
| Wear time of a cochlear implant (month), mean±SD | 42.5±9.3 | 33.8±8.6 | <0.001 |
| Time since cochlear implantation | 45.1±10.1 | 35.3±9.2 | <0.001 |
| Vocabulary index, mean±SD | 47.8±28.3 | 36.1±20.7 | <0.001 |
| Corrected hearing (dB), mean±SD | 33.2±5.2 | 34.5±4.8 | 0.850 |

TABLE III. THE SIGNIFICANCE OF VARIABLES' FOR THE ARTICULATION ACCURACY OF THE CHILDREN WITH COCHLEAR IMPLANTATION BASED ON THE RANDOM FOREST

| | |
|---|---|
| Wear time of a cochlear implant | 44.580 |
| Time since cochlear implantation | 39.905 |
| Vocabulary index | 35.734 |
| Household income | 23.108 |
| Age | 15.605 |
| Sex | 15.334 |

## V. DISCUSSION

Recently, the healthcare field tries to provide the foundation necessary for preventing the spread of a disease by identifying the source of it, providing a personalized treatment, predicting the future, and identifying a risk factor by applying the big data technology. This study developed a prediction model for predicting the articulation of school children with cochlear implantation by using a random forest, which is an ensemble based machine learning algorithm. This study established a model for predicting articulation production with considering multiple explanatory variables. It was found that the most important factor was the wear time of a cochlear implant, followed by the time since a cochlear implant operation and receptive vocabulary.

The age at the time of cochlear implantation is known to be an important factor affecting language development of hearing-impaired children. Numerous studies have shown that hearing-impaired children improved their speech production ability after they received cochlear implantation [36-38]. Moreover, the age at the time of implantation, the wear time of a cochlear implant, the remaining hearing before cochlear implantation were reported as factors affecting the ability to produce speech

after cochlear implantation [12]. These studies revealed that children who received cochlear implantation before age 5 had better speech intelligibility than those after 5 years old [36]. Moreover, the age at the time of cochlear implantation had the highest effects on speech intelligibility than other factors and the longer wear time of a cochlear implant showed better speech intelligibility [37]. Additionally, even for a study targeting children under 5 years old, subjects had faster language acquisition when they received the implantation earlier [38]. On the other hand, subjects who received the implantation after adolescence had poorer auditory performance than those who received it in childhood [12].

Vocabulary index is another major factor for predicting the improvement in communication ability after cochlear implantation. The linguistic ability has been reported as a significant variable for predicting the speech intelligibility along with the degree of hearing loss and age [38]. Particularly, the naming test and the vocabulary test were significantly correlated with the speech intelligibility of hearing-impaired children [39].

In summary, the results of this study suggested that the time of cochlear implantation and the wear time of a cochlear implant were more important factors than gender, household income, and age for predicting the speech intelligibility of children with cochlear implantation. Furthermore, the speech intelligibility after cochlear implant implies that linguistic ability such as receptive vocabulary is important in addition to the wear time of a cochlear implant.

The model, based on the random forest technique and developed for predicting the speech intelligibility of children with cochlear implantation, had better prediction power than the existing prediction models based on a regression model [24]. Moreover, it was reported that it provided more stable results because it is based on an ensemble algorithm, in addition to a good prediction performance [27, 40]. Therefore, it is believed that to use a random forest model would be more effective for estimating the importance of a variable among many variables than to use a regression model.

## VI. CONCLUSION

The results of this study on a prediction model suggested that it would be necessary to implement cochlear implantation and to develop a customized aural rehabilitation program considering the linguistic ability of a subject for enhancing the speech intelligibility of a child with cochlear implantation.

REFERENCES

[1] Health Insurance Review & Assessment Service, Statistical index of Health Insurance Review & Assessment Service, Health Insurance Review & Assessment Service, Seoul, 2017.

[2] Korea Centers for Disease Control and Prevention, Korean National Health and Nutrition Examination Survey, Korea Centers for Disease Control and Prevention, Sejong, 2017.

[3] F. R. Lin and J. K. Niparko, Measuring health-related quality of life after pediatric cochlear implantation: a systematic review. International Journal of Pediatric Otorhinolaryngology, vol. 70, no. 10, pp. 1695-1706, 2006.

[4] P. Schleich, P. Nopp and P. D'haese, Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant. Ear and hearing, vol. 25, no. 3, pp. 197-204, 2004.

[5]   T. Y. Ching, P. Incerti and M. Hill, Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. Ear and hearing, vol. 25, no. 1, pp. 9-21, 2004.

[6]   C. Morera, M. Manrique, A. Ramos, L. Garcia-Ibanez, L. Cavalle, A. Huarte, C. Castillo and E. Estrada, Advantages of binaural hearing provided through bimodal stimulation via a cochlear implant and a conventional hearing aid: a 6-month comparative study. Acta Oto-Laryngologica, vol. 125, no. 6, pp. 596-606, 2005.

[7]   K. Vermeire and P. Van de Heyning, Binaural hearing after cochlear implantation in subjects with unilateral sensorineural deafness and tinnitus. Audiology and Neurotology, vol. 14, no. 3, pp. 163-171, 2009.

[8]   R. Laszig, A. Aschendorff, M. Stecker, J. Müller-Deile, S. Maune, N. Dillier, B. Weber, M. Hey, K. Begall, T. Lenarz, R. D. Battmer, M, Bohm, T. Steffens, J. Strutz, T. Linder, R. Probst, J. Allum, M. Westhoten and W. Doering, Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-month tt results. Otology & Neurotology, vol. 25, no. 6, pp. 958-968, 2004.

[9]   M. Bond, S. Mealing, R. Anderson, J. Elston, G. Weiner, R. S. Taylor, M. Hoyle, Z, Liu, A. Price and K. Stein, The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: a systematic review and economic model. Health Technology Assessment, vol. 13, no. 44, pp. 1-330. 2009.

[10]  L. Spencer, N, Tye-Murray and B. Tomblin, The production of english inflectional morphology, speech production and listening performance in children with cochlear implants. Ear and Hearing, vol. 19, no. 4, pp. 310-318, 1998.

[11]  N. Tye-Murray, L. Spencer, E. Bedia and G. Woodworth, Differences in children's sound production when speaking with a cochlear implant turned on and turned off. Journal of Speech and Hearing Research, vol. 39, no. 3, pp. 604-610, 1996.

[12]  E. D. Thoutenhoofd, S. M. Archbold, S. Gregory, M. E, Lutman, T. P. Nikolopoulos and T. H. Sach, Paediatric cochlear implantation: Evaluating outcomes, Wiley Blackwell, Virginia, 2005.

[13]  R. Y. Litovsky, A. Parkinson and J. Arcaroli, Spatial hearing and speech intelligibility in bilateral cochlear implant users. Ear and hearing, vol. 30, no. 4, pp. 419-431, 2009.

[14]  T. A. Serry and P. J. Blamey, A 4 year investigation into phonetic inventory development in young cochlear implant users. Journal of Speech, Language, and Hearing Research, vol. 42, no. 1, pp. 141-154, 1999.

[15]  M. A. Svirsky, A. M. Robbins, K, I. Kirk, D. B. Pisoni and R. T. Miyamoto, R, Language development in profoundly deaf children with cochlear implants. Psychological science, vol. 11, no. 2, pp. 153-158, 2000.

[16]  C. Yoshinaga-Itano and A. Sedey, Early speech development of children who are deaf or hard of hearing: interrelationship with language and hearing. The Volta Review, vol. 100, no. 5, pp. 181-211, 2000.

[17]  E. A. Tobey, A. E. Geers, C. Brenner, D. Altuna, and G. Gabbert, G, Factors associated with development of speech production skills in children implanted by age five. Ear and hearing, vol. 24, no. 1, pp. 36-45, 2003.

[18]  S. H. Ferguson and D. Kewley-Port, Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. The Journal of the Acoustical Society of America, vol. 112, no. 1, pp. 259-271, 2002.

[19]  P. J. Blamey, J. Z. Sarant, L. E. Paatsch, J. G. Barry, C. P. Bow, R. J. Wales, M. Wright, C. Psarros, K. Rattigan and R. Tooher, Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing. Journal of Speech, Language, and Hearing Research, vol. 44, no. 2, pp. 264-285, 2001.

[20]  H. Byeon, D. Lee, and S. Cho, Assessment for the model predicting of the cognitive and language ability in the mild dementia by the method of data-mining technique. International Journal of Advanced Computer Science and Applications, vol. 7, no. 6, 2016.

[21]  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

[22]  Convergent with Art, Humanities, and Sociology, vol. 4, no. 2, pp. 303-312, 2014.

[22]  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.

[23]  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.

[24]  M. S. Yoon, H. S. Sim, S. O. Chang and C. S. Kim, Predictor variables of speech intelligibility after cochlear implant in Korean prelingually deafened children. Communication Sciences & Disorders, vol. 10, no. 3, pp. 57-70, 2005.

[25]  Y. T. Kim, G. H. Hong and K. H. Kim, Content and reliability analyses of the Receptive and Expressive Vocabulary Test (REVT). Communication Sciences & Disorders, vol. 14, no. 1, pp. 34-45, 2009.

[26]  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.

[27]  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, vol. 9, no. 9, pp. 494-499, 2018.

[28]  A. Singh, M. N. Halgamuge and R. Lakshmiganthan, Impact of different data types on classifier pe-rformance of random forest, naive bayes, and k-nearest neighbors algorithms. International Journal of Advanced Computer Science and Applications, vol. 8, no. 12, pp. 1-10, 2017.

[29]  H. Byeon, S. Yu, H. Jin and S. Cho, The development of mild cognitive impairment prediction model based on random forest with focusing on visuospatial function and language ability, Journal of Engineering and Applied Sciences, vol. 13, no. 2, pp. 2885-2888, 2018.

[30]  A. Statnikov, L. Wang and C. F. Aliferis, A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC bioinformatics, vol. 9, no. 1, pp. 319, 2008.

[31]  C. Strobl, A. L. Boulesteix, T. Kneib, T. Augustin and A. Zeileis, Conditional variable importance for random forests. BMC bioinformatics, vol. 9, no. 1, pp. 307, 2008.

[32]  L. Breiman, Random forests. Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[33]  D. R. Cutler, T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson and J. J. Lawler, Random forests for classification in ecology. Ecology, vol. 88, no. 11, pp. 2783-2792, 2007.

[34]  G. Biau, Analysis of a random forests model. Journal of Machine Learning Research, vol. 13, pp. 1063-1095, 2012.

[35]  A. Cutler, D. R. Cutler and J. R. Stevens, Random forests, In Ensemble machine learning, Springer, Boston, MA, 2012.

[36]  N. Tye-Murray, L. Spencer, and G. Woodworth, G, Acquisition of speech by children who have prolonged cochlear implant experience. Journal of Speech and Hearing Research, vol. 38, no. 2, pp. 327-337, 1995.

[37]  M. Mondain, M. Sillon, A. Vieu, M. Lanvin, F. Reuillard-artieres, E. Tobey and A. Uzeil, Speech perception skills and speech production intelligibility in French children with prelingual deafness and cochlear implants. Archives of Otolaryngology-Head & Neck Surgery, vol. 123, no. 2, pp. 181-184, 1997.

[38]  D. M. Houston and R. T. Miyamoto, Effects of early auditory experience on word learning and speech perception in deaf children with cochlear implants: implications for sensitive periods of language development. Otology & Neurotology, vol. 31, no. 8, pp. 1248-1253, 2010.

[39]  M. Yoon, Y. Lee, and H. S. Sim, The relationship between speech intelligibility and related factors of speakers in prelingually hearing impaired children using hearing aids. Korean Journal of Communication Disorders, vol. 5, no. 2, pp. 144-158, 2000.

[40]  H. Byeon, H. Jin and S. Cho, S. 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.