Comparing the Balanced Accuracy of Deep Neural Network and Machine Learning for Predicting the Depressive Disorder of Multicultural Youth

Haewon Byeon

Department of Medical Big Data

College of AI Convergence, Inje University, Gimhae 50834, Republic of Korea

Abstract—Multicultural youth are more likely to experience negative emotions (e.g. depressive symptoms) due to social prejudice and discrimination. Nevertheless, previous studies that analyzed the emotional aspects of multicultural youth mainly compared the characteristics of multicultural youth and those of the other youth or identified individual risk factors using a regression model. This study developed models to predict the depressive disorders of multicultural youth based on the Quick Unbiased Efficient Statistical Tree (QUEST), Classification And Regression Trees (CART), gradient boosting machine (G-B-M), random forest, and deep neural network (deep-NN) using epidemiological data representing multicultural youth and compared the prediction performance (PRED PER) of the developed models. Our study analyzed 19,431 youths (9,835 males and 9,596 females) aged between 19 and 24 years old. We developed models for predicting the self-awareness of health of youths by using QUEST, CART, G-B-M, random forest, and deep-NN and compared the balanced accuracy of them to evaluate their PRED PER. Among 19,431 subjects, 42.9% (5,838 people) experienced a depressive disorder in the past year. The results of our study confirmed that deep-NN had the best PRED PER with a specificity of 0.85, a sensitivity of 0.71, and a balanced accuracy of 0.78. It will be necessary to develop a model with optimal PRED PER by tuning hyperparameters (e.g., number of hidden layers, number of iterations, and activation function, number of hidden nodes) of deep-NN.

Keywords—*Quick unbiased efficient statistical tree; gradient boosting machine; deep neural network; classification and regression trees; balanced accuracy*

I. INTRODUCTION

South Korea has shifted to a multicultural society at a very rapid pace over the past 30 years. As a result, the number of multicultural youth has been increasing rapidly. The Ministry of Education (2019)[1] reported that the number of multicultural youth students was 137,225 people (2.5% of all students) in 2019, which is a 200% increase // an increase by 200% from 67,806 people in 2014. Particularly, the total fertility rate (TFR) decreased to 0.918 as of 2019 [2] due to the low fertility trend in Korean society, and the proportion of children with multicultural backgrounds in the school-age population is predicted to increase continuously [1].

Nevertheless, most of the policy interests targeting multicultural families (MUC-fam) in Republic of Korea have emphasized employment and welfare aspects, and there are not enough surveys on the mental health of MUC-fam [3,4]. In addition, studies conducted on multicultural youth have mainly focused on academic achievement or school adaptation, and emotional characteristics such as depression have been studied rarely [5].

Multicultural youth are more likely to experience negative emotions (e.g. depressive symptoms) due to social prejudice and discrimination [6]. Moreover, a national survey (epidemiological survey) [7] reported that the youth in low socioeconomic status such as low household income experienced a depressive disorder more. It was also implied that the multicultural youth were more likely to be exposed to negative emotionality since the proportion of MUC-fam in rural areas was higher than that in cities and only 9.7% of multicultural households made KRW 30 million or more per year. Above all, Byeon (2014)[8] revealed that approximately 15% of multicultural youth aged between 19 and 24 years old experienced social discrimination and children of international marriage families born in South Korea and immigrant children of international marriage families had difficulties in social adaptation due to the social characteristics of MUC-fam and rapid changes in adolescence. Since multicultural youth are highly likely to have multiple factors associated with a depressive disorder (e.g., sociodemographic factors and personal characteristics), they are believed to experience a depressive disorder more than the other youth. Nevertheless, previous studies [9,10,11,12] that analyzed the emotional aspects of multicultural youth mainly compared the characteristics of multicultural youth and those of the other youth or identified individual risk factors using a regression model.

It is necessary to develop a prediction model based on big data to grasp the characteristics of a depressive disorder because emotional issues are induced by multidimensional factors such as stress, social support, and environment. In recent years, machine learning (Mach Learn) has been widely used in various fields as a means to overcome the limitations of the Mach Learn regression model. This study developed models to predict the depressive disorders of multicultural youth based on quick unbiased efficient statistical tree (QUEST), classification and regression trees (CART), gradient boosting machine (G-B-M), random forest, and deep neural network (deep-NN) using epidemiological data representing multicultural youth and compared the prediction performance (PRED PER) of the developed models.

II. METHODS AND MATERIALS

A. Source of Data

This study analyzed the raw data of the National Survey of MUC-fam on MUC-fam residing in South Korea in 2012 jointly conducted by the Ministry of Health, Welfare and Family Affairs (MHWFA), the Ministry of Justice (MJ), and the Ministry of Gender Equality (MGE). The National Survey of MUC-fam was carried out to grasp the living conditions and welfare needs of MUC-fam to develop policies customized for MUC-fam [2]. This national survey consists of economic level, general characteristics, marital life, employment, health and health care. The National Survey of MUC-fam was conducted July 20 to October 31, 2012, and targeted all 154,333 marriage immigrants by using the status of foreign residents and the basic status of MUC-fam by the Ministry of Public Administration and Security (MPAS). The selection criteria for MUC-fam were based on the MUC-fam Support Act: this study targeted (1) families composed of marriage immigrants and South Korean citizens and (2) those composed of foreigners who obtained South Korean citizenship by acknowledgment or naturalization and South Korean citizens who obtained Republic of Korean citizenship by naturalization, birth or acknowledgment. This study analyzed 19,431 youths aged between 19 and 24 years old (9,835 males and 9,596 females) among the children of marriage immigrants (MI).

B. Measurement of Variables

A depressive disorder, the outcome variable, was divided into "experienced a depressive disorder" and "did not experience a depressive disorder" based on the question, "Have you ever felt sad or despair that lasted two weeks or longer in a row in the past year?", of the Diagnostic and Statistical Manual of Mental Disorders-five [13], a diagnostic criterion for major depressive disorder. Explanatory variables included gender, education, residence area (rural or urban), economic activity (yes or no), social discrimination experience (yes or no), career counseling experience (yes or no), Korean language education experience (yes or no), the experience of using a support center for multi-cultural families (yes or no), level of Korean reading (good, average, or poor), level of Korean writing (good, average, or poor), level of Korean listening, level of Korean speaking, and experience of Korea society adaptation training (yes or no) by referring to previous studies [9,10,11,12] related to a depressive disorder.

C. Developing Models for Predicting a Depressive Disorder: QUEST

QUEST [13] selects a significant variable among variables entered first, performs a quadratic discriminant analysis based on the selected variable, and selects a predictor for reducing the variable selection bias. The variable selection is made (1) by ANOVA F-statistics for continuous variables and (2) by choosing the various with the smallest probability as a classification variable for categorical variables using the chi-square test. It is characterized by selecting a threshold by finally conducting quadratic discriminant analysis for the classification variable selected in this process [14].

The QUEST algorithm for predicting and classifying variables is performed by the following procedure [15]. First, the significance probability (p-value) of an ordinal or continuous predictor is calculated by the F-test of ANOVA. Second, when a categorical predictor is selected, the p-value of the cross-validation is calculated in the contingency table of predictor and target variables. If the p-value is smaller than the adjusted Bonferroni's threshold, the corresponding variable is selected as a classification variable. If there are more than 3 categories, it is replaced with a binary classification based on two-means clustering. Third, if a separation variable cannot be selected by the above two procedures, the p-value of the Levene F-test is calculated for an ordinal or continuous classification variable, and it is compared with the adjusted Bonferroni's threshold to finally select the variable corresponding to the p-value as the classification variable. This procedure is repeated until the condition is fully satisfied. For data composed of categorical variables, if the categories of outcome variables are three or more, the two-mean clustering consisting of two categories is performed. At this time, quadratic discriminant analysis is used to find the optimal threshold of explanatory variables, and child nodes are formed by finding explanatory variables that classify the outcome variables best.

D. CART

CART measures impurity using the Gini Index (Gini impurity) and performs a binary split in which only two child nodes are formed from a parent node [16]. The Gini impurity refers to the probability that two elements randomly extracted from n elements belong to different groups. When the decrement of the Gini Index is calculated, the predictor and the optimal threshold that decreases the Gini Index most are selected as child nodes, as the final step.

E. G-B-M

G-B-M is a Mach Learn algorithm that creates a strong learner (S-learner) by combining weak learners (W-learner) of decision trees (Decis Tree) by using an ensemble technique [17]. This method generalizes models by generating models for each step and optimizing the loss function (loss-func) that can be arbitrarily differentiated, like other boosting methods. Even if the accuracy is low, a model is created, and the error of the generated model is complemented by the next model. Through this process, more powerful learner), than the previous a learner, is created. The basic principle is to increase accuracy by repeating this process. The concept of G-B-M is presented in Fig. 1.

Fig. 1. The Concept of G-B-M.

F. Random Forest

Random forest is a decision tree-based ensemble method that generates a large number of random samples (Rand Sam) from a training dataset through bootstrapping (random sampling with replacement of the same size from a given data), learns an independent Decis Tree for each sample (Eac Sam) set, and creates a final model by synthesizing the results. The structure of it is presented in Fig. 2.

G. Deep-NN

Deep-NN is an algorithm composed of an composed of independent variables, an composed of dependent variables, and two or more hidden layers between them. Independent node (Ind node)s are arranged in each layer of the input layer, the hidden layer, and the output layer. A group of nodes in each layer is connected by weighted neurons (connecting lines) (Fig. 3).

This study used H2O's deep-NN among various deep learning types. H2O's deep-NN is based on a multi-layer feedforward artificial neural network (A-NN) that is trained with stochastic gradient descent (Grad Des) using back-propagation (Back-propag). This study set two hidden layers with ten nodes in each layer (20 nodes in total) and five epochs (number of iterations). This study used the Rectifier Linear Unit (ReLU), default value, as the activation function of deep learning (Fig. 4).

H. Validation of the Models

This study developed models for predicting the subjective health of youth by using QUEST, CART, G-B-M, random forest, and H2O's deep-NN and compared the balanced accuracy of them to evaluate their PRED PER. Balanced accuracy is an index that considers sensitivity and specificity and presents the same value as the area under the curve (AUC), indicating the accuracy of a classification model. This study used 10-fold cross-validation (Cro-Valid) to validate the developed models.
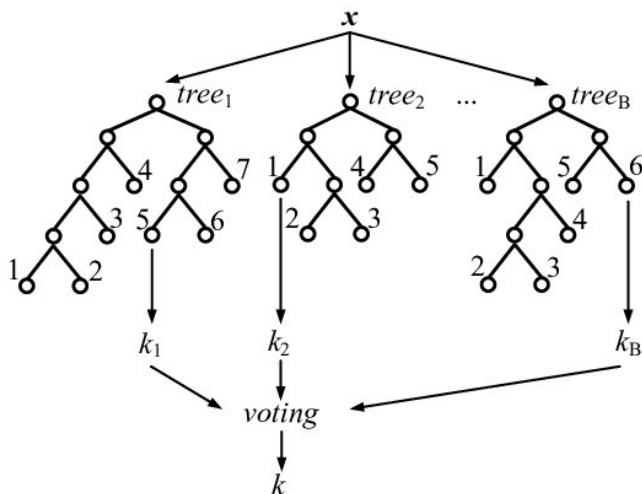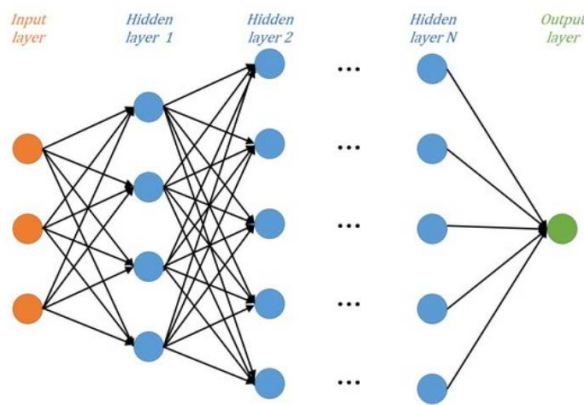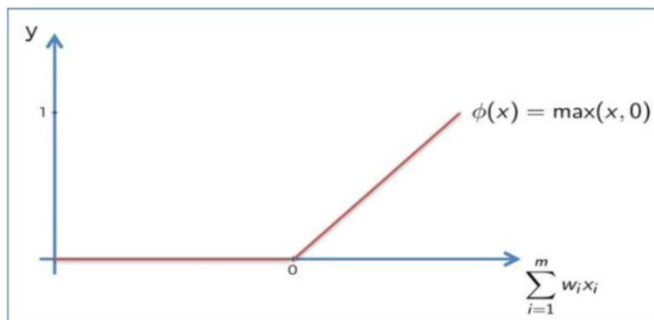

Fig. 3. Layers of Deep-NN [19].

Fig. 4. The Concept of Rectified Linear Unit Activation Function [20].

In this study, a model containing randomness, such as random forest, was developed while fixing the seed to "435435". This study defined the model with the highest accuracy as the model with the best PRED PER by comparing the PRED PER of these models. When the accuracy of multiple models was the same, the model with high sensitivity was defined as the model with the best PRED PER. All analyses were performed using R version 4.0.4 (Foundation for Statistical Computing, Vienna, Austria).

III. RESULTS

The general characteristics of subjects by depression experience are presented in Table I. Among 19,431 subjects, 42.9% (5,838 people) experienced a depressive disorder in the past year. The results of chi-square test showed the highest level of education, gender, economic activity, Korean language education experience, level of Korean speaking, Korean writing level, level of Korean listening, Korean reading level, Korea society adaptation training experience, career counseling experience, social discrimination experience, and experience of using a support center for MUC-fams were significantly ($p<0.05$) different between multicultural youth with a depressive disorder experience and those without a depressive disorder experience.

Fig. 2. Structure of Random Forest Model [18].

TABLE I.   CHARACTERISTICS OF SUBJECTS BASED ON EXPERIENCE OF DEPRESSIVE DISORDER, N (%)

Variables	Depressive disorder		p
	Yes (n=5,838)	No (n=13,593)	
Residence area			0.308
Rural	4,970 (30.2)	11,492 (69.8)	
Urban	869 (29.3)	2,101 (70.7)	
Level of education			<0.001
Elementary school or below	0	202 (100)	
Junior high school	725 (29.2)	1,754 (70.8)	
High school	4,376 (30.8)	9,812 (69.2)	
College or higher	738 (28.8)	1,825 (71.2)	
Gender			
Male	2,739 (27.8)	7,097 (72.2)	
Female	3,099 (32.3)	6,496 (67.7)	
Economic activity			0.001
Yes	2,311 (28.7)	5,732 (71.3)	
No	3,527 (31.0)	7,861 (69.0)	
Korean language education experience			<0.001
No	5,460 (30.6)	12,402 (69.4)	
Yes	378 (24.1)	1,191 (75.9)	
Korean writing level			0.012
Good	4,392 (29.7)	10,372 (70.3)	
Average	798 (29.5)	1,903 (70.5)	
Poor	648 (33.0)	1,318 (67.0)	
Korean reading level			<0.001
Good	4493 (29.7)	10,640 (70.3)	
Average	720 (28.5)	1,809 (71.5)	
Poor	625 (35.3)	1,144 (64.7)	
Korean speaking level			0.005
Good	4,456 (29.5)	10,659 (70.5)	
Average	941 (31.7)	2.024 (68.3)	
Poor	441 (32.6)	911 (67.4)	
Korean listening level			<0.001
Good	4,459 (29.0)	10,925 (71.0)	
Average	984 (35.1)	1,821 (64.9)	
Poor	395 (31.8)	847 (68.2)	
The experience of using a support center for MUC-fams			<0.001
No	5,641(30.7)	12,706 (69.3)	
Yes	197 (18.2)	887 (81.8)	
Korea society adaptation training experience			<0.001
Yes	1,810 (33.9)	3,524 (66.1)	
No	4,028 (28.6)	10,069 (71.4)	
Social discrimination experience			<0.001
No	4,196 (25.6)	12,200 (74.4)	
Yes	1,642 (54.1)	1,393 (45.9)	
Career counseling experience			<0.001
No	5,273 (29.0)	12,941 (71.0)	
Yes	565 (46.4)	653 (53.6)	

The balanced accuracy of five models (QUEST, CART, G-B-M, random forest, and H2O's deep-NN) for predicting a depressive disorder of multicultural youth is presented in Fig. 5. The results of our study confirmed that H2O's deep-NN had the best PRED PER with a specificity of 0.85, a sensitivity of 0.71, and a balanced accuracy of 0.78. The normalized importance of variables was analyzed using H2O's deep-NN. The major predictors of multicultural youth's depressive disorder were social discrimination experience, gender, level of Korean speaking, Korean writing level, and level of Korean listening (Fig. 6). Among them, social discrimination experience was the most important factor in predicting a depressive disorder (Fig. 6).
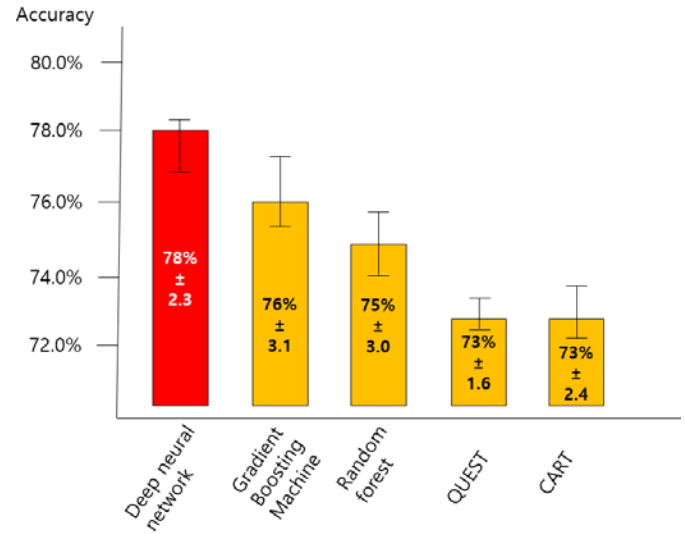

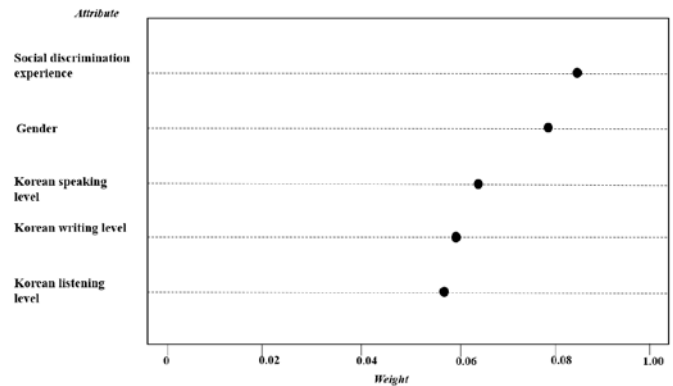Fig. 5.   Comparison of Balanced Accuracy of 5 Models for Predicting Depression, (%).

Fig. 6.   Importance of Variables (Impor Var) in the Depression.

IV. CONCLUSION

Our study compared the balanced accuracy of models for predicting the depressive disorder of multicultural youth and confirmed that H2O's deep-NN was the model with the best PRED PER among QUEST, CART, G-B-M, random forest, and H2O's deep-NN. On the other hand, tree-based Mach Learns such as QUEST and CART had relatively low balanced accuracy compared to other Mach Learn methods. These results agreed with da Krauss et al. [21], which showed that the

prediction accuracy of deep-NN was better than gradient-boosted-trees or random forest. On the other hand, Grolinger et al. [22] reported that H2O's deep-NN had worse PRED PER than support vector regression and took longer computation time. These results imply that the most optimal algorithm with the best PRED PER may vary depending on the data type. Therefore, more follow-up studies are needed to compare the performance of Mach Learn and deep-NNs using various datasets with different characteristics.

Currently, various open-source platforms (e.g., Torch, MXNet, TensorFlow, Caffe, Theano, and H2O) have been developed for deep learning. Among them, the H2O platform provides access to Mach Learn algorithms through common development environments (e.g., Python, R, and JAVA), big data systems (e.g., Hadoop and Spark), and various data sources (e.g., SQL, NoSQL, HDFS, and S3). This study developed a model to predict the depressive disorder of multicultural youth using the defaults of the H2O platform (i.e., number of hidden layers, number of iterations, number of hidden nodes, and activation function). It was confirmed that the balanced accuracy of predict the depression was superior to that of other Mach Learn algorithms. It will be necessary to develop a model with optimal PRED PER by tuning hyper-parameters (e.g., number of hidden layers, number of iterations, number of hidden nodes, and activation function) of H2O's deep-NN.

Our study developed a model for predicting the depressive disorder experience of multicultural youth using H2O's deep-NN and found that social discrimination experience, gender, level of Korean speaking, Korean writing level, and level of Korean listening were related to the depressive disorder of multicultural youth. Among them, social discrimination experience was the most influential factor in predicting a depressive disorder. Consequently, it is required to prepare a legal system that can help multicultural youth overcome discrimination and prejudice and give attention to them at the social level.

ACKNOWLEDGMENT

REFERENCES

[1] Ministry of Education, 2019 Basic Education Statistics. Ministry of Education, Sejong, 2019.

[2] National Statistical Office, Population Trend Survey. National Statistical Office, Daejeon, 2021.

[3] H. Kim, and J. G. Lee, Multicultural society: examining Korea and China's policies for ethnic minorities and immigrants. The Social Science Research Institute, vol. 12, no. 1, pp. 37-67, 2019.

[4] D. H. Lee, A study on multicultural families support act. The Journal of the Korea Contents Association, vol. 19, no. 7, pp. 650-658, 2019. doi: 10.5392/JKCA.2019.19.07.650.

[5] J. Y. Oh, Role of multicultural family support center for school adaptation of elementary school children of multicultural families. Cultural Exchange and Multicultural Education, vol. 7, no. 3, pp. 79-99, 2018.

[6] K. H. Han, Predictors of depressive symptom trajectories among multicultural family adolescents. Mental Health & Social Work, vol. 48, no. 1, pp. 56-83, 2020. doi: 10.24301/MHSW.2020.03.48.1.56.

[7] Ministry of Health & Welfare, Korean youth obesity survey. Ministry of Health & Welfare, Seoul, 2009.

[8] 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, vol. 4, no. 2, pp.303-312, 2014. doi: 10.14257/AJMAHS.2014.12.12.

[9] M. K. Park, Y. Cha, H. Lee, The effects of using the multicultural family support centers on the bicultural identity of the multicultural youths: focusing on propensity score matching. The Journal of Multicultural Society, vol. 12, no. 3, pp. 107-140, 2019. doi: 10.14431/jms.2019.10.12.3.107.

[10] R. H. Lee, Daily stress and smartphone overdependence among multicultural adolescents: comparisons by gender and age. Journal of Digital Convergence, vol. 16, no. 11, pp. 561-569, 2018. doi: 10.14400/JDC.2018.16.11.561.

[11] B. Seo, Analysis of factors and actual status of delinquency among multi-cultural youth In Korea. Contemporary Society and Multiculture, vol. 9, no. 2, pp. 123-150, 2019.

[12] K. H. Han, and S. K. Kahng, The relationship among acculturative stress, depression and self-esteem of multicultural adolescents-testing a theoretical model of the stress process. Mental Health & Social Work, vol. 47, no. 1, pp. 231-257, 2019. doi: 10.24301/MHSW.2019.03.47.1.231.

[13] J. C. Tolentino, and S. L. Schmidt, DSM-5 criteria and depression severity: implications for clinical practice. Frontiers in Psychiatry, vol. 9, pp. 450, 2018. doi: 10.3389/fpsyt.2018.00450.

[14] C. L. Lin, and C. L. Fan, Evaluation of CART, CHAID, and QUEST algorithms: a case study of construction defects in Taiwan. Journal of Asian Architecture and Building Engineering, vol. 18, no. 6, pp. 539-553, 2019. doi: 10.1080/13467581.2019.1696203.

[15] H. Byeon, and O. Lee, Development of prediction model for life satisfaction among Korean elderly using QUEST algorithm: findings from a community-based study. International Journal of Applied Engineering Research, vol. 10, no. 79, pp. 88-92, 2019.

[16] M. M. Ghiasi, S. Zendehboudi, and A. A. Mohsenipour, Decision tree-based diagnosis of coronary artery disease: CART model. Computer Methods and Programs in Biomedicine, vol. 192, pp. 105400, 2020. doi: 10.1016/j.cmpb.2020.105400.

[17] S. Touzani, J. Granderson, and S. Fernandes, Gradient boosting machine for modeling the energy consumption of commercial buildings. Energy and Buildings, vol. 158, pp. 1533-1543, 2018. doi: 10.1016/j.enbuild.2017.11.039.

[18] A. Gelzinis, A, Verikas, E. Vaiciukynas, M. Bacauskiene, J. Minelga, M. Hållander, V. Ulozas, and E. Padervinskis, Exploring sustained phonation recorded with acoustic and contact microphones to screen for laryngeal disorders. In 2014 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE), pp. 125-132, 2014. doi: 10.1109/CICARE.2014.7007844.

[19] C. Lin, Q. Chang, and X. Li, A deep learning approach for MIMO-NOMA downlink signal detection. Sensors, vol. 19, no. 11, pp. 2526, 2019. doi: 10.3390/s19112526.

[20] M. Bhurtel, J. Shrestha, N. Lama, S. Bhattarai, A. Uprety, and M. K. Guragain, Deep learning based seed quality tester. Science, Engineering and Technology (SET) Conference, pp.1-10, 2019.

[21] C. Krauss, X. A. Do, and N. Huck, Deep neural networks, gradient-boosted trees, random forests: statistical arbitrage on the S&P 500. European Journal of Operational Research, vol. 259, no, 2, pp. 689-702, 2017.

[22] K. Grolinger, M. A. Capretz, and L. Seewald, Energy consumption prediction with big data: Balancing prediction accuracy and computational resources. In 2016 IEEE International Congress on Big Data (BigData Congress), pp. 157-164, 2016. doi: 10.1109/BigDataCongress.2016.27.