



ISSN: 2508-7894

KJAI website: <http://acoms.kisti.re.kr/kjai>doi: <http://dx.doi.org/10.24225/kjai.2023.11.3.29>

Development of Big Data-based Cardiovascular Disease Prediction Analysis Algorithm

Kyung-A KIM¹, Dong-Hun HAN², Myung-Ae CHUNG³

Received: July 31, 2023. Revised: August 29, 2023. Accepted: September 05, 2023.

Abstract

Recently, the rapid development of artificial intelligence technology, many studies are being conducted to predict the risk of heart disease in order to lower the mortality rate of cardiovascular diseases worldwide. This study presents exercise or dietary improvement contents in the form of a software app or web to patients with cardiovascular disease, and cardiovascular disease through digital devices such as mobile phones and PCs. LR, LDA, SVM, XGBoost for the purpose of developing “Life style Improvement Contents (Digital Therapy)” for cardiovascular disease care to help with management or treatment We compared and analyzed cardiovascular disease prediction models using machine learning algorithms. Research Results XGBoost. The algorithm model showed the best predictive model performance with overall accuracy of 80% before and after. Overall, accuracy was 80.0%, F1 Score was 0.77~0.79, and ROC-AUC was 80%~84%, resulting in predictive model performance. Therefore, it was found that the algorithm used in this study can be used as a reference model necessary to verify the validity and accuracy of cardiovascular disease prediction. A cardiovascular disease prediction analysis algorithm that can enter accurate biometric data collected in future clinical trials, add lifestyle management (exercise, eating habits, etc.) elements, and verify the effect and efficacy on cardiovascular-related bio-signals and disease risk. development, ultimately suggesting that it is possible to develop lifestyle improvement contents (Digital Therapy).

Keywords : Bigdata Analysis, Cardiovascular Disease, Machine Learning, Algorithm, XGBoost, Digital Therapy

Major Classification Code : Artificial Intelligence, Bigdata, Machine Learning

1. Introduction

According to the World Health Organization (WHO), cardiovascular disease is the leading cause of death in 2019 and the second leading cause of death in Korea. As reported

by the statistical analysis of heart disease by the Health Insurance Review and Assessment Service in Korea, the number of patients with major heart disease in 2020 is about 1.62 million, an increase of 16.9% compared to 2016, and an average annual increase of 4%. The width is the largest at 42%.

* This paper was supported by IITP (Institute of information & Communications Technology Planning & Evaluation(www.iitp.kr), Korea) [Project Number: 2022-0-00317]. This paper was supported by the research grant of the KODISA scholarship foundation in 2023.

1 First Author. Doctoral student, Department of Medical Artificial Intelligence, Eulji University, Korea.
Email: kyungakim@naver.com

2 Second Author. Master's student, Department of Medical Artificial Intelligence, Eulji University, Korea. Email: d555v@naver.com

3 Corresponding Author. Professor, Department of BigData Medical Convergence, Eulji University, Korea.
Email: machung@eulji.ac.kr

© Copyright: The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cardiovascular disease is one of the diseases that require long-term treatment as a chronic disease, and requires national management due to medical expenses and indirect economic loss costs. Machine learning has generally been used in the field of computer science, although it has been actively applied in the clinical medical field recently (Choi et al., 2018). Therefore, effective diagnostic algorithms and early risk assessments that can simultaneously meet the dual goals of time and cost effectiveness and patient safety are critical.

The causes of cardiovascular disease are obesity, an increase in the elderly population, excessive energy intake, and a decrease in physical activity. Cardiovascular disease was defined as myocardial infarction, angina pectoris, and coronary artery disease, and the major risk factors for cardiovascular disease were hypertension, diabetes, hyperlipidemia, and obesity. These risk factors can be practically prevented through health changes such as smoking cessation, increased physical activity, and improved eating habits, as well as through early detection and management of diseases.

Machine learning (ML) is a popular approach used in the medical research context to automate data analytic modelling (Mohd Faizal et al., 2021). The purpose of this study is to find a machine learning algorithm model suitable for predicting cardiovascular disease by collecting data from patients with cardiovascular disease and analyzing the relationship between cardiovascular disease and risk factor variables based on the collected database. In the future, it is intended to be used as a predictive model necessary to verify the validity and accuracy of cardiovascular disease prediction.

2. Theoretical Background

2.1. Prior Studies Related to Patients with Cardiovascular Disease

Recently, studies to classify a patient's disease or predict a risk using a machine learning classifier based on big data are widely conducted (Oh et al., 2022).

Research on disease prediction and diagnosis of cardiovascular diseases based on medical data has also been conducted for a long time. As we move into the age of precision medicine, understanding the use of phenotypic data and methods to analyze already acquired information is of paramount importance (Venkatesh et al., 2017).

Park et al. (2018) a comparative study of machine learning algorithms for diagnosing ischemic heart disease was conducted, providing clinicians with an auxiliary diagnostic role to help explain the causal relationship of disease, and minimizing unnecessary tests to improve both doctors and patients. Suggested that diagnostic predictions that can increase time and cost efficiency are possible.

Bosma et al. (1998) proposed a predictive model for

coronary artery disease including age, gender, blood pressure, total cholesterol, serum high-density lipoprotein cholesterol, smoking, diabetes, and left ventricular hypertrophy. In Korea, Jee et al. (2004) a model was created to predict the risk of ischemic heart disease, including age, total cholesterol, systolic blood pressure, smoking, and diabetes. Hong et al. (2015) created a model to predict cardiovascular disease in community elders aged 60 years or older. Made a model that in Kim and Kim (2018), a prediction model for each heart was created using a deep learning-based multi-layer perceptron. Most cardiovascular disease predictive model studies are mainly studies using statistical techniques, and recently, many studies using machine learning algorithms and artificial intelligence have been conducted. This suggests that machine learning has an important clinical role in assessing the prognostic risk of individual patients with suspected cardiovascular disease (Motwani et al., 2017).

2.2. Machine Learning Analytics

The machine learning algorithm used in this study is as follows.

Logistic regression (LR) is an algorithm widely used in epidemiological studies, and in this study, it was used as a reference for comparison with other algorithms that analyze data. The purpose of LR is to express the relationship between the dependent variable and the independent variable as a specific function, like the goal of general regression analysis, and use it in future prediction models. Since the dependent variable of LR targets categorical data and divides the result into a specific classification, it can be seen as a classification technique.

Linear discriminant analysis (LDA) is the most commonly used algorithm in machine learning. It learns the distribution of data to determine the decision boundary. It is a method of classifying data by creating a decision boundary. When classifying given data into K classes, we aim to find a straight line where the center (mean) of each class is farther away and the variance is small.

Support Vector Machine (SVM) is a supervised learning model for pattern recognition and data analysis in the field of machine learning, and is a representative algorithm used in data mining techniques and artificial intelligence. It is an analysis technique that models a hyperplane by dividing the training data set into the widest range, and can handle high-accuracy and high-dimensional data (Noh & Park, 2020). There are classification algorithms such as decision tree and artificial neural network, but SVM shows better performance than these two in classifying data, especially binary data.

Ensemble BT is an algorithm that sequentially learns and predicts several weak learners using weights to make them strong learners. It is a method of gradually learning while

improving errors by assigning weights to incorrectly predicted data. Typically, a gradient boosting model is used, and many models with excellent predictive performance use this technique.

In this study, Logistic regression (LR), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), which are most commonly used in machine learning, XGBoost a cardiovascular disease prediction model was examined using an algorithm.

3. Research Methods

3.1. Data Collection and Variables

This study was conducted with the approval of the Eulji University IRB (EU22-74-01), and raw data from the Korea National Health and Nutrition Examination Survey (KNHANES) by the Korea Centers for Disease Control and Prevention a total of 115,608 people were studied, 61,400 from 2008 to 2014 and 54,208 from 2015 to 2021.

This data is open data that can be individually identified, and information codes such as 'patient name' and 'personal identification information' are deleted. Variable items of the data include gender, age, heart rate, systolic period, diastolic period, height, weight, waist circumference, body mass, smoking, current smoking, hypertension, duration of hypertension, diabetes, duration of diabetes, hyperlipidemia, duration of hyperlipidemia, myocardial infarction, duration of myocardial infarction, angina pectoris, duration of angina pectoris, coronary artery, blood data (cholesterol, etc.), exercise, eating habits, housing type, etc. Feature selection is essential for predictive modeling, and machine learning is particularly useful for it (Kipp et al., 2018).

Among these, as mentioned in the introduction, cardiovascular diseases and variable items corresponding to the main causes of cardiovascular diseases Fifteen were selected and a study was conducted to analyze the association between cardiovascular disease and risk factor variables.

3.2. Data Preprocessing

This study used the significant data of 4,874 patients with cardiovascular disease among 115,608 total data from 2008 to 2021 of the Korea National Health And Nutrition Examination Survey(KNHANES). The data used were coded as 1 or 0 with the physician's diagnosis of heart disease as a dependent variable.

Including seven factors (gender, age, systolic blood pressure, antihypertensive drug intake, smoking, diabetes, body mass index) used in the heart age prediction model of the Framingham Heart Study (FHS) to find out the

correlation with cardiovascular disease (Cho, 2019), 2 basic information (gender, age), 10 cardiovascular disease factors (systolic blood pressure, diastolic blood pressure, body mass index, fasting blood sugar, hypertension, dyslipidemia, stroke, diabetes, cancer, kidney failure), 3 lifestyle habits (smoking, smoking, Smoking, energy intake, and number of days of exercise) A total of 15 input variables were selected as independent variables of the predictive model. Additionally, all missing values for independent variable items were excluded.

3.3. Data Analysis and Statistical Methods

In this study, Python 3.11 (Python Software Foundation, Wilmington, De, USA) language, Scikit-learn 1.2.2, Scipy 1.3.2 library and IBM SPSS Statistics 20 were used for data analysis. The predictive model used logistic regression, linear discriminant analysis, support vector machine, and XGBoost machine learning algorithms.

First, preprocessing was performed, training data and verification data were divided at a ratio of 7:3 (Kang et al., 2022), the model was trained with the learning data and the model was evaluated with the test data (Lee et al., 2021). In order to analyze the statistical characteristics of the input variables and cardiovascular disease, the comparison of 'continuous variables' between the two groups (learning data, training data) was evaluated by unpaired t-test or Mann-Whitney rank test, mean \pm standard It was presented as a deviation, and the comparison of categorical variables between the two groups was evaluated with a chi-square test or Fisher's exact test, and presented as count and percentage. In addition, relative risk analysis was performed for each variable used to predict cardiovascular disease, and the statistically significant level was set at a P value of less than 0.05.

The performance evaluation of the learning model completed with the machine learning algorithm was evaluated by Accuracy, F1 Score, and ROC-AUC. ROC-AUC has a value between 0 and 1, and the closer it is to 1, the better the performance of the model can be judged.

3.4. Criteria and Methods for Evaluating Model Effects

In this study, Logistic regression (LR), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), XGBoost, which are most commonly used to develop a model for predicting cardiovascular disease using clinical data variables and machine learning technology algorithm was used.

The selection of variables for the creation of a cardiovascular disease prediction model was performed using the 'information gain attribute evaluation' method. These are basic classifiers for variable selection of the predictive model, and variables were selected with XGBoost, an Ensemble classification model

using a single-node decision tree. Information gain is a measure of the effectiveness of an attribute when classifying training data, and this value is the entropy of each class. Measured as a decreasing amount, it reflects the additional information about the class provided by the attribute (Motwani et al., 2017).

The results of classification are expressed as the probability of being included in each class. This represents the probability that an element x is in class y. In this study, attributes that obtain information of 0.001 or more were applied to the model.

For cross validation, 10-fold stratified cross-validation, which is preferred in the field of data mining, was used to estimate the model's performance and general error in the machine learning process. Randomly divides the data set into 10 intervals of equal size, so that each interval has the same number of events. In the cross-validation of the 10th quantile, each interval is used for validation and the remaining 9 intervals are used as training intervals, so the final 10 XGBoost model was created. The performance of the entire machine learning was measured by combining the validation results of the generated 10 models.

The end point of the study is to create a predictive model for cardiovascular disease and increase the predictive rate of the predictive model. It was presented as ROC 's AUC (The range area under the curve of receiver-operating characteristic analysis) (Hyun, 2022) and the performance evaluation of the cardiovascular disease prediction learning model generated by machine learning was The AUC of the ROC analysis was evaluated. AUC has a value between 0.5 and 1. A value of 0.5 means no predictive power, a value of 0.8 or more means good predictive power, and a value of 1 means perfect predictive power (Baek et al., 2018).

The internal verification accuracy of the learning model for cardiovascular disease prediction was set to ROC-AUC 0.85 (85%) or higher, and the external verification accuracy ROC-AUC 0.80 (85%) or higher.

4. Results and Discussion

Machine learning algorithm, the following results were obtained (see Figure 1).

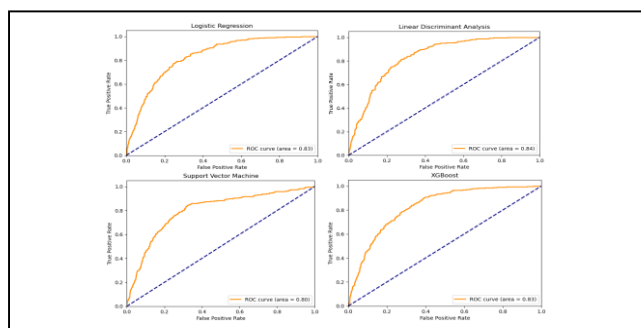


Figure 1: Machine Learning Model ROC Curve

Relative risk analysis was performed on each variable used to predict cardiovascular disease, and the statistical characteristics of the variables are described in Table 1(see Table 1).

Table 1: Baseline characteristics and odds ratios of cardiovascular disease variables

Variables	Total (N=4784)	CAD (N=1055)	No CAD (N=2945)	Odds ratio [95% CI]	P value
Sex, male	2620(54.7)	632(13.2)	1988(41.5)	1.225 [1.067 - 1.405]	<0.001
Age, yrs	54.6 ± 16.7	67.2 ± 9	50.9 ± 16.6	1.082 [1.074 - 1.091]	<0.001
Blood pressure: BP, mmHg					
Systolic BP	120.3 ± 17.7	126.6 ± 17.3	118.5 ± 17.4	1.006 [1.000 - 1.011]	<0.001
Diastolic BP	74.9 ± 10.5	73.4 ± 10.3	75.4 ± 10.5	0.985 [0.976 - 0.995]	<0.001
BMI, Kg/m ²	23.9 ± 3.4	24.7 ± 3	23.7 ± 3.5	1.083 [1.058 - 1.109]	<0.001
Hypertension	1708(35.7)	696(14.5)	1012(21.2)	4.850 [4.199 - 5.603]	<0.001
Diabetes	722(15.1)	308(6.4)	414(8.6)	3.181 [2.691 - 3.759]	<0.001
Dyslipidemia	1053(22.0)	444(9.3)	609(12.7)	3.561 [3.068 - 4.134]	<0.001
Stroke	219(4.6)	91(1.9)	128(2.7)	2.577 [1.952 - 3.403]	<0.001
Kidney disease	39(0.8)	21(0.4)	18(0.3)	4.071 [2.161 - 7.668]	<0.001
Cancer	277(5.7)	80(1.7)	197(4.1)	1.428 [1.091 - 1.869]	<0.001
Current smoking	2030(42.4)	570(11.9)	1460(30.5)	1.726 [1.506 - 1.979]	<0.001
Walking days, day	3.9 ± 2.4	3.5 ± 2.6	4 ± 2.4	0.979 [0.950 - 1.010]	<0.001
Energe intake, Kcal (HR:100Kcal)	1992 ± 807.9	1881.4 ± 726.3	2024.2 ± 827.4	1.000 [1.000 - 1.000]	<0.001
Fasting blood glucose, mg/dL	102.7 ± 25.3	108.6 ± 26.8	101 ± 24.6	1.001 [0.999 - 1.004]	<0.001

10-fold was performed using k-fold cross validation to evaluate the performance of the model. The mean accuracy of 10-fold cross validation was 0.80, and the standard deviation of accuracy was 0.016, verifying the suitability of the evaluation and learning models.

As a result of the prediction of the machine learning model in this study, Accuracy showed LR 0.8001, LDA 0.7938, SVM 0.7938, and XGBoost 0.8022, showing generally similar results, and F1 Score LR 0.7780, LDA 0.7774, SVM 0.7656. XGBoost was 0.7908, and the values of ROC AUC were LR 0.8368, LDA 0.8366, SVM 0.7991, and XGBoost 0.8285 (see Table 2).

Table 2: The Results of Model Comparison

Machine Learning Model	Accuracy	F1 Score	ROC-AUC
Logistic Regression(LR)	0.8001	0.7780	0.8368
Linear Discriminant Analysis(LDA)	0.7938	0.7774	0.8366
Support Vector Machine(SVM)	0.7938	0.7656	0.7991
XGBoost(Ensemble BT)	0.8022	0.7908	0.8285

Machine Learning Model ROC Curve Looking at the graph, it was found that the ROC-AUC values of the models of the machine learning algorithm were generally at a similar level, ranging from 80% to 84% (figure 1). Of these, XGBoost Since the algorithm model has an overall accuracy of around 80%, it can be said that the performance is the best.

Machine learning algorithm models used in this study, XGBoost showed the best performance. The importance of the variables of the model for the algorithm is as shown in Fig 2, and it was found that the correlation with cardiovascular disease was high in the order of age, hypertension, exercise, stroke, gender, diabetes, and

smoking (see Figure 2).

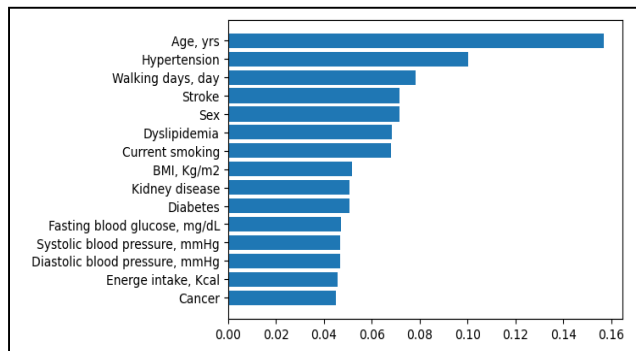


Figure 2: Feature importance score in XGBoost model

5. Conclusions

In this study, a cardiovascular disease prediction model was performed using four machine learning algorithms (LR, LDA, SVM, and XGBoost). Before the study, ROC-AUC 0.85(85%) was suggested to evaluate the performance of the algorithm prediction model, but as a result of the study, the overall accuracy was 80.0%, the F1 score was 0.77-0.79, and the ROC-AUC was 80% to 84%. Therefore, it was confirmed that it can be used as a reference model necessary to verify the validity and accuracy of cardiovascular disease prediction.

As a result of this study, exercise in patients with heart disease was found to be the third most important variable in the effect on cardiovascular disease. In the future, it suggested the possibility of developing a cardiovascular disease predictive analytics algorithm that can verify the effectiveness and efficacy on cardiovascular-related vital signs and disease risk by adding lifestyle management (exercise, diet, etc.) factors (variables).

In addition, with the growth of mobile health (mHealth) technology, it is possible to manage smoking, lack of physical activity, and eating habits, which are important risk factors for cardiovascular disease, by utilizing Web, APP, intelligent devices or social media, and wearable health tracking devices (Urrea et al., 2015). Therefore, it was suggested that the development of lifestyle improvement contents (digital treatment) is possible. Based on individual risk factors and predictive models, doctors and patients can help establish optimal personalized treatment plans and provide personalized treatment tailored to each individual's special circumstances and needs.

However, it is difficult to generalize the results of this study because the subject's data vary greatly for each variable. Therefore, since performance evaluation and comparison with prediction models of other algorithms have not been made, additional external verification and research

based on other machine learning models should be conducted in the future (Kim & Cho, 2023).

In future studies, we will continue to develop and improve the predictive model by increasing the accuracy of data by utilizing bio-signal data collected in clinical trials for patients with cardiovascular disease and by applying other types of algorithms other than those discussed in this study.

References

- Beak, S. K., Park, J. H., Kang, S. H., & Park, H. J. (2018). A study on the development of severity-adjusted mortality prediction model for discharged patient with acute stroke using machine learning. *Journal of Korea Academia-Industrial cooperation Society*, 19(11), 126-136. <https://doi.org/10.5762/KAIS.2018.19.11.126>
- Bosma, H., Peter, R., Siegrist, J., & Marmot, M. (1998). Two alternative job stress models and the risk of coronary heart disease. *American Journal of Public Health*, 88(1), 68-74.
- Choi, B. G., Rha, S. W., Kim, S. W., Kang, J. H., Park, J. Y., & Noh, Y. K. (2019). Machine Learning for the Prediction of New-Onset Diabetes Mellitus during 5-Year Follow-up in Non-Diabetic Patients with Cardiovascular Risk. *Yonsei Med J*, 60(2), 191-199. <http://doi.org/10.3349/ymj.2019.60.2.191>
- Cho, S. O., (2019). Change Pattern of Heart Age in Korean Population Using Heart Age Predictor of Framingham Heart Study. *Journal of the Korea Academia-Industrial cooperation Society*, 20(8), 331-343. <https://doi.org/10.5762/KAIS.2019.20.8.331>
- Hong, S. M., Byeon, H. W., Kim, J. S., & Mun, S. H. (2015). Development of Risk Prediction Model for Cardiovascular Disease Among Community-Dwelling Elderly. *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, 5(1), 37-46. <http://doi.org/10.14257/AJMAHS.2015.02.02>
- Hyun, J. K. (2022). Prediction of Diabetic Neuropathy Using Machine Learning Techniques. *J Korean Diabetes*, 23(4), 238-244. <https://doi.org/10.4093/jkd.2022.23.4.238>
- Jee, S. H., Song, J. W., Cho, H. L., Kim, S. Y., Jang, Y. S., & Kim, J. H. (2004). Development of Individualized Health Risk Appraisal Model of Ischemic Heart Disease Risk in Korea. *Journal of Lipid and Atherosclerosis*, 14(2), 153-168.
- Kang, S. A., Kim, S. H., & Ryu, M. H. (2022). Analysis of Hypertension Risk Factors by Life Cycle Based on Machine Learning. *Journal of the Korea Industrial Information Systems Research*, 27(5), 73-82. <https://doi.org/10.9723/jksii.2022.27.5.073>
- Kim, C. J., & Kim, J. S. (2018). A Study of Heart Disease Prediction Using Multilayer Perceptron based on Deep Learning. *Journal of Knowledge Information Technology and Systems(JKITS)*, 14(4), 411-419. <https://doi.org/10.34163/jkits.2018.13.4.001>
- Kim, S. H., & Cho, S. H. (2023). A Comparative Study of Predictive Factors for Hypertension using Logistic Regression Analysis and Decision Tree Analysis. *Phys Ther Rehabil Sci*, 12(2), 80-91. <https://doi.org/10.14474/ptrs.2023.12.2.80>
- Kipp, W. J., Soto, J. T., Glicksberg, B. S., Shameer, K., Miotto, R.

- Ali, M., & Joe. T. Dudley. (2018). Artificial Intelligence in Cardiology. *Journal of the American College of Cardiology(JACC)*, 71(23), 2668-2679.
<https://doi.org/10.1016/j.jacc.2018.03.521>
- Lee, Y. N., Lee, K. H., & Cho, W. S. (2021). Cost-Sensitive Learning for Cardio-Cerebrovascular Disease Risk Prediction. *The Korea Journal of BigData*, 6(2), 161-168.
<https://doi.org/10.36498/kbigdt.2021.6.2.161>
- Mohd Faizal, A.S., Thevarajah, T. M., Khor, S. M., & Chang, S. W. (2021). A review of risk prediction models in cardiovascular disease, conventional approach vs. artificial intelligent approach. *Computer methods and Programs in Biomedicine*, 207, 1-11.
<https://doi.org/10.1016/j.cmpb.2021.106190>
- Motwani, M., Dey D., Berman., Germano G., Achenbach S., & Al-Mallah M.H., Andreini, D., Budoff, M. J., et al. (2017). Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *European Heart Journal*, 38, 500-507. <http://doi.org/10.1093/eurheartj/ehw188>
- Noh, M. J., & Park, S. C. (2020). A Prediction Model of Asthma Diseases in Teenagers Using Artificial Intelligence Models. *J. Inf. Technol. Appl. Manag.*, 27(6), 171-180.
<http://doi.org/10.21219/jitam.2020.27.6.171>
- Oh, T. S., Kim, D. K., Won, C. W., Kim, S. Y., Jeong, E. J., Yang, J. S., Yu, J. H., Kim, B. S., & Lee, J. H. (2022). A Machine-Learning-Based Risk Factor Analysis for Hypertension; Korea National Health and Nutrition Examination Survey 20016-2019. *Korean j Fam Practice*, 12(3),173-178.
<http://doi.org/10.21215/kifp.2022.12.3.17.3>
- Park, D. W. (2023). Analysis of the Application Target and Scope of Artificial Intelligence Algorithm for the Development of 4th and 5th Industrial Revolution Technologies. *The Journal of Korean Institute of Communications and Information Sciences*, 48(1), 65-73. <http://doi.org/10.7840/kics.2023.48.1.65>
- Park, P. W., Kim, M. K., Lim, H. S., Yoon, D. Y., & Lee, S. W. (2014). A Comparative Study of Machine Learning Algorithms for Diagnosis of Ischemic Heart Disease. *Journal of KIISE*, 45(4), 376-389.
<http://doi.org/10.5026/JOK.2018.45.4.376>
- Urrea, B., Misra S., Plante, T. B., Kelli, H. M., Misra, S., Blaha. M. J., Martin S. S. (2015). Mobile Health Initiatives to Improve Outcomes in Primary Prevention of Cardiovascular Disease. *Curr Treat Options Cardio Med*, 17, 59.
<https://doi.org/10.1007/s11936-015-0417-7>
- Venkatesh, B. A., Yang, X., Wu, C. O., Kiu, K., & Hundley, G., et.al. (2017). Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circ Res.*, 121(9), 1092-1101.
<https://doi.org/10.1161/CIRCRESAHA.117.311312>