ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 12 (2022) No. 3, pp. 345-360
DOI: 10.33168/JSMS.2022.0317

A Study on Digital Healthcare Service in Big Data Environment: Focusing on Diagnosis of Hyperlipidemia Based on Diagnostic Testing

Jai-Woo Oh

Department of Health Management & Education, Kyungdong University, Korea

sbaby692001@naver.com

Abstract. The purpose of this study was to develop a hyperlipidemia prediction model by analyzing diagnostic test results and using machine learning techniques. To construct a prediction model, Excel was utilized for data analysis, the processed data was examined using the CatBoost algorithm, and a decision tree analysis algorithm using test factors and demographic components was applied. 69,687 pieces of valid data consisting of twenty-three diagnostic test factors and two demographic factors were obtained and applied to a platform called Hello Datta for big data analysis and machine learning. Feature extraction was performed through dimensional reduction for data normalization, and the pre-annotation function was used for data preprocessing. Following the annotation process, the data were segregated into learning data for the development and application of predictive models, and test data for the predictive model's performance evaluation. A random forest analysis algorithm was used to derive the test items that were important in disease prediction and to analyze the accuracy of the prediction. In the case of performance analysis, the accuracy, ROC area, confusion matrix, precision, and recall were analyzed. First, the model for predicting the presence of hyperlipidemia that was developed in this study was found to have a high prediction accuracy that surpassed 85% and a large AUC. Second, a cluster analysis of the test results that were positive (34%) and negative (66%) for hyperlipidemia showed that the collected data were useful in the predictive model research. Third, in the case of hyperlipidemia, which is strongly associated with blood circulation, it was deemed possible to use the developed model predict other similar diseases. Fourth, it was determined that if the data structure was altered to a format that displayed the test results for time series analysis and the time of disease diagnosis, useful findings might be produced. The diagnosis prediction model developed in this study, which uses big data analysis and machine learning technology, is expected to lead to a digital healthcare service model that combines healthcare data and artificial intelligence technology to reduce diagnostic errors in medical settings and provide prompt disease diagnosis for improved healthcare services.

Keywords: Digital healthcare service, big data analysis, machine learning, hyperlipidemia diagnosis, diagnostic test

1. Introduction

Healthcare, which should be equally enjoyed by all people, is a type of public service that cannot depend solely on the private sector or market principles. As a result of continuous development and industrialization, healthcare services have grown more abundant than ever, and nations burdened by rising medical expenses due to rapid population aging are looking to incorporate information and communications technology (ICT) into healthcare services to meet the growing demand for improved healthcare. Digital healthcare services are expected to become available in a wide range of products and services incorporated with innovative ICT for various applications in real life, and it will be possible for consumers to choose what suits their needs with ease. For this reason, digital healthcare service providers are taking on a strategic approach to service quality design so that consumers striving for a healthier life can use and purchase their digital healthcare services with brand loyalty. Accordingly, medical institutions are becoming increasingly interested in ICT-based healthcare services to implement a disease prevention strategy by diagnosing, treating, and managing diseases early on to prevent progression toward premature death, disability, or serious complications (Cho 2013). Apple's digital healthcare services development strategy was to create an application program with improved accuracy to increase user involvement and make the app easier to use. In other words, it was suggested that the development of digital healthcare services could revolutionize the interaction between patients and healthcare providers and that consumers could participate in the production of healthcare services in unprecedented ways in the history of medicine. Digital healthcare services will develop into products and services in various forms for application in real life with a combination of innovative ICT that will meet user needs. Such digital healthcare services are projected to be provided in the form of a variety of products and services that customers can choose from in a similar way to online shopping. There is currently a lack of relevant research from the perspective of digital healthcare services, so this study was conducted to classify the technical characteristics of technical service providers, user characteristics, and characteristics of the services provided by healthcare service providers among the stakeholders in digital healthcare services. This is in consideration of the need to develop various types of healthcare services to accommodate the demand for improved health through disease prevention and early detection of diseases (Kim 2010).

The goal of this study was to propose a digital healthcare service for health screening institutions or small and medium-sized hospitals that would improve the accuracy of diagnostic test results and check for inaccuracies in medical specialist diagnosis results. With this, a big data analysis-based hyperlipidemia test was proposed to address the issue of result judgment delays when performed by a specialist using diagnostic test findings and when the number of patients whose results must be assessed is greater. By verifying the significance of digital healthcare services and ICT-based healthcare services in a big data environment in the field of health data integration platform, which has become a key part of the digital healthcare industry, various development possibilities were confirmed for the digital healthcare industry.

2. Theoretical Considerations

2.1. Concept of digital healthcare service

Digital healthcare refers to the convergence of biotechnology (BT) and information and communications technology (ICT) for personalized health and disease management. With a paradigm shift in the digital healthcare industry, the term, "digital healthcare," is being used interchangeably with various other terms such as uHealth, smart health (s-Health), and mobile health (mHealth). uHealth refers to a healthcare service where wireless communication technology is applied to provide services without temporal or spatial constraints (Martínez et al., 2008; Jeon 2011). Smart health (sHealth) service is aimed at helping users manage their daily activities using wearable devices for personal health management (Catarinucci et al., 2015). Mobile health (mHealth), which started to undergo dramatic growth in 2013, typically refers to services where mobile devices such as wearables and smartphones are used as tools for healthcare (Charalampos 2010). The convergence of ICT and BT first lead to the development of uHealth, and as tech companies such as Google, Apple, Microsoft, and Samsung began to invest in the healthcare sector, smart health and mobile health services emerged. Digital healthcare encompasses all these services from the treatment-centered uHealth to smart health and mobile, the focus of which is on health management (Volk et al., 2015; Lee 2015, Zaragoza et al., 2017). The ICT applied goes beyond wearable and includes applications through which Electronic Health Records (EHR) can be accessed via smartphones and services that allow users to easily access their own health electronic data with the help of technology providers, such as Apple, and even engage in the provision of medical data. This has led to opportunities for the rise of new healthcare services to improve health. Studies on digital healthcare services have noted that digital healthcare is a "comprehensive national policy to ensure appropriate, sustainable, routine, and safe use of digital health technologies within the extensive domain of national health" (Frost et al., 2018) and defined "digital healthcare" as a new tool to provide patients with medical services and treatment results such as diagnosis and diagnostic results at point of care (POC) and to be used for work planning training and management for clinical decision-making and as a service provided by digitizing health system functions used routinely such as HER (Labrique 2018). Meanwhile, the Korea Creative Economy Research Network (KCERN) defined digital healthcare as "data-based healthcare that is connected by the Internet and made intelligent by artificial intelligence. It overcomes the constraints of reliability as an intelligence-based health management service connected by data by connecting time-based on standardized product compatibility of electronic medical records (EMR) and data among different hospitals, connecting spaces for telemedicine and telehealth management, and connecting social networks to promote collective intelligence and patient health, through which human connections are made based on data. It involves utilizing things and human data through wired and wireless Internet, and knowledge is structuralized by artificial intelligence based on data to overcome the limitations of professional expertise. As such, it is the future of healthcare that involves not only simple data utilization but also big data analysis and deep learning. The key element driving innovation in digital healthcare is data, and "digital healthcare service is an industrial area where personal health and disease management is pursued based on the convergence of healthcare industry and IC, and changes are brought across the healthcare industry, including medical care and health management, through the process of measuring, integrating, analyzing, and utilizing healthcare data" (Lee and Kim 2018).

In this study, "digital healthcare service" was defined as the "future of healthcare service for personal health and disease management in which healthcare and medical big data are combined with cutting-edge ICT and is aimed at providing innovative integrated services for personalized healthcare and healthcare industry innovation based on clinical data integration and integrated bio big data.

2.2. Concept of Big Data Analysis

The biggest change in the area of data in the era of the Fourth Industrial Revolution is the emergence of big data. Data science, which is the basis of big data analysis, is an academic discipline that deals with data analysis methods without any academic boundaries and numerous concepts such as artificial intelligence, machine learning, and deep learning are mentioned without distinction (Kim 2019). While data science is the study of how to perform analysis using data, artificial intelligence could be described as a means to allow computers to perform learning and thinking just like humans with intellectual abilities (Weiyu and Siau 2019). Machine learning is focused on methods of inputting data to apply learning logic to computers so that they can learn on their own (Weiyu and Siau, 2019). On the other hand, deep learning, which has recently garnered attention in the area of big data analysis, is more advanced than machine learning, and it allows computers to learn the patterns apparent in data on their own, without any input of data to be learned, to make predictions or decisions. Big data analysis uses a bottom-up inferential approach, which is an exploratory analysis method, rather than a top-down method of verifying a hypothesis-based cause-and-effect relationship, which is mainly used in the social sciences (Chen and Wojcik 2016). Big data analysis methods include text analysis, supervised learning, and unsupervised learning. Text analysis is widely applied across various fields such as the humanities, computer science, and computational linguistics (Weiss et al., 2010). Supervised learning and unsupervised learning apply data mining, which involves finding associations and patterns among the components in data to predict useful results. Supervised learning is to train algorithms on a training data set when the outcome of the target variable is known. After learning the relationship between a series of feature variables and the target outcome variable from the data, the goal is to characterize their association. Since the outcome is already known, the algorithm can be used to infer the most appropriate model using the data. The aim of such supervised learning is to build a model that can be used with future data to predict outcomes that are yet unknown. Unsupervised learning, on the other hand, is aimed at building a technical model for discovering interesting structures or patterns in data in which there are no target outcome variables indicated in the data, no information about the outcomes, and no given target outcome variables. Unsupervised learning is an exploratory method where the outcomes are generated from the data based on analysis. Examples of unsupervised learning include clustering and association rule learning, and in terms of target data, there is multimedia analysis that deals with a wide range of data formats such as audio, image, and video as well as network analysis where complex relationship patterns are summarized after a network calculation (Chen and Wojcik 2016).

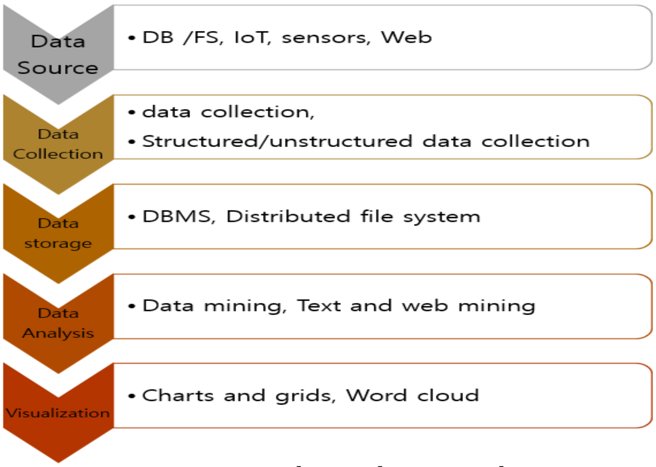


Figure 1. Big data analysis procedure

In this study, “big data analysis” was defined as "next-generation technology and architecture designed to extract necessary value in an economical way through high-speed capture, data exploration, and analysis from a large volume of data composed of various data.” As shown in [Figure 1], the big data analysis process includes a data collection step, a data storage step, a data analysis step that involves text mining, machine learning, etc., and a reporting and visualization step in which the results are communicated.

2.3. Concept of machine learning

Machine learning is a new technology and a method of implementing artificial intelligence where algorithms and programs are applied for a computer to learn data in order to make decisions or predictions about what will happen in the future. This refers to a series of processing processes in which the computer learns patterns from data and performs appropriate actions based on newly input data (Oh 2017).

Artificial intelligence, which is a comprehensive concept of machine learning, emerged with the application of neural network algorithms in the 1950s. Machine learning algorithms were first explored in the 1980s, and with the growth of hardware, machine learning started to be used in everyday life, albeit in a restricted way. Recently, the development of algorithms for deep learning, which is more advanced than machine learning, has led to a wider use of big data and the development of broader applications of machine learning. There has also been a growing interest in the application of machine learning technology. Machine learning algorithms are classified into supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning is a learning method that uses data with an output value corresponding to an input value, and it is implemented to teach the computer the correct answer. The aim is to predict the output value corresponding to the unknown input value using the model learned through supervised learning. In the case of supervised learning algorithms, learning is carried out by reducing the error in the value of the cost function or the predicted output value by using classification or regression to predict a specific value (Shin 2018). Key methodologies include random forest, support vector machine, boosting, artificial neural network, and deep neural network. On the other hand, an unsupervised learning algorithm is a type of grouping algorithm that groups similar data without giving a correct answer as there is no label, and it is more difficult than supervised learning as shapes or patterns are sought in unlabeled data. It is mainly applied to dimensional reduction or clustering as a way to find the shape, pattern, and relationship of data with the computer learning the data on its own without any outcome value (Shin 2018). The difference from supervised learning is that when there is given data, it directly guesses the nature of the data instead of creating a function to calculate a specific value. Main examples of unsupervised learning algorithms include principal component analysis and k-means clustering. Meanwhile, reinforcement learning, which could be described as the epitome of machine learning, is a slightly different concept from supervised learning and unsupervised learning. Reinforcement learning is a method in which the agent performing the action recognizes the current state and selects the action with the maximum reward among actions that can be chosen. There are no data that can be classified, and even if there are data, there is no correct answer. Instead, learning occurs based on the rewards received for one’s actions. This is a method in which the agent experiences a series of interactions among observation, action, and reward and learns by trial and error, as is the case with the way humans acquire knowledge, and it is mainly used for games, AI, and robot navigation. (Dy 2019). Reinforcement learning is a learning method that has long existed, but the algorithms in the past failed to produce results that were good enough to be applied in real life. However, since the introduction of deep learning, the application of neural networks to reinforcement learning has enabled its application to complex problems such as Go and autonomous vehicles. Good examples of successful application of deep learning to reinforcement learning are DQN and A3C.

In this study, an empirical analysis was used to supervised learning algorithms such as support vector machines, tree-based ensemble methods, and deep learning.

2.4. Hyperlipidemia diagnosis method

Hyperlipidemia is a condition in which there are more lipids in the blood than normal. Although the fat content in the blood itself does not directly cause diseases, if the condition persists, it increases the risk of stroke, myocardial infarction, angina pectoris, and other cardiovascular diseases. Of the several types of lipids found in the blood, cholesterol is often elevated above normal among those who have hyperlipidemia, and this condition is called hypercholesterolemia. Hypercholesterolemia is a key factor that causes not only arteriosclerosis and ischemic heart disease but also various other diseases. While cholesterol needs to be present in the human body at moderate levels for the synthesis of cells that make up the body as well as hormones, when it is present in excess, it sticks to the blood vessels resulting in obstruction, and can become the main cause of fatal diseases such as myocardial infarction and cerebral infarction due to the formation of blood clots. There are various causes of hyperlipidemia, but it is known that the condition, which mainly occurs in the elderly, is primarily caused by Westernized diet and lifestyle factors. Hyperlipidemia patients can largely be divided into familial hyperlipidemia patients and acquired hyperlipidemia patients. Recently, hyperlipidemia among young people has been on the rise, and most of these cases are caused by lifestyle choices. About 30% of cholesterol is synthesized from the food we eat, and 70% is produced in the liver. Cholesterol is carried in the bloodstream as lipoproteins, which are harmless in small amounts. However, when cholesterol levels rise owing to liver disease or obesity caused by excessive food intake, the lipoproteins block the blood vessels, resulting in hyperlipidemia. Hyperlipidemia, which is called the silent killer, is difficult to self-detect because it does not cause any immediate pain or noticeable symptoms, even if fat and bad cholesterol build up in the blood vessels. However, as cholesterol accumulates, there is narrowing of the blood vessels, leading to an onset of related diseases. Therefore, it is important to manage the cholesterol levels with regular health screenings and exercise. Hyperlipidemia symptoms include a significant rise in body weight, but accurate diagnosis requires regular diagnostic tests.

3. Research Method

The aim of this study was to develop a diagnosis and prediction model for hyperlipidemia by analyzing the diagnostic test results using big data analysis techniques. Hence, collection and supplementation of high-quality data, preprocessing of analytical data, predictive model development, and validity evaluation were carried out. The details of the steps carried out are as follows: First, 69,800 diagnostic test results were gathered after Institutional Review Board (IRB) approval for data collection and supplementation, and they were classified into positive test results for hyperlipidemia and normal test results. In order to evaluate the degree of impact among the diagnostic test items, data labeling was performed to rearrange the tests into 25 items. Null data among the test factors were corrected by applying the average value for each sex. Second, in the case of data preprocessing, the test factor data were normalized to judge and predict the results of 23 diagnostic tests. Considering the fact that the diagnostic test results from different medical institutions vary, feature extraction was performed through dimensional reduction, which involves determining the data to be used as primary data and secondary data for each disease. Third, in the data generation process, whether there is a presence of hyperlipidemia was corrected by inputting the data from the data preprocessing step into the data processing platform, Hello Data. In this step, data were quickly generated according to the target through the pre-annotation and data analysis functions provided on Hello Data. The fourth step was the data augmentation process. Since a large amount of data is required for machine learning algorithm, a distribution function was set based on the data obtained during the data generation process. In this step, there is a need to augment the data so that the machine learning algorithm is not affected, even if there are no secondary diagnostic test results, by referencing the feature extraction results from the data preprocessing process. Thus, sufficient data was obtained to train the machine learning algorithm as quickly as possible by forming a distribution function for each of the 23 diagnostic test results and generating tens of thousands of virtual data to maintain the correlation between these distributions and their independence. The fifth step involved developing and training machine learning algorithms. The data and system development necessary for this study was completed by finding the structure of a machine learning program suitable for the data and, conducting training and then repeating the entire process by using the results as feedback.

Excel was used for data analysis, the processed data were analyzed using CatBoost algorithm, and a decision tree analysis algorithm with test factors and demographic factors was applied to develop a predictive model. For predictive model performance analysis, TP rate, FP rate, accuracy, precision, and receiver operating characteristic (ROC) area indicators were used to measure and analyze the performance of each predictive model. In order to measure and analyze the performance of the machine learning prediction model, it is necessary to understand the meaning of true positive (TP), false positive (FP), false negative (FN), and true negative (TN), which were used to analyze accuracy, recall, precision, F-measure, and the ROC area to measure performance.

3.1. Analysis of quality control consistency prediction model

Positive cases of hyperlipidemia accounted for 32% of all cases in the dataset, and its proportion was found to be very useful in designing the model. The data were deemed appropriate for use in developing the predictive model, and the decision tree analysis clearly revealed major variables, making it possible to create a model as shown in [Figure 2].

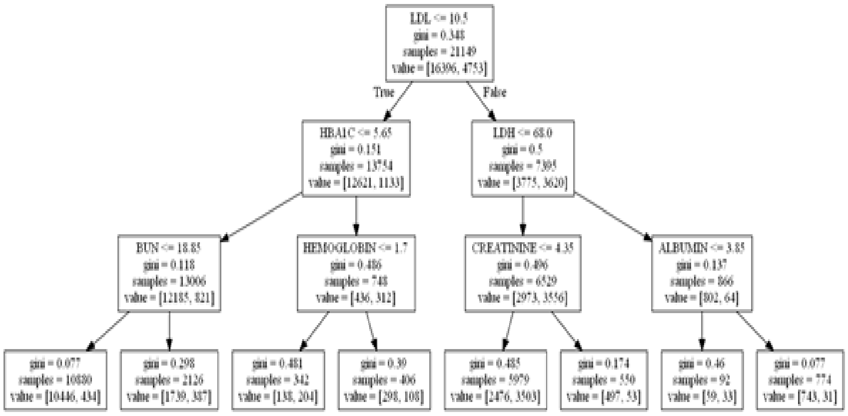


Fig. 2: Decision tree for the hyperlipidemia prediction model

The results of analyzing the importance of 25 factors affecting hyperlipidemia by using the random forest algorithm are shown in [Figure 3]. The accuracy of predicting hyperlipidemia was found to be 0.87, and the important features were analyzed to be LDL, BUN, and Hba1c out of 23 test factors. The demographic factor that had the highest level of importance was age, followed by sex. The LD and BUN results above a certain level indicated the presence of hyperlipidemia.

Second, the loss and accuracy were analyzed, and the result was found to be 0.77. The graph of the results is shown in [Figure 4]. As a result of analyzing the ROC curve, which is a method of evaluating the performance of the binary classifier system, the AUC was found to be 0.87, as shown in [Figure 5], indicating that high predictive power.

```
Accuracy is:  0.8681581097216112
              precision    recall  f1-score   support

         0.0       0.90      0.93      0.91     11731
         1.0       0.78      0.70      0.74      4182

    accuracy                           0.87     15913
   macro avg       0.84      0.81      0.82     15913
weighted avg       0.87      0.87      0.87     15913
```

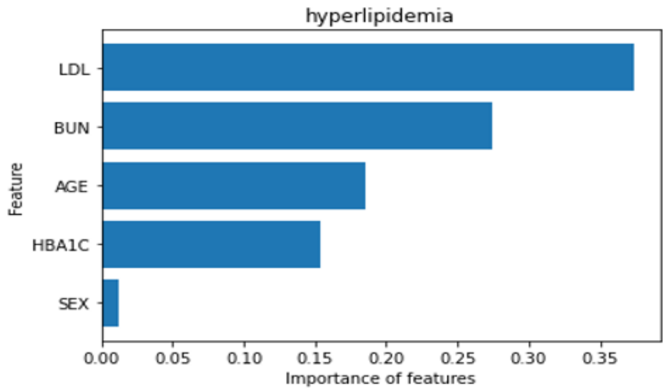
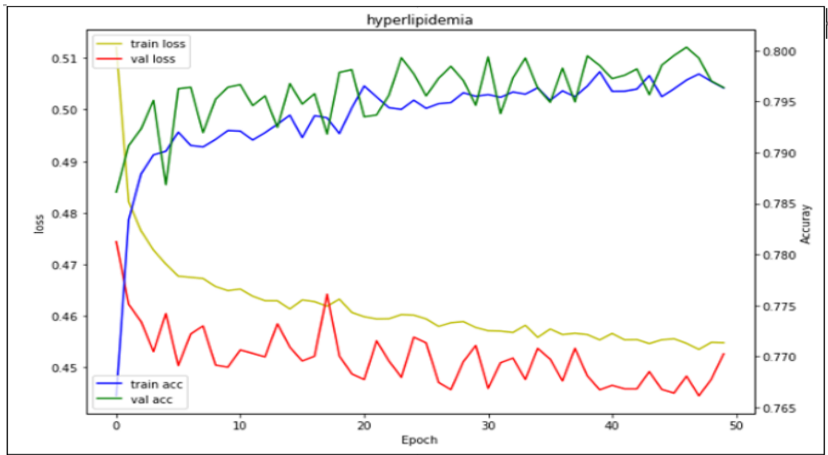


Fig. 3: Random forest analysis results




```
1  model_hyperlipidemia.evaluate(X_test_hyperlipidemia,Y_test_hyperlipidemia_categ)
15913/15913 [==============================] - 1s 62us/step
[0.5158775770813292, 0.7732671400816445]
```

Figure 4. Loss and accuracy of the hyperlipidemia prediction model

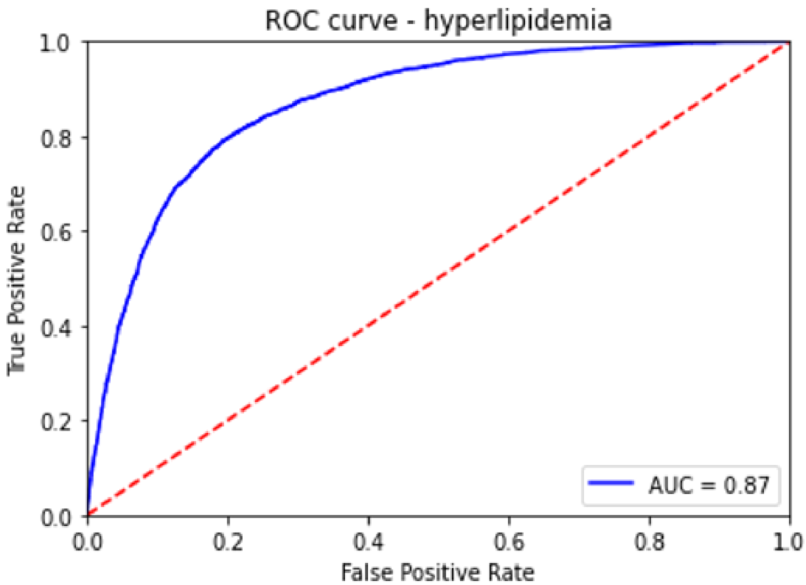


Fig. 5: ROC curve for the hyperlipidemia prediction model

The results of the confusion matrix analysis are shown in [Figure 6], with a TP rate of 0.83, FP rate of 0.25, TN rate of 0.75, and FN rate of 0.17. Precision and recall were 76.8% and 83%, respectively. As such, precision was deemed low and recall satisfactory.

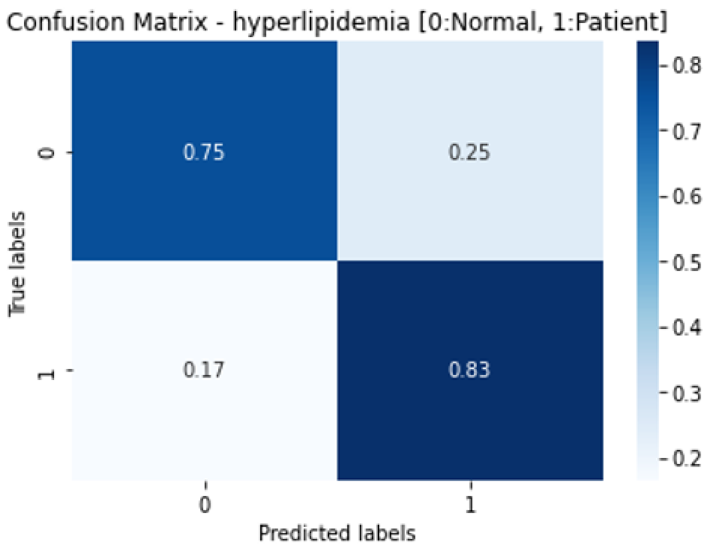


Fig. 6: Confusion matrix for the hyperlipidemia prediction model

3.2. Result

The indicators for determining the accuracy of the hyperlipidemia predictions made based on diagnostic tests were accuracy (%), ROC curve (%), sensitivity (%), specificity (%), precision (%), and F1 score (%). The evaluation results on the pre-separated test set not used in the modeling set were applied to judge the performance of the predictive model. The standard deviation of the errors was calculated based on 10-fold cross-validation to determine the possibility of difference due to the sample. The CatBoost boosting algorithm and random forest algorithm were used to process 23 test factors and demographic factors, such as sex and age, to analyze the hyperlipidemia prediction model, and the pre-annotation and data analysis functions provided by Hello Data, a data processing tool, were used to obtain sufficient data to train the machine learning algorithm. The model was developed by repeating the process of finding the structure of a machine learning model suitable for the data in question, training the model, and providing the results as feedback, and it may be used as a baseline model for future model development.

The predictive model developed was analyzed to determine the accuracy (87%), ROC curve (87%), sensitivity (83%), specificity (75%), precision (76.8%), F1 score (80%), and recall (83%). Although there is yet to be an objective international numerical standard for accuracy, prior studies have reported that a predictive model with an accuracy of 85% to 90% has practical applications to commercial platforms. As a result, the hyperlipidemia prediction model developed in this study was found to be 87 percent accurate and could be applied to a commercial platform.

4. Conclusion

There has been a recent paradigm shift in digital healthcare services from treatment to prevention based on the convergence of healthcare and ICT for real-time and rapid results management. Medical institutions have been striving to upgrade their diagnostic testing systems, which could be said to be the most data-intensive of all systems used by medical institutions, in order to meet diverse demands for health management, despite the growing interest in applying artificial intelligence-based on healthcare data. The direction of healthcare service development in the era of the Fourth Industrial Revolution is primarily geared toward the fields where healthcare data are combined with ICT such as ICBM (IoT, Cloud, Big Data, Mobile) and artificial intelligence. This study was carried out to examine the possibility of applying big data-based digital healthcare services for the diagnosis of hyperlipidemia by using diagnostic test results. In order to develop the modules necessary for judging and predicting the presence of a disease using machine learning, a platform called Hello Data was used. For the development of a predictive model and evaluation of its performance, the data were divided into training data and test data. A random forest analysis algorithm was employed for the predictive model, and associated variables such as accuracy, Receiver Operating Characteristic (ROC) area, confusion matrix, precision, and recall were studied for the predictive model performance analysis.

The accuracy of the model in predicting the presence of hyperlipidemia was at least 87%, based on which it was judged to be appropriate for application as a commercial platform model. It was found that the diagnostic test factors that were most important in the disease prediction were LDL, BUN, and Hba1c, and the demographic factor with the strongest influence was age, followed by sex. Just as various test results are taken into consideration to make a diagnosis in clinical settings, multiple test results were used to make the predictions, instead of using just a single test result, in this study. This study enabled an analysis of the impact of the test items on disease prediction, which will be beneficial in future studies in which additional tests are conducted and the key test items for each disease type are examined. For the purpose of analyzing the performance of the predictive model of the learning module that was developed in this study, an ROC curve analysis was carried out, and the AUC was found to be 87%, indicating that the predictive model had high utility. Also, in order to analyze the prediction performance after learning, a confusion matrix analysis was performed, and the precision was analyzed to be 0.75, indicating that the precision was too low to be used as a prediction model. The recall rate was analyzed to be 0.83, which was deemed appropriate for use as a hyperlipidemia prediction model. However, the false negative rate and the false positive rate were found to be around 17% and 25%, respectively, and thus, it is judged that additional research is required to lower these rates by supplementary preprocessing of collected data and repeating learning. As for the implications of this study, it is judged that the learning model developed in this study could be expanded in terms of application to predict other diseases. This could potentially lead to a new healthcare service model that can enhance the reliability and productivity of health screenings by improving the health screening environment and addressing health screening issues. In the case of chronic diseases, there is a strong correlation between highly similar diseases in terms of onset, so it may be possible to diagnose multiple chronic diseases at the same time by changing the collected data. One limitation of this study was that the time lag between health screening and disease diagnosis was difficult to reflect in the data collected for the study, which impeded the improvement of precision and recall rate.

Follow-up research is expected to improve accuracy, AUC, precision, and recall by analyzing the time series data on the collected data and supplementing the learning module for predicting multiple diseases.

Acknowledgment

This research was supported by Kyungdong University Research Fund, 2020.

References

Charalampos Doukas, Thomas Pliakas, Ilias Maglogiannis. (2010). Mobile healthcare information management utilizing Cloud Computing and Android OS. *Annu Int Conf IEEE Eng Med Biol Soc*, 2010, 1037-1040. DOI: 10.1109/IEMBS.2010.5628061

Cho, B. R., & Ann, E. M. (2013). Present Status and Problems of Health Screening Program in Korea. *Health and welfare policy forum*. 198, 48-54. http://repository.kihasa.re.kr/handle /201002/10004

Eric Evan Chen and Sean P Wojcik. (2016). A practical guide to big data research in psychology. *Psychological methods*, 21(4), 458-474. doi:10.1037/met0000111

Martínez, I., Escayola, J., Martínez-Espronceda, M., Serrano Senior, L., Trigo, J., Led, S. & García, J. (2008). Standard-based middleware platform for medical sensor networks and u-health. *Proceedings of the 17th International Conference on Computer Communications and Networks, IEEE ICCCN 2008*, August 2008: 1-6. DOI:10.1109/ICCCN.2008.ECP.135

Jeon, H. A. (2011). Study on the Smart Healthcare Definition and Market Trends. *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*. 2011(6): 841-842. http://www.dbpia.co.kr/journal/article Detail?nodeId=NODE02181960

Kim, C. T. (2019). Studying Psychology using Big Data. *The Korean Psychological Association*, 38(4): 519-548. http://www.dbpia.co.kr/journal/articleDetail?No deId=NODE09292062

Kim, H. S. (2010). A study on the efficient policy of health examination based on comparing private health sector with public health sector [dissertation]. [Seoul]: Kyunghee University, http://khu.dcollection.net/public_resource/pdf/2000000605 50 _20220309214722.pdf

Labrique, A. B, Vasudevan, L., Kochi, E., Fabricant, R., Mehl, G. (2018). mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. *Global Health*. 1(2): 160-171. https://www.globalhealthlearning. org/course/mhealth-innovations-health-system-strengthening-tools-12

Lee, B. K. (2015). Key to opening up smart healthcare market, mobile medical devices. *The optical journal*. 156(2015), 41 - 51. https://scienceon.kisti.re.kr/srch/ selectPORSrchArticle.do?cn=JAK O201554753131 410&SITE=CLICK

Lee, D. E & Kim, S. K. (2018). Digital Healthcare Innovation trends and policy implications. *Science and Technology Policy Institute*. 48, 1-31. http://www.dbpia.co.kr/journal/ articleDetail?nodeId=NODE 07464623

Luca Catarinucci, Danilo de Donno, Luca Mainetti, Luca Palano, Luigi Patrono, Maria Laura Stefanizzi, Luciano Tarricone. (2015). An IoT-Aware Architecture for Smart Healthcare Systems. *IEEE Internet of Things Journal*. 2(6), 515-526. DOI:10.1109/JIOT.2015.2417684

Michael J Frost, Jacqueline B Tran, Fatema Khatun, Ingrid K Friberg, Daniela C Rodríguez. (2018). What Does It Take to Be an Effective National Steward of Digital Health Integration for Health Systems Strengthening in Low- and Middle-Income Countries? *Global Health Science and Practice*. 6(1): 18-28. doi:10.9745/GHSP-D-18-00270

Mojca Volk, Janez Sterle, Urban Sedlar. (2015). Safety and Privacy Considerations for Mobile Application Design in Digital Healthcare. *International Journal of Distributed Sensor Networks*. 11(10), 1. https://doi. org/10.1155/2015/549420

Na, D. Y. (2019). Machine learning algorithm and IoT technique research for animal welfare smart farm [dissertation]. [Seoul]: Konkuk University, http://kku.dcollection.net/public_resource/pdf/200000175642_20220309235826.pdf

Park, J. –W. & Hwang, B. -D. (2019). Analysis of Factors Affecting Catastrophic Healthcare Expenditure. *International Journal of IT-based Public Health Management*. Global Vision Press. 6(1), 35-40. doi:10.21742/IJIPHM.2019.6.1.06.

Oh, M. A., Choi, H. S., Kim, S. H., Jang, J. H., Jin, J. H., Chun, M. G. (2017). A study on social security big data analysis and prediction model based on machine learning. Sejong City: Korea Institute for Health and Social Affairs, http://repository.kihasa.re.kr/handle/201002/29093

Weiss, S., Indurkhya, N., Zhang, T., & Damerau, F. J. (2010). Text Mining: Predictive Methods for Analyzing Unstructured Information. Springer Science & Business Media.

Zaragoza, M. G., Kim, H.-K., & Chung, Y. (2017). U-healthcare Big Data Analytics Process Control. *International Journal of Control and Automation*, NADIA, ISSN: 2005-4297 (Print); 2207-6387 (Online), 10(11), 165-174, http://dx.doi.org/10.14257/ijca.2017.10.11.15.