Kütahya Dumlupınar University
Institute of Graduate Studies

Journal of Scientific
Reports-A
E-ISSN: 2687-6167
Number 51, December 2022

RESEARCH ARTICLE

IN HEALTHCARE APPLICATIONS of MACHINE LEARNING ALGORITHMS for PREDICTION of HEART ATTACKS

Tansu SESLİER[1], Mücella ÖZBAY KARAKUŞ[2,*]

[1]Yozgat Bozok Universirty, Graduate Education Institute, Yozgat, tansu.sslr@gmail.com, ORCID: 0000-0002-8990-1512
[2]Yozgat Bozok University, Faculty of Engineering-Architectury, Department of Computer Engineering, Yozgat, mucella.karakus@yobu.edu.tr, ORCID: 0000-0003-0599-8802

Receive Date:19.10.2022 *Accepted Date: 16.11.2022*

ABSTRACT

Due to changing lifestyles in the world and in our country, the account of chronic diseases (CD) is rising day after day. CD is one of the most widespread reason of death. About 46% of the death of people in the world, excluding communicable diseases and accidents, are because of cardiovascular diseases (CVDs), according to this study, and 7.4 million of her 17.5 million deaths from these diseases are due to heart attacks. It was something. The number of deaths from cardiovascular disease is estimated to reach 22.2 million in 2030. The fact that most of the agents that are the reasons of the heart disease (HD) that can be prevented and treated is an important phenomenon in reducing cardiovascular disease deaths. Accurate and timely diagnosis of HD is therefore plenty important. Used machine learning (ML) techniques to determine heart attack risk in this study. Therefore, heart attack risk assessment was performed with a less expensive and effective approach. In this study, Logistic Regression, Support Vector Machines (SVM), Nearest Neighbor Algorithms, NaiveBayes, and Random Forest, ML techniques were applied to a data set containing 303 patient records and 14 variables. As a result of the application, the SVM technique achieved the best accuracy outcomes as 87.91%.

Keywords: *Machine Learning, Heart Attack Prediction, SVM*

1. INTRODUCTION

CVDs are involve the way of the blood is pumped and circulated throughout the body, which unfavorably affects the vessels and heart. In Cardiovascular diseases, the decisions which are wrong or delayed, are likely to cause death [1]. Coronary heart disease (CHD) is caused by narrowing of the coronary arteries, reduced blood flow to the heart, and decreased oxygen supply. The purpose of this study was to investigate the condition known as "heart attack" and the factors that cause it [2]. CVD is the leading cause of death worldwide, killing approximately 17.9 million people each year. More than 80% of cardiovascular deaths are owing to strokes and heart attacks, and and more than 30% of these deaths consist of people that under 70 years old. The World Health Organization (WHO) identifies

those at highest risk for CVD and ensures that appropriate treatment can prevent premature death [3]. When examining cardiovascular diseases, patients' age, gender, hypertension, high blood cholesterol, feeding habits, physical inactivity, smoking, alcohol use, obesity, diabetes takes into account the same qualities as CVDs can be due to various characteristics. According to the literature studies, it has been seen that reducing the risk factors that cause heart disease (HD) can actually help in the prevention of heart diseases [4].

Today, the diagnosis of CVDs is mostly made by following complex and high-priced medical processes carried out in well accoutred hospitals and health institutions. In addition, these processes often oblige the implementation of invasive procedures by highly skilled medical proficients. Although this method gives excessive diagnostic accuracy, the number of sufferers with get entry to this approach may be very restricted. Therefore, it is highly desirable to develop new, economical and easily reachable techniques for diagnosing cardiovascular diseases [5]. Convenient computerized knowledge and/or decision support systems can assist in performing clinical testing at lower cost. The efficient and correct implementation of the automated system requires a comparative study of the various available techniques [6].

It is intelligent computational process of huge knowledge sets employing a integration of machine learning, applied math analysis, and information technology. It conjointly aims to get helpful patterns and rules to guide choices concerning future actions. Machine learning can be identified as a scientific technique for planning and developing algorithms that enable computers to meet and develop reeltime problems based on historical statistics and solve real-time problems under specific instructions and rules [7].In our study, type of chest pain, ECG results, slope of ST, fasting blood glucose, sex, resting blood pressure, age, cholesterol level, resting maximum heart rate, angina that exercise-welding, ST depression that exercise-welding, defect type, and major vessel the target variable in the quantity is used in machine learning. There are many researches on the prevention of heart disease risk. More studies in the field of heart diseases will provide more opportunities to prevent heart diseases [4].

To determine coronary heart disease, a subset of variables was selected with K-Means clustering and Particle Swarm Optimization algorithms by Verma et al. [8]. In the study, a mixed model was created Logistic Regression Fuzzy by using Artificial Neural Network, unordered precept reduction and C4.5. In this model, the researchers tested on clinical data consisting of 26 variables and 335 patient records. The highest classification accuracy with 88.4% was obtained from the polynomial logistic regression algorithm [8].

Pandy et al. [9] were developed a model that predict heart disease to help healthcare proficients estimate the situation of HD based on patients' medical data. There are 14 variables. A prediction model was obtained by using J48 decision tree to classify HD according to variables. Different pruning approaches have been applied in the development of the system. An accuracy rate of 75.73% was obtained from the decision tree algorithm with reduced error pruning [9].

Kim et al. [10] tried to predict coronary heart disease. There were a total of 10 variables in the study, including 9 input variables and 1 output variable, with 748 patient records. Variables such as HDL, cholesterol, LDL, smoke, blood pressure that systole-induced, age, blood pressure that diastole-

JSR A
Journal of Scientific Reports

induced, gender and presence of diabetes were included. They are important factors commonly used to predict the CHD. 525 of 748 patients were allocated as training data and 223 as test data. Fuzzy logic was used to tide over the issues related with the instability of CHD estimation. Correct answer rate and ROC analysis were used to evaluate the constructed model. The correct answer rate for the model was 69.51%, and the ROC curve value was determined to be 0.594 [10].

Ahmed et al. [1] used five common machine learning techniques based on audited learning to detect HD early. The aim of the study was finding the selected classification method's optimal performance. For this purpose, the performances of the five classifiers were interpreted for 270 data in total in the data set by using the confusion matrix and ten-fold cross-validation method. The rate of accuracy 84% was achieved as highest by using ANN method [1].

The same dataset with this study was used by Yin et al. [11], to apply with the machine learning method. Machine learning methods are frequently used for model prediction. By analyzing the collected data, extracting valid data, and making relevant adjustments to the model, an estimator capable of making high-quality predictions is generated. A decision tree has been created that can effectively predict heart disease risk. In order to create a decision tree, by creating a new dataset by randomly selecting blocks of data, and then a technique that named "decision tree forest" was created from valuable decision trees that were created plenty of times. Then 25% of the residuary records are extracted to corroborate the predictive performance of the used method. Eventually, the estimated accuracy of our method on valid dataset is obtained as 0.857. The consequences indicate that this method is a consistent and dependable prognostic model [11].

2. MATERIALS and METHOD

2.1. Data Set

The dataset that consisting of the values of individuals with high heart attack risk and low heart attack risk taken from Kaggle [12] that was collected by Rashik Rahman. The dataset includes 303 patient records and 14 variables. The variables used were chest pain, ECG results, slope of ST, fasting blood glucose, sex, resting blood pressure, age, cholesterol level, resting maximum heart rate, angina that exercise-welding, ST depression that exercise-welding, number of major vessels, type of defect, and the target is variable. An exhaustive information about the variables is given in Table 1.

Table 1. Variables and Explanations.

#Number	Variables	Variable Types	Attribute Description
1	Age	Digital	Patient's age
2	Sex	Categorical	Patient's gender
3	Cp	Categorical	Chest pain type
4	Trestbps	Digital	Resting blood pressure
5	Chol	Digital	Cholesterol value
6	Fbs	Categorical	Fasting blood sugar

JSR A
Journal of Scientific Reports

Seslier, T. and Özbay-Karakuş, M., Journal of Scientific Reports-A, Number 51, 358-370, December 2022.

7	Restecg	Categorical	Resting electrocardiographic result
8	Thalachh	Digital	Maximum heart rate
9	Exang	Categorical	Exercise-induced angina
10	Oldpeak	Digital	Exercise-induced ST depression
11	Slope	Categorical	ST segment
12	Ca	Digital	Number of main veins
13	Thal	Categorical	Defect type
14	Output	Categorical(Target Variable)	

2.2. Experimental Methods

2.2.1. Logistic regression

The model that predicts the probability of an incident occurring by fitting the graph of data to a logistic curve is called the binary logistic regression (LR) model [13]. The dependent variable determines which independent variable(s) is an important risk factor and to what extent these independent variable(s) affect the estimation of the dependent variables' values. [14]. The achieved parameters for the expository variables can be used to predict the probability ratios for each explanatory variable in the model [13].

Special form of logistic regression model:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots\beta_p x_p}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots\beta_p x_p}} \quad (1)$$

Here; β_0, the constant is the $\beta_1, \ldots, \beta_p$ regression coefficients and the $x_1, x_2, \ldots, x_p$ independent variables. The conditional probability that the $\pi(x)$ dependent variable is equal to 1, depending on x.

$\pi(x)$ Applying the logit transform to the LR model based on, the following model is obtained.

$$\ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_p x_p \quad (2)$$

Logit transform is a linear function that can take values between -∞ and +∞ depending on the range of x.

2.2.2. Support vector machine

For regression analysis and statistical classification, a popular used technique which called “SVM” is a supervised learning method [15]. Classical learning methods aim to diminish the error ratio in the training dataset. This method is termed as empirical risk minimization (ERM). These learning techniques operate the principles of ERM, and ANNs are the most widespread model of ERM. Moreover, SVMs are grounded on the Structural Risk Minimization (SRM) principle grounded on

JSR A
Journal of Scientific Reports

"statistical learning theory". With better generalization ability, SRM is achieved by minimizing an upper bound on the generalization error [16]. It allows drawing a decision boundary on the plane of the training data that is furthest from the members of the two separate classes as shown in Figure 1.

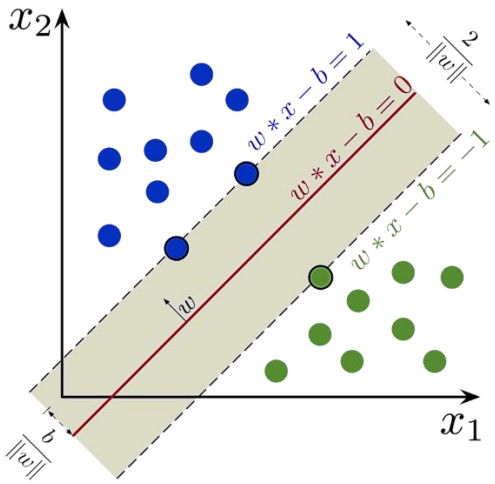

Figure 1. Margins and maximum-margin hyperplane of a trained SVM model which is trained with specimens from two separate classes [17].

Each point of the data is expressed as in Equation 3.

$$\left\{(x_i, y_i) | x_i \in R^d, y_i \in \{-1,1\}\right\}_{i=1}^{n} \tag{3}$$

Each entry in the formula, denoted by x, and y, specifies a class of data points represented by -1 or 1. Each point on the plane is wx-b indicated as a point. Here w is the normal of the multiplane and b is the trend value. The machine learning method that enables finding the separation boundary using the quadratic optimization method is the SVM [18].

2.2.3. Naive bayes

Naive Bayes strategies create the belief that everyone attributes area unit freelance. This conjecture of conditional independence is called naive because it is rarely valid in real-world applications. For this reason the algorithms characterized as naive, conversely they tend to learn quickly in many controlled classification problems [19]. They are probabilistic learning algorithms that are grounded on the theorem of Bayes.

$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)} \tag{4}$$

In this equation (4);
Y: Attributes

X: Class

P(X\Y): It shows the probability of the X event given the Y event.

P(Y\X): It shows the probability of event Y given that event X has occurred.

P(X): It shows the probability of an X event occurring.

P(Y): It indicates Y, the probability that the event will occur.

2.2.4. Random forest

Random Forest (RF) turned into stepped forward with the aid of using Leo Bremen. RF regulations can be created as supervised classification rules [20]. However, ML (machine learning) algorithm that is one supervised, random forest, is used for regression as well as for classification, it is primarily prefered for classification goal. The Random Forest algorithm works as follows: Firstly the samples are gathered from the dataset randomly, then a decision tree is generated for each sample from the existing tree, and the tree with the best predictions is selected. The greater the number of decision trees, the more accurate the prediction results [21].

The process of improving a Random Forest algorithm consists of two steps. The first stage is creating a random forest, and the second stage is building predictions over generated Random Forest classifier. Judging by the Random Forest generation Pseudocode:

a. At first selecting the "R" attributes from the whole "m" attributes that are R<<m.
b. Secondly compute knots using the best split points among the 'R' features.
c. Thirdly, separating the knot into additional knots by using the best split.
d. In fourth step, repeating this steps from a to c till the number of nodes “l” is achieved.
e. In the last step, repeating the previous steps from a to d to create forest for number “a” to create “n” which is tree number.

Finally, when the classifier that based on Random Forest is achieved, the estimation is made.

2.2.5. K-nearest neighbor algorithm

Classification is done using the proximity between a selected feature and the closest feature. The value of K included here is expressed as a number such as 3 or 5.

The formula in Equation 5 is used to determine distances between objects.

$$(i,j) = \sqrt{\sum_{k=1}^{p}(X_{ik} - X_{jk})^2} \quad (5)$$

When a new object comes to be identified, with respect to the identified data, when examining the working style, firstly, "K" is checked. The "K" number is often selected to be odd number to prevent

equality [22]. When calculating the interval between the nearest neighbor number K and data points, Manhattan, Euclidean, Minkowski or similar methods are used [23].

2.2.6. Decision trees

Decision Trees (DT), the oldest and most important type of MLA, are tree-like structures that model decision logic, test results, and categorize data items. Nodes in a DT typically have levels that multiple, with the node which is the initial or the highest is being named the "root node". For input variants and/or characteristics, all nodes that are internal nodes, express tests. According to the results of the test, in the classification algorithm, the testing and branching procedure is repeated, until the leaf reaches the node by branching at every step to the convenient lower node. Terminal or leaf nodes correspond to the outcome of decisions. DT which is a widely used supplement in medical diagnoses has proven to be easily assess and fast learning [24].

3. RESULTS

Disease diagnosis is the most important health function. It can save lives if the disease is diagnosed before the usual or planned period. Machine learning-based classification methods can support the healthcare industry by enabling rapid and reliable disease diagnosis. As heart disease is difficult to diagnose, so it's a good time for doctors and patients. We review the indicated machine learning and image fusion classification methods. The aim of the project is understanding whether a patient who has CVD or not. For information preprocessing, data processing classification techniques were applied, especially call trees and naive Thomas Bayes to the dataset which is divided to training and test sets. This section presents the results of this classification model performed at the expense of Python programming. Results are generated for each individual coaching knowledge set and check data set.

One of the frequently applied measures in the comparison of classification methods is the accuracy criterion. Generally, confusion matrix(CM) is used to define the delicacy criteria. The CM interprets the items about the work done. The confusion matrix components for the heart disease classification process examined in the study are defined as follows;
In the proposed study, various metricals are utilized to compute the performances of MLA techniques. These metrics are calculated on CM.

TP (True Positive): Data count, which is essentially true and categorized as true.

FN (False Negative): Data count, which is essentially true but categorized as false.

FP (False Positive): Data count, which is essentially false but categorized as true.

TN (True Negative): Data count, which is essentially false and categorized as false.

Table 2, 3, 4, 5, 6, 7, 8; in the following tables, 0 means "no heart attack risk", 1 means "heart attack risk".

JSR A
Journal of Scientific Reports

Table 2. Complexity Matrix.

	0	**1**
0	TP	FN
1	FP	TN

Accuracy: Usually a measure of how often classifiers are estimated correctly. The accuracy rate of the developed algorithm is calculated according to Equation 6.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{6}$$

Sensitivity (Recall): Indicates the rightly classified positive values from the positive samples' total count.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{7}$$

Precision: It shows how many of the values we estimated as positive are essentially positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{8}$$

F1 Score: Precision and sensitivity are the harmonic mean, so it's a powerful metric for calculating model performance. The precision in calculating the F-Score is represented as P and R respectively.

$$\text{F1} - \text{Score} = 2 * \frac{\text{P*R}}{\text{P+R}} \tag{8}$$

There are a total of 303 records in the study. 30% of the records were used for testing and 70% for training. The confusion matrices of the algorithms in Table 3 below:

Table 3. Complexity Matrix of LR Algorithm.

	0	**1**	**Σ**
0	31	7	38
1	5	48	53
Σ	36	55	91

Table 4. Complexity Matrix of Support Vector Machines Algorithm.

	0	**1**	**Σ**
0	31	7	38
1	4	49	53

JSR A
Journal of Scientific Reports

Σ	35	56	91

Table 5. Naive Bayesian Complexity Matrix.

	0	**1**	**Σ**
0	32	6	38
1	6	47	53
Σ	38	53	91

Table 6. Random Forest Algorithm Complexity Matrix.

	0	**1**	**Σ**
0	30	8	38
1	7	46	53
Σ	37	54	91

Table 7. Complexity Matrix of K-Nearest Neighbor Algorithm.

	0	**1**	**Σ**
0	31	7	38
1	7	46	53
Σ	38	53	91

Table 8. Decision Tree Algorithm Complexity Matrix.

	0	**1**	**Σ**
0	28	10	38
1	12	41	53
Σ	40	51	91

Accuracy (Ac), Sensitivity (S), Precision (P) and F1-Score (F1-S) were used to compare classification algorithms. The results obtained using these methods in Table 9, were evaluated and the attained results.

Table 9. Application outcomes of classification algorithms.

Algorithm	**Ac**	**S**	**P**	**F1-S**
LR	0.87	0.91	0.87	0.89
SVM	0.88	0.92	0.88	0.90
NB	0.87	0.89	0.89	0.89

RF	0.84	0.87	0.85	0.86
K-NN	0.84	0.87	0.85	0.86
DT	0.76	0.77	0.80	0.79

Machine learning algorithms have an important place in determining the heart attack risk and risk factors. In our study, Logistic Regression, Support Vector Machines, Naive in determining the risk of heart attack NB, RF, K-NN and DT classification methods were utilizated. As a result of the study, 88% accuracy rate was obtained with the LR method, 88% with the SVM method, 88% with the NB method, 84% with the RF method, 84% with the k-NN method, and 76% with the DT method. The algorithm with the highest accuracy rate was the SVM algorithm. Classification algorithms; accuracy, sensitivity Precision and F1-score were compared. On the other hand, SVM method has a moderate accuracy rate, and can give physicians and researchers the opportunity to interpret. For these reasons, which model to choose in applications can be determined by considering the special conditions of the application.

4. CONCLUSION

In the long term, the operating of the raw health data to identify cardiac info will aid save lives and descry cardiac abnormalities early. In this work, machine literacy ways are used to reuse the raw data and enable new and new demarcation related to heart complaint. Predicting heart complaint is grueling and veritably important in the medical field. still, if the complaint is determined earlier and preventative evaluations are taken in no time, mortality can be greatly controlled. It's largely desirable to further expand this exploration to concentrate on datasets of reelworld, rather than just theoretical styles and simulations.

SVMs were set up to be fairly accurate in prognosticating heart complaint. unborn assignments in this exploration can be fulfilled through colorful combinations from machine literacy ways to better vaticination ways. likewise, new point selection styles can be improved to gain a broader sense of important qualifications, thereby perfecting the performance of heart complaint vaticination.

Heart attack is a vital health problem in human society. This article summarizes the existing techniques and available methods for predicting this disease. SVM, an emerging field of artificial intelligence, has shown some promising results in other fields of high-precision medical diagnosis. In heart disease prediction, it remains an open field awaiting implementation. Some ML methods that can be utilizated for heart disease prediction are discussed, as well as breakthrough machine learning algorithms. Analytical comparisons are made to find the best available algorithms for medical datasets. In the future, we aim to continue work on ad hoc medical datasets, where the dataset changes over time and requires retraining on the dataset. As a result of this study, it is thought that researchers working in the field of cardiology will guide their clinical decisions. With the developed machine learning algorithms, medical errors, unwanted application types, medical costs are reduced and patient safety and quality of life are increased. In this study, classification algorithms were applied on 14 variables in the dataset. In future studies, classification algorithms can be applied by examining the variables in the data set in more detail by using optimization algorithms.

The SVM classifier predicted cardiovascular disease patients with 88% accuracy. Therefore, we conclude this project by stating that SVMs are best and most suitable for processing medical information sets. In the future, the system designed using the machine learning algorithm classifier used will be used to predict or diagnose alternative diseases. This work can be extended or improved for automation of cardiovascular disease analysis and deep learning algorithms.

5. REMARKS

ML has been reviewed in the medical field, especially for predicting heart disease. Algorithm completion has been determined. The proposed system is available in Availability Zones. More specific feature selection methods can be used to improve algorithm accuracy and get reliable results. When certain types of heart disease are diagnosed, patient care must be tailored to that specific condition. In essence, we conclude that a dataset of reasonably collected and reliable data is used to create a predictive model of heart disease. Therefore, the most important half of preprocessing is to prepare the data set to be used in the MLA program and to preprocess the data set to get better results. Associate in Nursing Requires the use of appropriate algorithmic programs to develop predictive models that produce correct results. ANNs have been found to have meaningful implications for predicting heart conditions in most models. Therefore, it helps all caregivers and patients. It is still growing in the field of nursing, thanks to the vast access to patient information in hospitals and clinics. As far as data set standards are important, prognostic accuracy should be considered. More hospitals should be encouraged to publish high-quality datasets (while respecting patient privacy). Create a model and come up with reasonable outcomes to help people exploit and treat heart disease in its early stages.

ACKNOWLEDGEMENT

The authors received no specific grant for the research, authorship, and/or publication of this article. This paper was presented as oral in 2nd International Conference on Engineering and Applied Natural Sciences ICEANS 2022.

REFERENCES

[1] Ahmed, M. R., Mahmud, S., Hossin, M., Jahan, H., and Noori, S, (2018), A Cloud Based Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms., IEEE 4th International Conference on Computer and Communications, pp. 1951-1955.

[2] Krishniah, V. V., Sekar, D., and Rao, D., (2012), Predicting the heart attack symptoms using biomedical data mining techniques, The International Journal of Computer Science and Applications, vol. 1, no. 3, pp. 10-18.

[3] WHO, https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1 Accessed on: Jan. 24, 2022

[4] Taşcı, M. E., and Şamlı, R. (2020), Veri Madenciliği İle Kalp Hastalığı Teşhisi, Avrupa Bilim ve Teknoloji Dergisi, pp. 88-95.

[5] Caliskan, A., and Yuksel, M., (2017), Classification of coronary artery disease data sets by using a deep neural network. The EuroBiotech Journal, vol:1 no.4, pp. 271-277.

[6] Soni, J., Ansari, U., Sharma, D., and Soni, S., (2011), Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction, International Journal of Computer Applications, vol. 17, no. 8, pp. 43-48.

[7] Chandna, D., (2014), Diagnosis of Heart Disease Using Data Mining Algorithm, (IJCSIT) International Journal of Computer Science and Information Technologies, vol. 5, no. 2, pp. 1678-1680.

[8] Verma, L., Srivastava, S., and Negi, P., (2016), A Hybrid Data MiningModel to Predict Coronary Artery Disease Cases Using Non-Invasive Clinical Data, Journal of Medical Systems, vol. 40, no. 7.

[9] Pandey, A. K., Pandey, P., Jaiswal, K., and Sen, A. A., (2013), Heart Disease Prediction Model using Decision Tree, IOSR Journal of Computer Engineering (IOSR-JCE), vol. 12, no. 6, pp. 83-86.

[10] Kim, J., Lee, J., and Lee, Y., (2015), Data-mining-based coronary heart disease risk prediction model using fuzzy logic and decision tree, Healthcare İnformatics Research, vol. 21, no. 3, pp. 167-174.

[11] Yin, W., Yao, Y., Gu, Y., Bao, W., Cheng H., (2021), Prediction of Heart Disease Probability Based on Various Body Function., pp. 267-275.

[12] Available:https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset

[13] Senaviratna, N. A., and Cooray, T., (2019), Diagnosing Multicollinearity of Logistic Regression Model, Asian Journal of Probability and Statistics, vol. 5 no. 2, pp. 1-9.

[14] Kurt, İ., and Türe, M., (2005), Tıp Öğrencilerinde Alkol Kullanımını Etkileyen Faktörlerin Belirlenmesinde Yapay Sinir Ağları ile Lojistik Regresyon Analizi'nin Karşılaştırılması, Trakya Üiversitesi Tıp Fakültesi Dergisi, vol. 22 no. 3, 142-153.

[15] Zhou, J., Qiu, Y., Zhu, S., Armaghani, D., Li, C., Nguyen, H., and Yagiz, S., (2021), Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate, Engineering Applications of Artificial Intelligence.

[16] Widodo, A., and Yang, B.-S., (2007), Support vector machine in machine condition monitoring and fault diagnosis, Mechanical Systems and Signal Processing, no. 21, pp. 2560–2574.

[17] Islam, M., Mahmud, A.A., Machine learning approaches for modeling spammer behavior, Asia information retrieval symposium, pp. 251-260, Springer, Berlin, Heidelberg.

[18] Sevli, O., (2019), Göğüs Kanseri Teşhisinde Farklı Makine Öğrenmesi Tekniklerinin Performans Karşılaştırması, Avrupa Bilim ve Teknoloji Dergisi no. 16, pp. 176-185.

[19] Saritas, M. M., and Yasar, A., (2019), Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification, International Journal of Intelligent Systems and Applications in Engineering, vol. 7, no. 2, pp. 88-91.

[20] K.Vijiya Kumar., (2019), Random Forest Algorithm for the Prediction of Diabetes, Proceeding of International Conference on Systems Computation Automation and Networking.

[21] Katarya, R., and Srinivas, P., (2020), Predicting Heart Disease at Early Stages using Machine Learning: A Survey, Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 302-305.

[22] Kılınç, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimşek, M., and Özçift, A., (2016), KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi, Marmara Fen Bilimleri Dergisi no.3, pp. 86-94.

[23] Saçlı, B., Aydınalp, C., Cansız , G., Joof, S., Yılmaz, T., Çayören, M., Akduman , İ., (2019), Microwave dielectric property based classification of renal calculi: Application of a KNN algorithm, Computers in Biology and Medicine, no.112.

[24] Uddin, S., Khan, A., Hossain, M., and Moni, M., (2019), Comparing different supervised machine learning algorithms for disease prediction, Mechanical Systems and Signal Processing, vol. 19, no. 281, pp. 1–16.