MACHINE LEARNING BASED FEATURE OPTIMIZATION AND EARLY DETECTION SYSTEM IN HEART DISEASES

Erman Özer
Deparment of Computer Engineering
Recep Tayyip Erdogan University
Rize, Turkey
erman.ozer@erdogan.edu.tr

Hasan Aydos
Department of Computer Engineering
Recep Tayyip Erdogan University
Rize, Turkey
hasan_aydos20@erdogan.edu.tr

2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS) | 979-8-3503-8306-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISAS60782.2023.10391754

Abstract:

In today's world, humanity faces a myriad of challenges, many of which pose significant threats to our well-being. Chief among these challenges are health-related issues. Among these health problems, heart diseases stand out as the leading cause of mortality. Consequently, the early diagnosis of heart diseases plays a pivotal role in mitigating mortality rates and enhancing people's overall quality of life. This study aims to employ machine learning algorithms to enhance the early detection capabilities of heart disease. A dataset comprising the health records of 253,680 patients with heart disease is analyzed using five distinct machine learning algorithms: Logistic Regression, K-Nearest Neighbors Classifier, Decision Tree Classifier, Naïve Bayes, and Linear Support Vector Machine (Linear SVM). The dataset is partitioned, with 80% allocated for training the algorithms and the remaining 20% for testing. Furthermore, the study's evaluation employs four different metrics: accuracy, precision, recall, and the F1-measure. Initially, early diagnosis of heart disease is attempted using the complete set of features in the dataset. However, this approach results in excessive costs and time consumption. Subsequently, a feature reduction process is implemented to optimize resource utilization, yielding an improved early detection rate. The research findings indicate that Logistic Regression outperforms the other algorithms, achieving the highest success rate with an accuracy score of 90.67%. These research results underscore the substantial contribution of machine learning algorithms to the early detection of heart disease, ultimately enhancing the quality of life for individuals.

Keywords—heart disease, machine learning, corr function.

I. INTRODUCTION

In the developing world of today, the influence and position of computer software on people's lives cannot be ignored. It is well known that artificial intelligence is one of the most important components of computer programs. This has attracted the attention of many researchers, educators and even many companies. Today, the integration of artificial intelligence into healthcare is the subject of research [1,2]. In general, it is believed that artificial intelligence can be used for early diagnosis of diseases, which is one of the major concerns of people today [3].

This research is about using machine learning algorithms for early diagnosis of heart diseases that could cost many lives by 2050 [4]. For training machine learning algorithms, a dataset consisting of individuals who have already had heart disease in their medical history and individuals who have never had heart problems before will be used. Data from more than 250,000 people will be used for the proposed research, and the accuracy of the results will be compared against the most popular machine learning algorithms. Corr function will be used to reduce data features in the dataset to save time and unit costs. Agile project development methods will be used during the project. Sprint planning and processes are dynamically managed to control the project process.

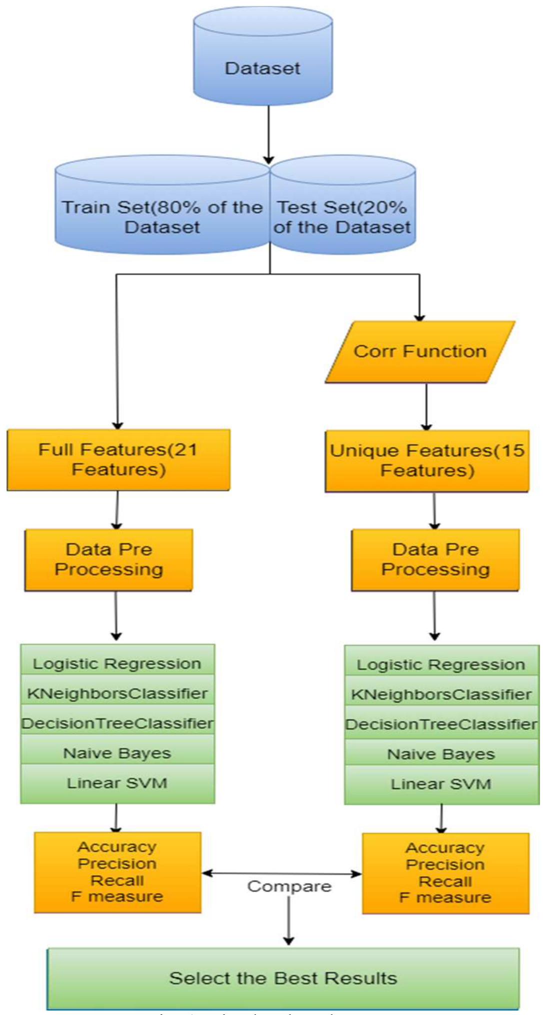

Fig. 1. The developed system

979-8-3503-8306-5/23/$31.00 ©2023 IEEE

As a method for early detection of heart attack risk, the system we developed is illustrated in Figure-1 and works as follows:

- Firstly, a dataset consisting of real patient data appropriate for diagnosis is used. 80% of the dataset is used for training and 20% for testing.
- Corr function is used to examine the correlation of all features in the dataset, and feature reduction is achieved. The success rate before and after using the corr function is compared, and the gains in terms of time and cost are evaluated.
- In this system, five different machine learning algorithms are applied: Logistic Regression, KNeighbors Classifier, DecisionTreeClassifier, Naïve Bayes, Linear SVM.
- In the final stage, the results obtained from the machine learning algorithms are compared with four different metrics (Accuracy, precision, recall, f1-measure).

II. LITERATURE REVIEW

There are many studies on heart disease using traditional models and a small number of patients. Venkatalakshmi and Shivsankar et al. Numerous experiments were conducted with the same datasets with the Weka tool. The dataset they use consists of 13 features and 294 records. They used naive bayes and decision tree algorithms. In their studies, 85.03% accuracy was obtained with Naive Bayes and 84.01% accuracy with Decision Tree, respectively [5]. Jindal et al. obtained results using the data of 304 patients of different ages using the World Health Organization. This dataset consists of 14 different features. Performances were compared with KNN, Logistic Regression, Random Forest Classifier. The best accuracy result was obtained with KNN with 88.52% [6]. Haq et al. [7] proposed a heart disease identification method using feature selection and expectations. Sequential Back Selection Algorithm (SBS FS) performance for the accuracy result (KNN) is used for the special selection. Despite good results, the number of patients used is very small. Data from 294 patients were used.These three studies obtained data from a relatively small number of patients compared to other studies.

Samuel et al. propose an approach using both artificial neural network and logic methods as a hybrid. According to the results obtained, approximately 4.4% higher results were obtained than ANN [8]. Sathya et al analyze different classifier features such as heart disease SVM, KNN and MLP using WEKA 3.8 software. The SVM Classifier is superior to KNN and MLP, achieving 85.9% accuracy with a minimum of 15.3% false positives. Although it achieves high accuracy, the false positive rate is quite low [9]. Prince et al. made comparisons with 6 different machine learning algorithms and 4 different metrics. According to the results obtained, the best result was obtained with 73% Decision Tree [10].

According to the studies, the number of patients used in some studies is low, the accuracy or false positive rate obtained in some is very low, and the feature reduction method is not used in some. In this study, both feature reductions were used and accurate high results were obtained. With feature reduction, very high gains in terms of time have been obtained.

III. MACHINE LEARNING ALGORITHMS

The algorithms integrated into this study for the diagnosis of heart attack are as follows:

A. Logistic Regression

A mathematical machine learning algorithm that finds relationships between two different data to predict another data. This data analysis technique analyzes the relationship between two different data and produces a result. The resulting output is 0 or 1. "0" represents the answer "no" while "1" represents the answer "yes". Logistic Regression is faster, less computationally intensive, easier to troubleshoot, and less complex than other machine learning techniques. Therefore, it is often preferred [11]. Its mathematical function is:

$$\mathrm{Logit(p)} = \ln(\mathrm{p}/1 - \mathrm{p}) \tag{1}$$

B. K-Nearest Neighbors Classifier

Compares unclassified data with data in the dataset to determine a specific outcome. The K-Neighbors Classsifier algorithm determines the number of clasest neightbors that will be used to determine which class the data belongs to by specifying the values of k. In this study, the value k is chosen as 20 since it yielded the best results. The value for each neighbor is calculated, and it is determined which class it belongs to based on which class has the most similarity [12]. Its mathematical function is:

$$d = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \tag{2}$$

C. Decision Tree Classifier

A machine learning algorithm that divides large sets of data into small sets by applying decision rules to datasets with a large amount of data in order to reach accurate results [13]. In short, the basic aim is to create a model that predicts a variable by making the simplest inferences from the data characteristics [14].

D. Naïve Bayes

A machine learning algorithm that analyzes data based on probability principles. First, a certain amount of data is taught to the machine learning algorithm, and then it quickly shows which class newly added data belongs to based on the learned data [15]. Its mathematical expression is:

$$p(class\ |data) = \frac{p(data|class)x\ p(class)}{p(data)} \tag{3}$$

E. Linear Support Vector Machines

Draws a line to separate data placed on a specific plane. This machine learning algorithm shows which class newly added data should belong to by ensuring that the data is at its farthest point from each other. The area within this line is called the "margin". Values between "+1" and "-1" are taken within this area [16].

$$If\ y = 0\ is\ w^T.x + b < 0 \tag{4}$$

$$If\ y = 1\ is\ w^T.x + b \geq 0 \quad (5)$$

(Where W is the weight vector, x is the input vector, and b is the deflection.

IV. DATASET

The dataset that we will use to train machine learning algorithms for early diagnosis of heart disease is the Behavioral Risk Factor Surveillance System (BRFSS), a health-related survey conducted annually by the CD. These survey results have been collected in the United States since 1984 and are continually updated each year. This dataset on heart disease includes 253,680 instances, with 23,893 individuals having the disease and 229,787 individuals not having the disease. The dataset contains 22 variables, one of which indicates whether an individual has the disease. This variable plays an important role in the early diagnosis of the disease. [17]

TABLE I. FEATUERS AND DESCRIPTIONS

Feature	**Description**
*Heart Disease	Whether or not the individual had a heart attack, with "1" representing those who have had a heart attack and "0" representing those who have not.
*High BP	Whether or not the individual has high blood pressure, with "1" representing those who have high blood pressure and "0" representing those who do not.
*High Chol	Whether or not the individual has high cholesterol, with "1" representing those who have high cholesterol and "0" representing those who do not.
*Chol Check	Whether or not the individual had cholesterol screening, with "1" representing those who have had cholesterol screening and "0" representing those who have not.
*BMI	Body mass index (BMI) of the individual.
*Smoker	Whether or not the individual smokes, with "1" representing smokers and "0" representing non-smokers.
*Stroke	Whether or not the individual had a stroke, with "1" representing those who have had a stroke and "0" representing those who have not.
*Diabetes	Whether or not the individual has diabetes, with "1" representing those who have diabetes and "0" representing those who do not.
PhysActivity	Whether or not the individual engages in physical activity, with "1" representing those who do engage in physical activity and "0" representing those who do not.
Fruits	Whether or not the individual consumes fruits, with "1" representing those who do consume fruits and "0" representing those who do not.
Veggies	Whether or not the individual consumes vegetables and "0" representing those who do not.
HvyAlcohol Consump	Whether or not the individual has high alcohol consumption, with "1" representing those who have high alcohol consumption and "0" representing those who do not.
*AnyHealthCare	Whether or not the individual received any health support, with "1" representing those who have received health support and "0" representing those who have not.
*NoDocbcCost	Whether or not the individual has a family history of heart disease, with "1" representing those who have a family history and "0" representing those who do not.
*GenHlth	Whether or not the individual has a family history of heart disease, with "1" representing those who have a family history and "0" representing those who do not.
*MenHlth	Whether or not the individual has a mental health condition, with "1" representing those who have a mental health condition and "0" representing those who do not.
*PhysHlth	Whether or not the individual has a physical health condition, with "1" representing those who have physical health condition and "0" representing those who do not.
*DiffWalk	Walking limitations of the individual.
*Sex	Gender of the individual, with "0" representing males and "1" representing females.
*Age	Age of the individual.
Education	Education level of the individual.
Income	Income level of the individual.

V. METRICS

The parameters used for classification criteria are listed below:

- True Positive (TP): Predicting that a person has heart disease when they have it.
- False Positive (FP): Predicting that a person has heart disease when they don't.
- True Negative (TN): Predicting that a person does not have heart disease when they don't.
- False Negative (FN): Predicting that a person does not have heart disease when they do.

Additionally, it is compared the result of the algorithms with four different metrics:

A. *Accuracy*

A metric that calculates the accuracy that calculates the accuracy rate of a model.

$$\frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

B. *Recall*

It is a metric that tries to estimate how positive the transactions that only want to be positive are.

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

C. *Precision*

It is a metric that shows how many of the values that only positive are positive.

$$Precion = \frac{TP}{TP + FP} \quad (8)$$

D. *F1-Measure*

A metric that shows the harmonic average of precision and recall values.

$$\mathrm{F}_1 = 2 * \frac{precision * recall}{precision + recall} \quad (9)$$

VI. RESULTS

In this study, a dataset of 253,680 records of heart patients is used. Instead of traditional algorithms, machine learning algorithms that provide more accurate and faster results are used. These algorithms are logistic regression, KNeighbors Classifier, DecisionTree Classifier, Naïve Bayes, and Linear SVM. Four different metrics are used to test these five different algorithms: (Accuracy, precision, recall, f1-measure).

The results of the four metrics obtained are given in table form:

TABLE II. RESULTS OF FULL FEATUERS

	Full Features (21 feature)			
Algorithm / Metrics	***Acc***	***P***	***R***	***f***
Logistic Regression	90.62	91.13	98.27	94.56
KNeighbors	90.51	91.23	99.12	95.01
DecisionTree	85.2	92.17	91.35	91.75
Naïve bayes	81.99	92.41	91.21	91.8
Linear SVM	90.60	91.85	99.45	95.49

unique features (marked * in Table 1) that remain after applying the Corr function. It can be observed that the reduction of features has both time and cost benefits. The Corr function used to determine the correlation is shown in Figure 2.

$$r = \frac{n.\sum X_i.Y_\mathrm{i} - \sum Y_\mathrm{i}.\sum Y_\mathrm{i}}{\sqrt{[n.\sum X_i^2 - (\sum X_i)^2].[n.\sum Y_i^2 - (\sum Y_i)^2]}} \quad (10)$$

The results of the algorithms and metrics computed with the 15 different features obtained with the Corr function are compared in the following table:

HeartDiseaseorAttack	HighBP	High Chol	Chol Check	BMI	Smoker	Stroke	Diabetes	Phys Activity	Fruits	Veggies	HvyAlcohol Consump	AnyHealth care	NoDocbcCost	GenHlth	Ment Hlth	Phys Hlth	DiffWalk	Sex	Age	Education	Income
1,00	0,21	0,18	0,04	0,05	0,11	0,20	0,18	-0,09	-0,02	-0,04	-0,03	0,02	0,03	0,26	0,06	0,18	0,21	0,09	0,22	-0,10	-0,14
0,21	1,00	0,30	0,10	0,21	0,10	0,13	0,27	-0,13	-0,04	-0,06	0,00	0,04	0,02	0,30	0,06	0,16	0,22	0,05	0,34	-0,14	-0,17
0,18	0,30	1,00	0,09	0,11	0,09	0,09	0,21	-0,08	-0,04	-0,04	-0,01	0,04	0,01	0,21	0,06	0,12	0,14	0,03	0,27	-0,07	-0,09
0,04	0,10	0,09	1,00	0,03	-0,01	0,02	0,07	0,00	0,02	0,01	-0,02	0,12	-0,06	0,05	-0,01	0,03	0,04	-0,02	0,09	0,00	0,01
0,05	0,21	0,11	0,03	1,00	0,01	0,02	0,22	-0,15	-0,09	-0,06	-0,05	-0,02	0,06	0,24	0,09	0,12	0,20	0,04	-0,04	-0,10	-0,10
0,11	0,10	0,09	-0,01	0,01	1,00	0,06	0,06	-0,09	-0,08	-0,03	0,10	-0,02	0,05	0,16	0,09	0,12	0,12	0,09	0,12	-0,16	-0,12
0,20	0,13	0,09	0,02	0,02	0,06	1,00	0,11	-0,07	-0,01	-0,04	-0,02	0,01	0,03	0,18	0,07	0,15	0,18	0,00	0,13	-0,08	-0,13
0,18	0,27	0,21	0,07	0,22	0,06	0,11	1,00	-0,12	-0,04	-0,06	-0,06	0,02	0,04	0,30	0,07	0,18	0,22	0,03	0,19	-0,13	-0,17
-0,09	-0,13	-0,08	0,00	-0,15	-0,09	-0,07	-0,12	1,00	0,14	0,15	0,01	0,04	-0,06	-0,27	-0,13	-0,22	-0,25	0,03	-0,09	0,20	0,20
-0,02	-0,04	-0,04	0,02	-0,09	-0,08	-0,01	-0,04	0,14	1,00	0,25	-0,04	0,03	-0,04	-0,10	-0,07	-0,04	-0,05	-0,09	0,06	0,11	0,08
-0,04	-0,06	-0,04	0,01	-0,06	-0,03	-0,04	-0,06	0,15	0,25	1,00	0,02	0,03	-0,03	-0,12	-0,06	-0,06	-0,08	-0,06	-0,01	0,15	0,15
-0,03	0,00	-0,01	-0,02	-0,05	0,10	-0,02	-0,06	0,01	-0,04	0,02	1,00	-0,01	0,00	-0,04	0,02	-0,03	-0,04	0,01	-0,03	0,02	0,05
0,02	0,04	0,04	0,12	-0,02	-0,02	0,01	0,02	0,04	0,03	0,03	-0,01	1,00	-0,23	-0,04	-0,05	-0,01	0,01	-0,02	0,14	0,12	0,16
0,03	0,02	0,01	-0,06	0,06	0,05	0,03	0,04	-0,06	-0,04	-0,03	0,00	-0,23	1,00	0,17	0,19	0,15	0,12	-0,04	-0,12	-0,10	-0,20
0,26	0,30	0,21	0,05	0,24	0,16	0,18	0,30	-0,27	-0,10	-0,12	-0,04	-0,04	0,17	1,00	0,30	0,52	0,46	-0,01	0,15	-0,28	-0,37
0,06	0,06	0,06	-0,01	0,09	0,09	0,07	0,07	-0,13	-0,07	-0,06	0,02	-0,05	0,19	0,30	1,00	0,35	0,23	-0,08	-0,09	-0,10	-0,21
0,18	0,16	0,12	0,03	0,12	0,12	0,15	0,18	-0,22	-0,04	-0,06	-0,03	-0,01	0,15	0,52	0,35	1,00	0,48	-0,04	0,10	-0,16	-0,27
0,21	0,22	0,14	0,04	0,20	0,12	0,18	0,22	-0,25	-0,05	-0,08	-0,04	0,01	0,12	0,46	0,23	0,48	1,00	-0,07	0,20	-0,19	-0,32
0,09	0,05	0,03	-0,02	0,04	0,09	0,00	0,03	0,03	-0,09	-0,06	0,01	-0,02	-0,04	-0,01	-0,08	-0,04	-0,07	1,00	-0,03	0,02	0,13
0,22	0,34	0,27	0,09	-0,04	0,12	0,13	0,19	-0,09	0,06	-0,01	-0,03	0,14	-0,12	0,15	-0,09	0,10	0,20	-0,03	1,00	-0,10	-0,13
-0,10	-0,14	-0,07	0,00	-0,10	-0,16	-0,08	-0,13	0,20	0,11	0,15	0,02	0,12	-0,10	-0,28	-0,10	-0,16	-0,19	0,02	-0,10	1,00	0,45
-0,14	-0,17	-0,09	0,01	-0,10	-0,12	-0,13	-0,17	0,20	0,08	0,15	0,05	0,16	-0,20	-0,37	-0,21	-0,27	-0,32	0,13	-0,13	0,45	1,00

Fig. 2. Corr Function Results

The results are obtained by analyzing and recognizing 21 different features. The main innovation of this study compared to other studies is the use of the Corr function to study the correlation of all features, which led to a reduction of features. In order to calculate the correlation coefficient between the features of the dataset, a code was developed in Python to order the feature strengths in the range [-1, 1] using the Corr function. After calculating the correlation coefficient matrix, each feature was compared with its mean correlation. The feature change and the target change are compared to determine whether they are compatible or not. If the value of the coefficient is negative, there is an inverse relationship between the variables. Here, after the min-max transformation, the uncertainties of the features in our dataset are removed. Before applying the Corr function, a comparison is made between the 21 different features in the dataset and the 15

TABLE III. RESULTS OF REDUCED FEATUERS

	Unique Features (15 Feature)			
Algorithm / Metric	***Acc***	***P***	***R***	***f***
Logistic Regression	90.67	90.77	97.25	93.89
KNeighbors	90.47	91.05	98.85	94.78
DecisionTree	87.05	92.25	94.53	93.37
Naive bayes	82.80	92.27	94.35	93.29
Linear SVM	90.50	91.15	99.23	95.01

As can be seen in Table 2 and Table 3, although the number of features has decreased, the results obtained are quite good. In this study, by reducing the number of features, a temporal gain of about 20% is obtained. This provides us with early detection of heart disease.

VII. CONCLUSION

In today's world, which is constantly evolving and changing, heart disease is a threat to people's lives. This study aims to use machine learning algorithms for early diagnosis of heart disease. Five different algorithms are used for early diagnosis, and data is trained to provide results by teaching them symptoms of heart disease. The success rate of these algorithms are 90.67% for Logistic Regression, 90.47% for KNeighborClassifier, 82.80% for Naïve Bayes, 87.05% for DecisionTree, and 90.50% for Linear SVM. The results obtained are compared with four different metrics (accuracy, precision, recall, f1-measure). The Corr function is used to reduce the features by ensuring the correlation of all data, saving time and unit costs.

REFERENCES

[1] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R. Nour, "An Intelligent Learning System Based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection," IEEE Access, vol. 7, pp. 180235– 180243, 2019, doi: 10.1109/ACCESS.2019.2952107

[2] G. Manogaran, D. Lopez, C. Thota, M. Kaja Abbas, S. Pyne, and R. Sundarasekar, ''Big data analytics in healthcare Internet of Things,'' in Innovative Healthcare Systerm. Cham, Switzerland: Springer, 2017, pp. 263–284, doi: 10.1007/978-3-319-55774-8_10.

[3] L. M. Dang, M. J. Piran, D. Han, K. Min, and H. Moon, ''A survey on Internet of Things and cloud computing for healthcare,'' Electronics, vol. 8, no. 7, p. 768, Jul. 2019, doi: 10.3390/electronics8070768.

[4] F. Ahmed, ''An Internet of Things (IoT) application for predicting the quantity of future heart attack patients,'' Int. J. Control Automat., vol. 164, no. 6, pp. 36–40, Apr. 2017, doi: 10.5120/ijca2017913773.

[5] Venkatalakshmi B, Shivsankar MV. Heart disease diagnosis using predictive data mining. International Journal of Innovative Research in Science, Engineering and Technology. 2014 Mar;3(3):1873-7.

[6] Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. Heart disease prediction using machine learning algorithms. In IOP Conference Series: Materials Science and Engineering (Vol. 1022, No. 1, p. 012072). IOP Publishing(2021).

[7] A. U. Haq, J. Li, M. H. Memon, M. H. Memon, J. Khan, and S. M. Marium, ''Heart disease prediction system using model of machine learning and sequential backward selection algorithm for features selection,'' in Proc. IEEE 5th Int. Conf. Converg. Technol. (ICT), Mar. 2019, pp. 1–4.

[8] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, ''An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction,'' Expert Syst. Appl., vol. 68, pp. 163–172, Feb. 2017.

[9] K. Sathya and R. Karthiban, "Performance Analysis of Heart Disease Classification for Computer Diagnosis System," 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1-7, doi:10.1109/ICCCI48352.2020.9104089.

[10] R. J. P. Princy, S. Parthasarathy, P. S. Hency Jose, A. Raj Lakshminarayanan and S. Jeganathan, "Prediction of Cardiac Disease using Supervised Machine Learning Algorithms," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 570-575, doi: 10.1109/ICICCS48265.2020.9121169

[11] D. B. Springer, L. Tarassenko and G. D. Clifford, "Logistic Regression-HSMM-Based Heart Sound Segmentation," in IEEE Transactions on Biomedical Engineering, vol. 63, no. 4, pp. 822-832, April 2016, doi: 10.1109/TBME.2015.2475278.

[12] C. C, "Prediction of Heart Disease using Different KNN Classifier," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2021, pp. 1186-1194, doi: 10.1109/ICICCS51141.2021.9432178.

[13] L. M. R. Tarouco, L. M. Bertholdo, L. Z. Granville, L. M. R. Arbiza, F. Carbone, M. Marotta, and J. J. C. de Santanna, ''Internet of Things in healthcare: Interoperatibility and security issues,'' in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2012, pp. 6121–6125.

[14] M. A. Jabbar, B. L. Deekshatulu and P. Chndra, "Alternating decision trees for early diagnosis of heart disease," International Conference on Circuits, Communication, Control and Computing, Bangalore, India, 2014, pp. 322-328, doi: 10.1109/CIMCA.2014.7057816.

[15] S. K. L. Sameer and P. Sriramya, "Improving the Accuracy for Prediction of Heart Disease by Novel Feature Selection Scheme using Decision tree

comparing with Naive-Bayes Classifier Algorithms," 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2022, pp. 1-8, doi: 10.1109/ICBATS54253.2022.9758926.

[16] C. M. Chethan Malode, K. Bhargavi, B. G. Gunasheela, G. Kavana and R. Sushmitha, "Soft set and Fuzzy Rules Enabled SVM Approach for Heart Attack Risk Classification Among Adolescents," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697650.

[17] alex teboul, “Heart disease health indicators dataset notebook,” Kaggle, https://www.kaggle.com/code/alexteboul/heart-disease-health-indicators-dataset-notebook (accessed Oct. 25, 2023).