Improving the Accuracy of Heart Disease Prediction Approach of Machine Learning Algorithms

Conference Paper · February 2023

DOI: 10.1109/DELCON57910.2023.10127250

4 authors, including:

Mohammed Nasir Uddin
Jagannath University - Bangladesh
41 PUBLICATIONS **218** CITATIONS
SEE PROFILE

Syada Tasmia Alvi
Daffodil International University
17 PUBLICATIONS **134** CITATIONS
SEE PROFILE

Chowdhury Abida Anjum Era
Daffodil International University
2 PUBLICATIONS **1** CITATION
SEE PROFILE

Improving the Accuracy of Heart Disease Prediction Approach of Machine Learning Algorithms

Md. Belal Hossain
Dept. of CSE
Jagannath University
Dhaka, Bangladesh
belaljnu58@gmail.com

Mohammed Nasir Uddin
Dept. of CSE
Jagannath University
Dhaka, Bangladesh
nasir@cse.jnu.ac.bd

Syada Tasmia Alvi
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
syada.cse@diu.edu.bd

Chowdhury Abida Anjum Era
Dept. of CSE
Daffodil International University
Dhaka, Bangladesh
chowdhury.cse@diu.edu.bd

2023 2nd Edition of IEEE Delhi Section Flagship Conference (DELCON) | 979-8-3503-2206-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/DELCON57910.2023.10127250

***Abstract*—The work is about forecasting heart disease. First and foremost, we gathered data from various sources and divided it into two portions, one of which is 80% and the other is 20%, where the first part is for training and the remainder is reserved for the test dataset. After collecting this dataset, we applied the pre-processing formula and different classifier algorithms. K-Nearest Neighbor, Support Vector Machine, Decision Tree, Random Forest, Naive Bayes & Logistic Regression are the techniques utilized here. When compared to other algorithms, Logistic Regression, KNN, and SVM provided the same or superior accuracy. Precision, Recall, F1 score, and ERR are used to measure accuracy. Gender, Glycogen, BP, and Heartrate are some of the prefixes used while training and found to be different major vulnerable factors of heart diseases. The direction of this work is real-life experiments and clinical trials using different devices.**

***Keywords*—machine learning, data analysis, big data logistic regression, healthcare**

I. Introduction

Life depends on the competent function of the heart since the heart is a necessary part of our body. When the heart fails to pump sufficient b lood, i t c an h arm t he l iver, kidneys, and other organs [1]. Every year, many people are affected by this type of disease and cannot obtain a good diagnosis before becoming afflicted. A h uge n umber o f individuals throughout the world die as a result of this sickness, and a large number of people face this problem every year [2]. Among the several life-threatening disorders, heart disease has piqued the attention of medical researchers. Heart disease diagnosis is challenging but can automatically forecast the patient's status for better treatment. The diagnosis of heart disease typically depends on the patient's clinical symptoms. Many factors raise the risk of developing heart diseases [3], for intense smoking, cholesterol levels, heart disease in the family, overweight, high blood pressure, and lack of physical activity [4].

World Health Organization (WHO) stated that the fundamental human right is a person's good health. Heart disease is a frequent ailment among people nowadays, and diagnosing it by hand is very expensive and did not develop anywhere near the patient's symptoms before the hand-to-hand diagnosis. Machine learning is now widely employed in the healthcare system for medical objectives such as predicting heart disease, cardiovascular illness, and breast cancer, among other things [2].

The goal of this work is to predict cardiac illness and evaluate the performance of different machine learning models using a healthcare dataset, while also pinpointing influential factors in the data. Normally, when symptoms of heart disease appear, people seek care at a hospital or clinic for various tests and therapies [5]. As these tests and therapies are costly for most people, we can use the results of this study to get insight into the patient's position before they develop these diseases in the human body.

The following structure describes the way the paper is organized: The literature on heart disease detection is reviewed in section II. Section III outlines the suggested system. Section IV evaluates the proposed system from different perspectives. Section V draws conclusions based on the research. Finally, section VI outlines future work for the system.

II. Related Works

Gao et al. [6] put forward a model based on an ensemble approach to forecasting a patient's cardiac disease as well as heart disease using unique grouping techniques. They took 1025 data points with 13 attributes from Kaggle, then ran two feature selection techniques (linear discriminant analysis and principal component analysis) to uncover the most meaningful features. Five algorithms were then applied to build the model: K-Nearest Neighbors, Support Vector Machines, Decision Trees, Random Forests, and Naive Bayes. Utilizing the bagging ensemble technique with a decision tree yielded an accuracy score of 98.6%. There are numerous methods for predicting coronary illness, and various studies have been

completed. Using a strategic calculation, 77.0% precision was obtained in the early investigations in the field o f coronary illness [7]. Another study employed the CLASSIT concept bunch structure [8], and the exhibition score was shown to be 78.9% in a comparable year. The overall result was calculated using a coronary disease dataset of the University of California, Irvine (UCI) [9]. Authors suggested a hybrid computation that combines KNN and ID3 calculations. It is also known as a pre-handled calculation because it is pre-prepared data using the KNN. In this classifier, ID3 is used to predict coronary sickness, and KNN calculation is to characterize the amazing evaluation [10]. Latha and Jeeva [11], used an ensembling classification model to find serious risk factors with the best accuracy. This work has achieved similar accuracy of 85.48% for all their models. In [12], authors offered a cardiovascular disease prediction using boundary mapping techniques and other classification algorithms. This research finds perfect accuracy using random forest, which is about 84.81%. Nashif et.al [13] offered a cloud-based approach for predicting heart illness that uses artificial intelligence to distinguish current cardiovascular disease. The framework was constructed with Arduino and received results at regular in-tervals, such as pulse, pulse, and temperature. The framework improved precision with SVM. Phasinam et al. [14] conducted an analysis to assess the efficiency of the approaches that predict diseases. They aim to assess the general effectiveness of ML techniques to enhance accuracy. Kadam et al. [15] developed a system where they studied different machine-learning approaches utilizing Image Processing, Sound Signal Processing, and other classification algorithms to analyze data. The Random Forest classifier yielded the highest accuracy rate of 90.16%.

III. Proposed System

We present a model using machine learning approaches to identify heart disease, incorporating various techniques. We create a model that uses pre-processing, feature selection, data splitting, and classifiers algorithm as shown in Fig. 1. It is a crucial task for classification. We use hard negative mining to reduce false-positive and get better classification performance. In this study, various supervised machine-learning approaches have been employed. By applying these algorithms, we found the intended accuracy of which one is best compared to others. First of all, we collected data from the UCI Machine Learning Repository, and then we employed various pre-processing and feature selection approaches. At last, we use different models and then run that preprocessed data into the model and get accuracy based on the dataset [9]. Random forest, Decision tree, SVM, Logistic Regression, NB, KNN, and XBoost are the algorithms in question. So several steps are included in the prediction of heart disease.

A. Dataset Collection from UCI

UCI [9], which has 16 sections and 4238 lines, was used to gather data on heart illness. Each segment has its own trademark that can be used directly or indirectly to treat coronary illness. So, principally, we utilize all ascribed, and after that, in view of hazard elements and expectations, we diminish the trait to improve execution and the significant risk attribute. Every one of those highlights has age and sex, with sexual orientation allocated to 0; what's more, 1 for males and 0 for females individually. Here a few items are clustered by 0 or 1. 0 and 1 separately denote nonattendance or presence. The 16 characteristics that are used to predict heart disease are shown in Table I.

B. Data Preprocessing Technique

a) Data Cleaning: In general, true information will be weak, noisy, and conflicting. The data-gathering instruments utilized may have problems. Human or computer faults could have slowed down the flow of data. There is also the possibility of data transmission errors. For the mining system, dirty information might cause havoc. Data cleaning schedules are designed to address gaps in information, reduce any irregularities, and fix any mistakes in the data. In "Fig. 2" the methods of data cleaning are illustrated.

b) Data Integration : Information integration involves examining data from several sources and combining it into logical data storage, similar to information warehousing.

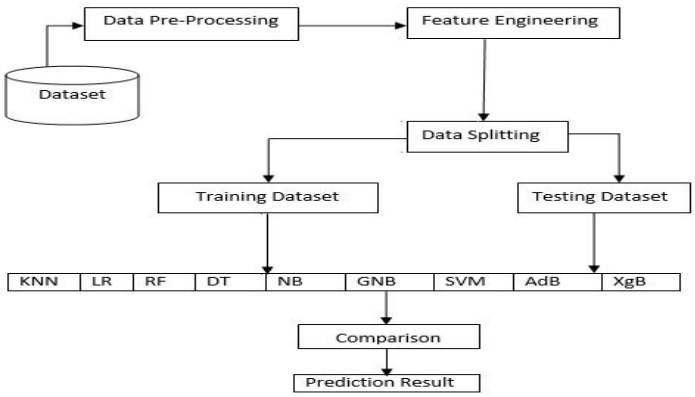

Fig. 1. Model of the proposed system for heart disease prediction

TABLE I
ATTRIBUTE DESCRIPTION

Sr.No.	Attribute Name	Description	Data Type
1	Gender	male 1 or female 0	Binary
2	Age	Age of the person in year	Numeric
3	Education	Primary, Secondary, Higher Secondary and Graduation	Numeric
4	Diabetes	If it exists, then 1, Otherwise 0	Numeric
5	CurrentSmoker	If it exists, Yes-1, Otherwise, No-0	Numeric
6	CigsPerDay	Per day taking cigarette number	Numeric
7	BPMeds	Blood Pressure Medication-0, 1	Numeric
8	PrevalentStroke	Stroke occurred in life-0, 1	Numeric
9	PrevalentHyp	Hypertension existence-0, 1	Numeric
10	TotalCholest	Whole cholesterol level	Numeric
11	SyBP	Blood Pressure(Systolic)	Numeric
12	Diat BP	Blood Pressure(Diastolic)	Numeric
13	B.M.I	Body-Mass-Index	Numeric
14	HeartRate	Heart BeatRate	Numeric
15	Glucose(GE)	Normal Blood GE Level	Numeric
16	TenYearsCHD	Ten Years Coronary Heart Disease In 0,1	Binary

Pattern reconciliation is an important consideration in data integration. Repetition is another key concern. It's possible that a characteristic taken from another table is excessive. Inconsistencies in trait or measurement names can also lead to redundancies in the subsequent informative index. Our system's data integration is depicted in "Fig. 3".

c) Data Transformation-DBSCAN : DBSCAN is a clustering approach that distinguishes dense clusters from sparse ones and is adept at dealing with outliers. K-means clustering may be used as an alternative when noise and outliers are absent. In "Fig. 4", DBSCAN is shown. Here, information is transformed into appropriate mining types. The following are included in the information transformation:

- Normalization is a technique that levels property data to fall inside a specific range, such as -1 to 1 or 0 to 1.
- Smoothing is a technique for removing turbulence from data. Such approaches include binning, bunching, and relapsing.
- Aggregation is applying outline or collection actions to data.
- Low-level or crude data is substituted by higher-level concepts via idea chains of command to generalize the data.
- To begin the clustering process, a certain number of

```
seaborn.countplot(dataset['TenYearCHD'])
dataset.isnull().values.any()
dataset.isna().sum()
dataset.mean()
new_df=dataset.fillna({
    'education':2,
    'cigsPerDay':0,
    'BPMeds':0.0,
    'totChol':234,
    'BMI':25,
    'heartRate':75,
    'glucose':78,
})
new_df
```

t[15]:

	male	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP	BMI	heartRate	glucose
0	1	39	4.0	0	0.0	0.0	0	0	0	195.0	106.0	70.0	26.97	80.0	77.0
1	0	46	2.0	0	0.0	0.0	0	0	0	250.0	121.0	81.0	28.73	95.0	76.0
2	1	48	1.0	1	20.0	0.0	0	0	0	245.0	127.5	80.0	25.34	75.0	70.0
3	0	61	3.0	1	30.0	0.0	0	1	0	225.0	150.0	95.0	28.58	65.0	103.0
4	0	46	3.0	1	23.0	0.0	0	0	0	285.0	130.0	84.0	23.10	85.0	85.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4233	1	50	1.0	1	1.0	0.0	0	1	0	313.0	179.0	92.0	25.97	66.0	86.0
4234	1	51	3.0	1	43.0	0.0	0	0	0	207.0	126.5	80.0	19.71	65.0	68.0
4235	0	48	2.0	1	20.0	0.0	0	0	0	248.0	131.0	72.0	22.00	84.0	86.0

Fig. 2. Methods of data cleaning

```
dd=new_dataset.join(mm, lsuffix="_left", rsuffix="_right")
dd
```

	0	1	2	3	4	5	male	BPMeds	prevalentStroke	prevalentHyp	diabetes	TenYearCHD
0	0.149406	0.106383	0.232804	0.277024	0.104520	0.184211	1	0.0	0	0	0	0
1	0.242784	0.177305	0.349206	0.319680	0.101695	0.368421	0	0.0	0	0	0	0
2	0.234295	0.208038	0.338624	0.237518	0.084746	0.421053	1	0.0	0	0	0	0
3	0.200340	0.314421	0.497354	0.316045	0.177966	0.763158	0	0.0	0	1	0	1
4	0.302207	0.219858	0.380952	0.183228	0.127119	0.368421	0	0.0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
4233	0.349745	0.451537	0.465608	0.252787	0.129944	0.473684	1	0.0	0	1	0	1
4234	0.169779	0.203310	0.338624	0.101066	0.079096	0.500000	1	0.0	0	0	0	0
4235	0.239389	0.224586	0.253968	0.156568	0.129944	0.421053	0	0.0	0	0	0	0
4236	0.174873	0.203310	0.412698	0.087736	0.107345	0.315789	0	0.0	0	0	0	0
4237	0.275042	0.236407	0.370370	0.143723	0.189266	0.526316	0	0.0	0	0	0	0

4238 rows × 12 columns

Fig. 3. Cleaned dataset after data integration

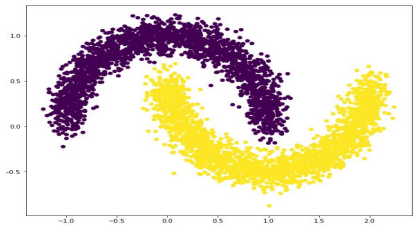

Fig. 4. DBSCAN

points within the neighborhood is required.

After finishing the process, start afresh with a new, unvisited spot, which will result in the identification of other clusters or noise, and each point is labeled as a cluster or noise at the end of the process.

d) Data Reduction: Conducting intricate investigations and scouring large data sets for pertinent information can be a lengthy process, thus making it unfeasible or out of the question to do so in an immediate manner. Information reduction methods are beneficial for studying the lessened portrayal of an informational collection without sacrificing the credibility of the initial data but instead supplying subjective data. The following procedures for data reduction are included:

- Conglomeration tasks are applied to the information in producing an information solid shape in Data Cube Aggregation.
- Superfluous, pitifully significant, or repetitive credits or measurements can be identified and deleted using dimension reduction.
- Encoding components are used in Data Compression to reduce the size of the informational collection. Wavelet Transform and Principle Component Analysis are two data compression algorithms used.
- In Numerosity Reduction, information is supplanted or assessed by elective and more modest information portrayals like parametric models, which store just the model boundaries rather than the genuine information.

C. Feature Selection

Using the Classifier Subset Evaluator, a highlight extraction approach was utilized to degree the exactness of these subsets for all utilized classifiers utilizing prepared classification information. This method is employed to decrease the number of attributes. Reducing the redundant attributes would reduce time and computational complexity. Then, it selects the attributes with a strong relationship with the target attribute. Feature Selection involves identifying and selecting the most relevant characteristics from a given dataset. The model that exploited the informative subset to the point where a classification model constructed just on this subset would have greater predicted accuracy.

D. Data Splitting

For this configuration, the UCI dataset has been divided into a training set that comprises 80% of the data and a testing set that comprises 20% of the data. The testing set aims to evaluate the models, while the training set is used in order to train them.

E. Machine Learning Algorithm

Various machine learning methods such as SVM, KNN, DTC, RFC, GNB, BNB, and LR are used to classify cardiac illness. In addition, these approaches are used to classify heart disease.

- K- Nearest Neighbor (KNN): It is a nonparametric algorithm that allows for the prediction of fresh sample classification. It can be used in both regression and classification forecasting situations. Using this model in this study, we found an accuracy of 87.15%.
- Support Vector Machine (SVM): It is employed to locate an appropriate hyperplane to classify the dataset with the greatest possible margin inside the data. In this investigation, we obtained an accuracy of 86.91% by using this model.
- Logistic Regression (LR): The binary form of a target variable is predicted using logistic regression, a supervised learning process. It can be applied to various situations, such as disease prediction and cancer detection. Using this model in this study, we reached an accuracy of 86.91%.
- Random Forest (RF): According to RF, a ”random decision forest” is a machine learning (ML) technique that can address classification and regression problems. Using this model in this study, we reached an accuracy of 86.32%.
- Naive Bays (NB): The Bayes theorem is applied with the unambiguous assumption of (naive) independence between characteristics in the NB family of fundamental probabilistic classifiers. Using this model in this study, we reached an accuracy of 86.78%.

F. Summary

First and foremost, we collected data from many sources and cleaned that data. After that, we pre-processed that data and finally created the model using different supervised machine learning algorithms: RFC, DTC, SVM, LR, NB, and KNN. All of these algorithms were used to create the model, and finally, we trained the model and tested it based on their requirements.

IV. Experimental Evalution

First of all, we divided that data into two different parts of the dataset: one was for training, and another was for testing, at 80% and 20%, respectively. After that, we experimented with the entire model in that training dataset, and then by testing the dataset, we got the result of the accuracy of those models.

A. Confusion Matrix

A disarray lattice is a method of presenting the AI arrangement calculations. We can compute Accuracy, Recall, Precision, Error Rate, and F1 Score by utilizing the disarray grid. Classification accuracy can be deceiving if the classes in

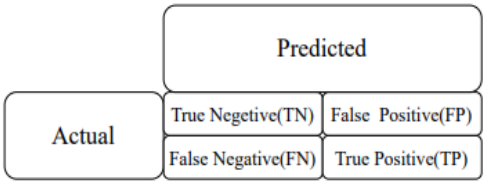

Fig. 5. Confusion Matrix

the dataset have different numbers of observations or if there are more than two classes. Calculating the confusion matrix shown in "Fig. 5" might help to determine which categorization models are correct and which are making mistakes. In the test dataset, a prediction was made for each row. It counts the number of correct and erroneous guesses for each class, grouped by the predicted class from the expected outcomes and forecasts.

B. Accuracy Calculation

From the confusion matrix, four values of predicted results are found. Based on those values (TN, TP, FN, and FP), we have calculated the Accuracy, Recall, Precision, F1 score, and Error rate of the implemented algorithms. The Accuracy can be defined using the "(1)".

$$Accuracy = (TP + TN)/(FP + FN + TP + TN) \quad (1)$$

C. Recall

The recall is the proportion of our model effectively distinguishing True Positives, as shown in "(2)". In this manner, for every one of the patients who has a coronary illness, the review reveals to us the number of those we effectively recognized as having a coronary illness.

$$Recall = TP/(TP + FN) \quad (2)$$

D. Precision

In the least complex terms, Precision calculates the proportion between the True Positives and every one of the Positives by the "(3)". For our difficult assertion, that would be the proportion of patients that we accurately distinguish as having a coronary illness out of the multitude of patients who have it.

$$Precision = TP/(FP + TP) \quad (3)$$

E. F1 Score Measure

F1 score is the standard for exactness and evaluation. As a result, both a bogus negative and a bogus positive action were taken. Precision is less useful than F1 score. Precision is better if both fictional positive and imaginary negative expenses are basically the same, but in all other cases, F1 score is fair. Also, DTC has the highest F1 score of about 22.9% and SVM has the worst at about 1.7%. These values are obtained by using the "(4)".

$$F1Score = (Recall * 2 * Precision)/(Recall + Precision) \quad (4)$$

F. Error Rate Calculation:

The error rate (ERR) is obtained by dividing the sum of incorrect predictions (FN + FP) by the total no of observations (P + N). The lower the error rate, the better the performance of the procedure. The K-nearest neighbor was found to have the lowest error rate of 12.8%, while the Decision Tree Classifier had the highest at 22.1%. The second-highest error rate of SVM and Logistic Regression is 13%. Numerically it is shown in "(5)".

$$ERR = (FN + FP)/(FP + FN + TP + TN) \quad (5)$$

In Table II, the implemented models are analyzed by the value of Accuracy, Recall, Precision, F1 Score, and Error Rate. Among all models, KNN predicted heart diseases with higher accuracy.

G. Area Under The Curve(AUC)

The AUC evaluates the performance of a model in distinguishing between positive and negative classifications, with a higher AUC representing better performance, the AUC value is 0.51 as shown in "Fig. 6".

H. Performance Comparison of the methods

The F1 score balances precision and recall on the positive class, whereas accuracy considers correctly identified positive and negative observations. RFC has required the highest precision at 37%, and LR has achieved the second position at 33.3%. KNN has achieved the lowest precision at zero. Also, DTC shows the highest recall at 25.6%, and SVM has the second highest recall at 0.91%. On the contrary, KNN has 0%.

While KNN gives the best accuracy, other performances in the recall, precision, and F1 score are about zero. DTC has the highest F1 score of about 22.9%, and SVM has the worst score of about 1.7%. The error rate of the K-nearest neighbor is the lowest at 12.8%, whereas DTC has the worst at 22.1%. To compare all of the performances, KNN is the best classification algorithm, and Support Vector Machine is the second-best

TABLE II
ACCURACY, RECALL, PRECISION, F1 SCORE, ERROR RATE

No.	Method	Accuracy	Recall	Precision	F1 Score	Error Rate
1	SVM	86.91	0.009	0.250	0.017	0.130
2	DTC	77.831	0.256	0.207	0.229	0.221
3	RFC	86.32	0.092	0.370	0.147	0.136
4	GNB	82.78	0.156	0.239	0.188	0.172
5	LR	86.91	0.018	0.333	0.035	0.130
6	BNB	82.78	0.155	0.239	0.189	0.172
7	KNN	87.15	0.0	0.00	0.00	0.128

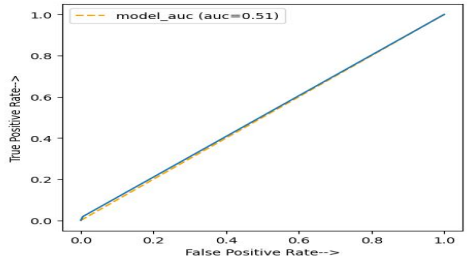

Fig. 6. True Positive Rate VS False Positive Rate

classification algorithm in this study. The comparison curve is shown in "Fig. 7".

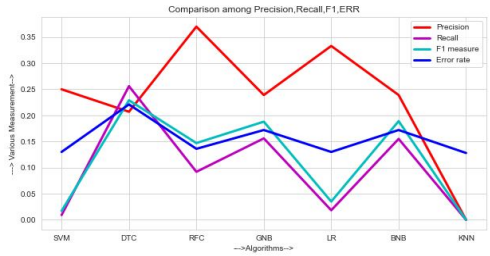

Fig. 7. Comparison curve of several techniques

This study has approximately 4200 data points and 16 qualities. After analyzing the data, we discovered that heart disease ailment affects just 644 individual groups for various reasons, whereas 3594 people are unaffected. For these reasons, we found that the primary causes are smoking, hypertension, diabetes, and stroke. The combined influence of these causes is most responsible for that. Finally, an outline of this breakdown claims that these major reasons account for the cause of heart disease.

V. Conclusions

In this work, a model was developed by us utilizing supervised machine learning algorithms and also found the vulnerable factor for this disease. We have the result that using the previous dataset that makes a decision for clinics and medical centers to predict heart disease based on the previous dataset by machine learning algorithms. A patient can also make a decision before real-life tests and experiments, and based on the previous assumption, we can be sure whether they are affected or not. In this work, the most and least important factors for heart disease are discussed, and their combined effects are for heart disease. Finally, we have examined which machine learning algorithm works well based on the used dataset.

VI. Future Work

Our research has demonstrated that supervised machine learning algorithms can be utilized to forecast heart disease and its associated risk factors from a dataset. From now on, we could investigate the effectiveness of more sophisticated machine learning techniques, such as deep learning, to refine the accuracy of the heart disease prediction procedure. Additionally, we would experiment with distinct datasets, as this project is an actual-world experiment and clinical tests using various instruments. Consequently, we plan to acquire real-time data. Moreover, we will try to acquire an image-based dataset to predict the heart disease of a patient.

References

[1] M. J. Ganesh, S. Madhuranjini, J. M. Monica, P. Renuka, and C. Priyanka, "Predict sympathy infection using naive bayesian algorithm," *International journal of engineering research and technology*, vol. 6, 2018.

[2] "Machine learning." https://www.coursera.org/learn/machine-learning.

[3] G. Guttikonda, S. Cherukuri, C. N. Sravanthi, M. Irfanullah, and M. Korlapati, "Prediction of heart disease and strategic decision making for phi of medical dataset," *International Journal of Latest Trends in Engineering and Technology*, vol. 8, pp. 045-050, 2018.

[4] P. Singh, S. Singh, and G. S. Pandi-Jain, "Effective heart disease prediction system using data mining techniques," *International Journal of Nanomedicine*, vol. 13, pp. 121 – 124, 2018.

[5] "Heart disease description." https://www.cdc.gov/heartdisease/about.htm.

[6] X.-Y. Gao, A. Amin Ali, H. Shaban Hassan, and E. M. Anwar, "Improving the accuracy for analyzing heart diseases prediction based on the ensemble method," *Complexity*, vol. 2021, 2021.

[7] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," *The American Journal of Cardiology*, vol. 64, no. 5, pp. 304-310, 1989.

[8] J. Gennari, P. Langley, and D. Fisher, "Models of incremental concept formation," Artif. Intell., vol. 40, pp. 11–61, 1989.

[9] J. A. Asuncion and D. Newman, "Uci machine learning repository," 2007.

[10] B. Kaur and W. Singh, "Review on heart disease prediction system using data mining techniques," *International journal on recent and innovation trends in computing and communication*, vol. 2, no. 10, pp. 3003–3008, 2014.

[11] C. B. C. Latha and S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," *Informatics in Medicine Unlocked*, vol. 16, p. 100203, 2019.

[12] P. Rubini, C. Subasini, A. V. Katharine, V. Kumaresan, S. G. Kumar, and T. Nithya, "A cardiovascular disease prediction using machine learning algorithms," *Annals of the Romanian Society for Cell Biology*, pp. 904–912, 2021.

[13] S. Nashif, M. R. Raihan, M. R. Islam, and M. H. Imam, "Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system," *World Journal of Engineering and Technology*, vol. 6, no. 4, pp. 854–873, 2018.

[14] K. Phasinam, T. Mondal, D. Novaliendry, C.-H. Yang, C. Dutta, and M. Shabaz, "Analyzing the performance of machine learning techniques in disease prediction," *Journal of Food Quality*, vol. 2022, 2022.

[15] M. A. Kadam, S. Patil, P. Pethkar, R. Shikare, and S. Sarnayak, "A cardiovascular disease prediction system using machine learning," *Journal of Pharmaceutical Negative Results*, pp. 7216–7225, 2023.