Improving Performance of Heart Disease Prediction with Bayesian Network and Feature Reduction

Usman Abdullahi Musa and Muhammad Sirajo Aliyu

Computer Science Department
Federal University Dutse – Nigeria

Correspinding Author: usman.m@fud.edu.ng

Abstract
Cardiovascular diseases are disease affecting the general well-being of the heart. It also causes severe morbidity and disability in people. The usage of electronic health record (EHR) systems has increased the amount of healthcare data available for analysis and forecasting. The need to make accurate predictions of heart disease through the use of machine learning algorithms is as a result of many factors the human mind cannot process. Several machine learning methods, including Random Forest, Logistic Regression, Artificial Neural Network (ANN), K-Nearest Neighbor, Support Vector Machine (SVM), have been applied on Cleveland heart datasets however, not very much was done on modeling with a Bayesian Network (BN). This study used the widely used Cleveland heart data collected from the UCI repository. Different feature reduction techniques where used and Bayesian Networks is used to make prediction on the reduced dataset. The results show that using feature reduction approaches improves the classifier's prediction performance. The proposed approach was 89% accuracy.

Keywords— Machine Learning, Bayesian Network , Naïve Bayes, Heart Disease

INTRODUCTION

Cardiovascular diseases are conditions that influence the constructions or capacity of your heart, for example, Abnormal heart rhythms/arrhythmias, Aorta infection, and Marfan disorder, Congenital coronary illness, Heart assault, Heart disappointment, Heart muscle sickness (cardiomyopathy) and Stroke, and so on [1]. Millions of people die annually from Noncommunicable Diseases accounting for about 63% of all global deaths. Low- and middle-income countries are known to bear 86% of the number of these premature deaths giving rise to estimated cumulative economic losses of US$7 trillion over the next 15 years[2].

The increase in the amount of health data gathered through the electronic health record (EHR) systems makes the use of strong analysis tools necessary. Numerous machine learning algorithms such as Random Forest, Logistic Regression, ANN, K-Nearest Neighbor, SVM, etc. have been applied on different heart disease dataset to make predictions.

The paper is organized as follows: Section 2 describes the previously used Machine Learning techniques used in this study. Section 3 presents the methodology, Section 4 summarizes the findings, and Section 5 concludes.

BACKGROUND OF THE RESEARCH

Several research have been done in relation to heart disease in machine learning, some have work on predicting the early stage of heart disease [3]–[5], increasing the efficiency of diagnosis and reduce the misclassification cost [6] and on the relationship between certain factors and to the presence of heart disease[7] as well as enhancing the selection of features(attributes) that are in causal relationship with presence of heart disease[8].

In a paper by [7] a probabilistic graphical model was used to understand the causal relationship among attributes of the Cleveland heart disease dataset from University of California Irvine (UCI) to predict heart disease using Bayesian network.
In another research by [9] AdaBoost, Bagging, Random Forest, and Voting Ensemble (Decision Trees, Logistic Regression and Support Vector Machines) were used to analyze different heart disease prediction system models for designing an automated medical diagnosis.

In another paper by [10], Convolutional Neural Network (CNN) was used to make heart disease prediction. Convolutional network is very useful in the field of image recognition. A Convolutional Neural Network (CNN) is known to provide better results than other machine learning algorithms if it is better tuned and fed a large amount of data. so much processing time if the computer doesn't have a good GPU, to the point where it takes a large dataset to process and train the network.

Many researches used different machine learning models on Cleveland dataset to predict heart disease, however, not very much was done on modeling with a Bayesian Network (BN). Furthermore, previous research used different datasets with many factors or input features which often make predictive modeling task more challenging to model. The existence of an enormous number of features, a learning model tends to overfit, leading to degradation of performance of the learning model[11][12]. To resolve these aforementioned issues, the proposed method in this study is the use of wrapper feature selection technique as a dimension reduction technique for extracting important features and the use of Naïve Bayes, Bayesian Network, KNN, and Logistic Regression to make predictions. The performance of all the models is also measured to make a comparison.

METHODOLOGY

The research work purposely focuses on predicting heart disease through feature reduction and the use of a machine learning model currently receiving attention in healthcare called Bayesian Network model. The methodology in this research in the study consist of some number of phases depicted in Figure 1 below. The phases include; data collection, data pre-processing, dimensionality /feature reduction, data training and testing for dimensionality/feature reduction, model testing, comparison of different feature reduction classification performance.

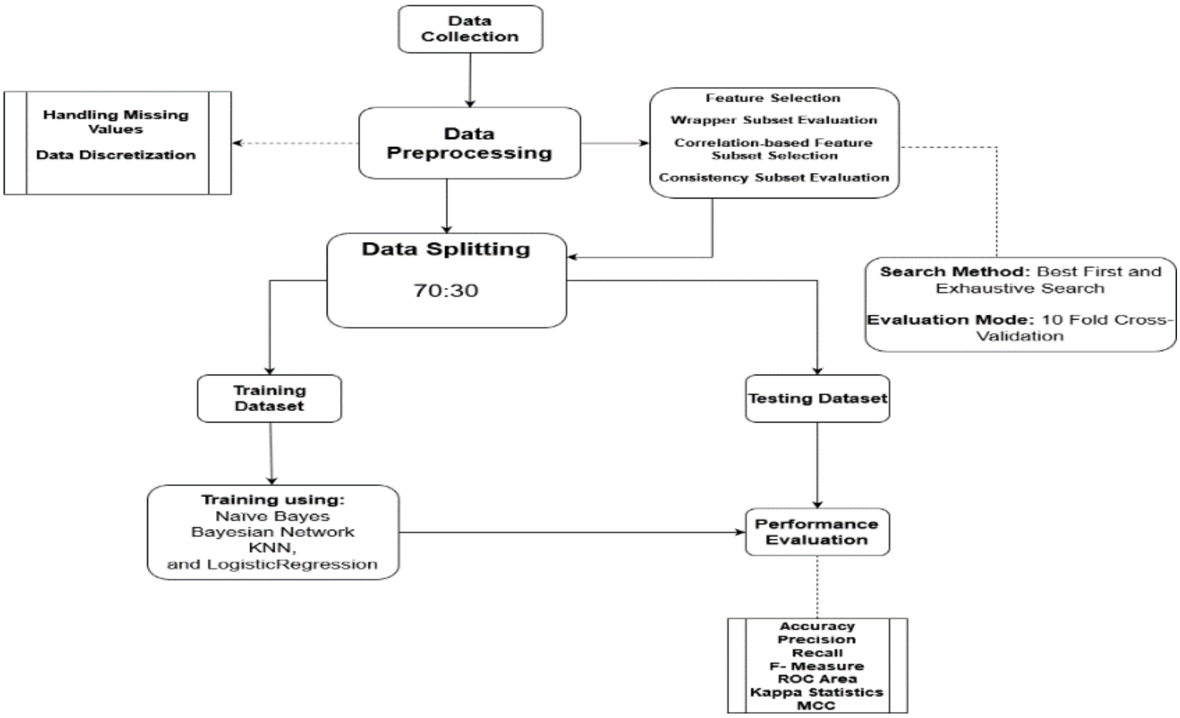


Figure 1: System architecture

Attribute	*Description*	*Type of Attribute*	*Attribute Value Range*
Age	Age in years	Numeric	29 to 77
Sex	Gender	Nominal	0 = female, 1 = male
cp	Chest pain type	Nominal	1 = typical angina, 2 = atypical angina, 3 = non-angina pain, 4 = asymptomatic
trestbps	Resting blood pressure in mm Hg on admission to the hospital	Numeric	94 to 200
chol	Serum cholesterol in mg/dL	Numeric	126 to 564
fbs	Fasting blood sugar > 120 mg/dL	Nominal	0 = false, 1 = true
restecg	Resting electrocardiographic results	Nominal	0 = normal,1 = ST-T wave abnormality, 2 = definite left ventricular hypertrophy by Estes' criteria
thalach	Maximum heart rate achieved	Numeric	71 to 202
exang	Exercise induces angina	Nominal	0 = no 1 = yes
oldpeak	ST depression induced by exercise relative to rest	Numeric	0 to 6.2
slope	The slope of the peak exercise ST segment	Nominal	1 = upsloping, 2 = flat, 3 = down sloping
ca	Number of major vessels colored by fluoroscopy	Nominal	0–3
thal	The heart status	Nominal	3 = normal, 6 = fixed defect, 7 = reversible defect
num	Prediction attribute	Nominal	0= Unlikely to obtain heart disease 1= Likely to obtain heart disease

Table I: Cleveland dataset attributes

Dataset Collection

This research used the heart disease dataset from the UCI machine learning repository called Cleveland Heart Disease Dataset. Cleveland Heart Disease Dataset is a publicly available supervised dataset provided by the Cleveland Clinic Foundation was used for the ML model. The dataset used in this research is shown in Table I below:

A. Data Preprocessing

Data preprocessing is also known as cleaning data. It is one of the most important steps to achieve the best from the dataset. This is a process whereby data inconsistencies such as missing values, out of range values, unformatted data, and noise are removed from the data. Our preprocessing would involve handling missing values, and data discretization

1) Handling Missing Values

Missing data values is a common problem faced by analysts. This occurs due to different reasons such as incomplete extraction, corrupt data, failure to load the information, etc. These ways of handling missing data include; deleting rows, replacing with mean/median/mode, assigning a unique category, predicting the missing values and using algorithms which supports missing values. This research would adopt the method of handling missing values that proves best.

2) Data Discretization

Data discretization is the procedure of changing continuous data attribute values into a limited set of intervals with minimum information loss[13]. Discretization can help improve significantly the classification performance of some as algorithms like Naïve Bayes that are sensitive to the dimensionality of the data[14].

Features Selection Techniques and Search Method

Feature Selection is considered part of data preprocessing. The process determines a smaller subset with nearly equal predictive ability [15]. As mentioned by[16], feature selection enhances data quality and accuracy of data mining algorithms, decreasing space and time complexity. In order to search for optimal feature list using feature selection technique, different search methods are used, these methods include; Best first search, Exhaustive search, evolutionary search, etc. This paper focuses three different feature subset evaluation techniques with best first search and exhaustive search methods. The best-First search method searches the space of attribute subsets by greedy hill climbing augmented with a backtracking facility[17] while the exhaustive search Performs an exhaustive search through the space of attribute subsets starting from the empty set of attributes[18]. These feature selection techniques used in the study are:

i. Wrapper Subset Evaluation
ii. Correlation-based Feature Subset Selection
iii. Consistency Subset Evaluation

i. Wrapper Subset Evaluation

Wrapper model approach uses the method of classification itself to measure the importance of features set; hence the feature selected depends on the classifier model used. Wrapper methods generally result in better performance than filter methods because the feature selection process is optimized for the classification algorithm[19].

ii. Correlation-based Feature Subset Selection

CFS is basically a filter approach that appraises the merits of subset attributes by classifying the feature ability according to the amount of redundancy between them and the feature subset selection. The CFS is a filter approach that removes the irrelevant features having high correlation with class and redundant features having high correlation with remaining features to obtain the highest merit feature subsets [12].

iii. Consistency Subset Evaluation

The consistency-based subset evaluation method generates a random subset, S, from the feature subset space (N) in every round of the process. If the number of features (C) contained by S is less than the current best subset, the inconsistency rate of data prescribed in S is checked against the existing best subset[20].

B. Machine Learning Models used in the Research

i. Naive Bayes

Naïve Bayes or stupid Bayes is used to handle binary (two-class) and multiclass classification challenges, It has its name because the probabilities for each hypothesis are simplified to make its calculation tractable [21]. This is expressed mathematically as:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

Where P(A) is the prior distribution of parameter A, P(A|B) denotes the posterior distribution, which denotes the probability of A given new data B, and P(B|A) denotes the likelihood function, which denotes the probability of B given existing data.

ii. Bayesian Network

A Bayesian network B = < N, A,θ> is a directed acyclic graph (DAG) <N, A> with a conditional probability distribution (CP) for each node, collectively represented by θ. Each arc a ∈ A between nodes represents a probabilistic dependency, and each node n ∈ N represents a domain variable. In general, a BN can be used to compute the conditional probability of one node, given values assigned to the other nodes; hence, a Bayesian Network can be used as a classifier to calculate the posterior probability distribution of a classification node given the values of other characteristics.[22]. A Bayesian network, for example, might show disease-symptom relationships. The network can predict the existence of illnesses given a collection of symptoms.

iii. Logistic regression

The logistic function, also known as the sigmoid function, was created by statisticians to characterize the properties of population increase in ecology, such as how it rises swiftly and eventually reaches the environment's carrying capacity. LR still considers the dependent variable to be a bi-categorical variable. It is mostly used to forecast and calculate the likelihood of success. Molding the equation into the form of needed data entry is also part of LR. A basic equation is used here:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \cdots \beta_n X_n \qquad (2)$$

iv. K-Nearest Neighbour

KNN makes predictions using the training dataset directly. For each new data point, predictions are formed by exploring the whole training set for the k most similar examples (neighbors) and summing the output variable for those k instances. The most similar of the k instances in the training dataset to a new input is determined using a distance metric. The most frequent distance measure for real-valued input variables is Euclidean distance. The square root of the total of the squared discrepancies between two points a and b across all input qualities i is used to calculate Euclidean distance(E.D).

$$\mathrm{E.D}(a, b) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \qquad (3)$$

C. PERFORMANCE METRICS USED IN THE RESEARCH

Performance metrics are used to evaluate how different algorithms perform based on various criteria such as accuracy, precision, recall, etc. They are discussed below.

1) Accuracy

Accuracy is the ratio of the number of correctly classified instances to all the cases. It is the sum of TP and TN divided by the total number of instances.

$$Accuracy = \frac{\mathrm{TP + TN}}{\mathrm{TP{+}TN{+}FP{+}FN}} \qquad (4)$$

2) Precision

Precision is the proportion of true positive instances that are classified as positive. It shows how near the projected values are to each other.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP{+}FP}} \qquad (5)$$

3) Recall/ Sensitivity

The recall is the proportion of positive examples that are correctly classified as positive. Recall is known as sensitivity.

$$\mathrm{Sensitivity} = \frac{\mathrm{TP}}{\mathrm{TP + FN}} \qquad (6)$$

4) F1 Score

F1 score combines both precision and recall and finds a balance between both. In other words, it computes the harmonic mean of precision and recall.

$$\mathrm{F - measure} = \frac{2 * \mathrm{Precision + Recall}}{\mathrm{precision + Recall}} \qquad (7)$$

5) MCC

MCC known as Mathew correlation coefficient. Mathematically, the Pearson product-moment correlation coefficient between actual and expected values is calculated using a contingency matrix range in the interval [−1, +1], with extreme values −1 and +1 reached in case of perfect misclassification and perfect classification, respectively.

$$\mathrm{MCC} = \frac{\mathrm{TP{\times}TN{-}FP{\times}FN}}{\sqrt{((\mathrm{TP{+}FP}){\times}(\mathrm{TP.FN}){\times}(\mathrm{TN{+}FP}){\times}(\mathrm{TN{+}FN}))}} \qquad (8)$$

6) AUROC curve

It's a graph depicting the ratio of false positives to real positives. The area assesses discrimination, or the classifier's ability to accurately classify the test data.

7) Kappa Statistics

The kappa measure of agreement is the ratio

$$K = \frac{P(A) - P(E)}{1 - P(E)} \quad (9)$$

Where P(A) denotes the percentage of times the k raters agree, i.e., the percentage agreement between the classifier and the ground truth, and P(E) is the proportion of times the k raters are expected to agree by chance alone i.e., the chance agreement. K=1 indicates perfect agreement and K=0 indicates chance agreement. The value greater than 0 means classifier is doing better. The classifier's result improves as the kappa statistic value rises.

IV. RESULTS AND DISCUSSION

A. RESULTS

In this section, 3 different feature selection method with the classification models results and strategies are presented and discussed. For feature selection, wrapper method, Consistency Subset Evaluation and Correlation-based Feature Subset Selection were used with best first search and exhaustive search method. To determine how training and testing is to be done, on 70:30 train and test split ratio of the dataset on the selected models as it will be sufficient proportion due to the size of the dataset.

The result of performing the different feature reduction techniques to determine the selected attributes by each subset evaluation, is shown in table II presents the selected attributes(features) for each of the feature selection technique.

Table II: Attributes selected on feature reduction techniques

Feature Reduction Technique	Search Method	Features
Dataset Before reduction	-	Age, Sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope,ca, thal
Wrappers for feature subset selection	Best First	age, sex, cp, restecg, exang, oldpeak, slope, ca, thal
WrapperSubsetEval	Exhaustive Search	Sex, cp, fbs, restecg, exang, oldpeak, slope,ca, thal
Consistency Subset Evaluation	Best First	age, sex, cp, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal
Consistency Subset Evaluation	Exhaustive Search	age, sex, cp, fbs, restecg, thalach, exang, oldpeak, slope,ca, thal
Correlation-based Feature Subset Selection	Best First	sex, cp, restecg, thalach, exang, oldpeak, slope, ca, thal
Correlation-based Feature Subset Selection	Exhaustive Search	Sex, cp, restecg, thalach, exang, oldpeak, slope,ca, thal

In total for machine learning classifiers, 303 records were used each with 14 total attributes. As shown in Table II show the list of features selected using the three different feature selection techniques. Both wrapper feature selection and correlation-based feature selection

showed that 9 out of the 14 features played a part in determining the last feature or the heart disease diagnosis. Consistency subset evaluation showed that only two do not play much relevance in determining the last feature indicating diagnosis of heart disease.

In order to determine the feature selection technique with better performance, there is need to evaluate it on different metrices. Table III shows how the performance of prediction before the application of feature reduction on the dataset. It also shows the performance of the feature selection techniques used in the study with their various performances.

TABLE III: Performance of different feature reduction techniques

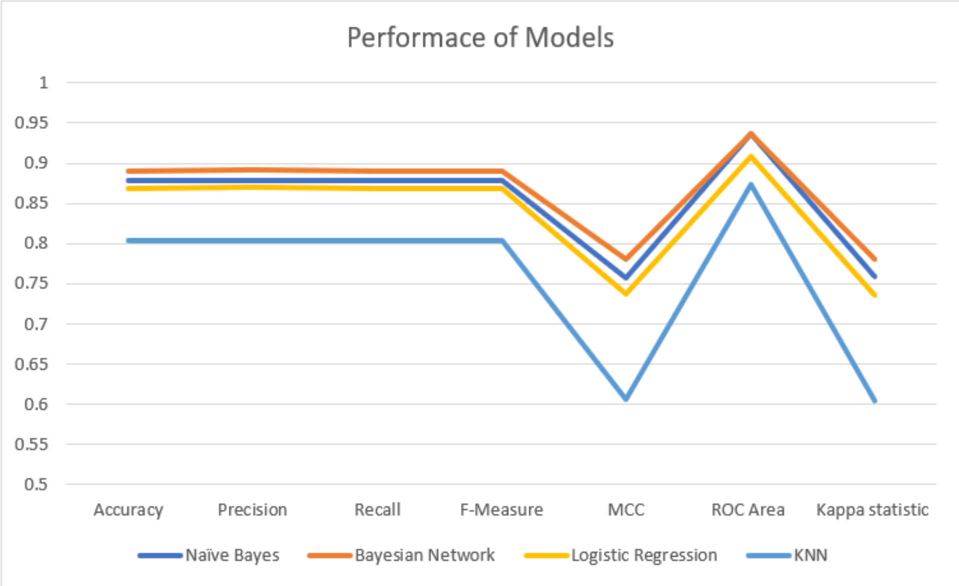
DATASET	PRECISION	RECALL	F-MEASURE	ROC AREA	ACCURACY
Dataset Before reduction	0.869	0.869	0.869	0.927	86.8852
Wrappers for feature subset selection with Best First	**0.891**	**0.89**	**0.89**	**0.936**	**89.011**
Wrappers for feature subset selection With exhausive search	0.868	0.857	0.862	0.928	85.7143
Consistency Subset Evaluation	0.857	0.857	0.857	0.925	85.7143
Consistency Subset Evaluation With exhausive search	0.857	0.857	0.857	0.925	85.7143
Correlation-based Feature Subset Selection	0.856	0.846	0.851	0.921	84.6154
Correlation-based Feature Subset Selection With exhausive search	0.856	0.846	0.851	0.921	84.6154

TABLE IV: Model(s) Performance

Model	**Accuracy**	***Precision***	***Recall***	***F-Measure***	**MCC**	**ROC Area**	**Kappa statistic**
Naïve Bayes	87.9121	0.879	0.879	0.879	0.758	**0.936**	0.7583
Bayesian Network	**89.011**	**0.891**	**0.89**	**0.89**	**0.781**	**0.936**	**0.7803**
Logistic Regression	86.8852	0.870	0.869	0.868	0.738	0.908	0.7362
KNN	80.3279	0.804	0.803	0.803	0.606	0.873	0.6043

Considering the wrapper feature selection with best first search method on those reduced attributes stated in table II have the highest performance, there is need to measurer and compare the model with the selected models on different metrics such as precision, recall, MCC score etc. Table III below shows the performance of Bayesian network, Naïve Bayes, KNN and logistics on 70:30 ration with different evaluation metrices.

Figure 2: Performance of the models used



Looking at the performance of the various models in table IV, each model had proved to have a good predictive power to heart disease However, looking at the ROC Area score, Bayesian Network is having a similar performance to that of Naive. However, on the other evaluation technique, Bayesian network has the best performance. Thus, it is deemed as the best algorithm out of the 4 tested.

B. DISCUSSION

In trying to extract the most relevant feature from the dataset and obtain the performance of each selection to the full dataset Table II gives the list of the attributes selected using different 3 different feature selection techniques i.e., Wrapper method, Correlation-based Feature Subset Selection and Consistency Subset Evaluation with the two search methods i.e., best fist search and exhaustive search. Consistency and correlation evaluation produce identical attribute selection for the two search ways since they are both filter methods of attribute selection, however the wrapper approach produces distinct attribute selection because the search technique for how the machine selects is different. To obtain their performance on different evaluation metrices (i.e., Accuracy, Precision, recall, F-Measure, MCC and ROC Area), Table III give the detail performance. This performance was modeled using Bayesian network which has been our model of choice. Ince the attribute selected for Correlation-based Feature Subset Selection and Consistency Subset Evaluation have virtually similar, the result performance of the is the same. The wrapper method with best first search had outperformed the other feature selection method s with 89.011% accuracy as shown in Table III. This performance of wrapper methods evidently shows feature reduction improving the performance of prediction models.

Table IV give the performance of Bayesian network over Naïve bayes, logistic regression and KNN. This shows Bayesian network having the highest performance inaccuracy (89.011%), precision (0.891), Recall (0.891), F-measure (0.891), MCC (0.781) Kappa Statistics (0.7803) and ROC Area (0.936). It is observed that the ROC Are of naïve bayes is similar to that of Bayesian network because both of them are probabilistic models and works almost similar.

V. CONCLUSION

Dimension reduction techniques used to improve the prediction of heart disease have proven to be effective. The method used in this study was the WEKA wrapper method of data selection to choose the best subset of features from the Cleveland dataset. This method was chosen because it was more accurate and efficient than other methods. The selected features are 9 in numbers and they include: Age, sex, cp, restecg, exang, oldpeak, slope, ca, thal. The proposed method in the study has been evaluated with various metrics, and its performance results are compared with explores different machine learning algorithms.
This research presents a thorough, effective, and highly preferred Machine Learning-based model that helps doctors recognize cardiac issues early so patients can take precautionary actions in a correction window. The paper used Naïve Bayes, Bayesian Network, KNN, and Logistic Regression on the reduced features. The same features are used to both train and test the dataset. The outcome reveals that these data mining techniques can predict heart disease early with an accuracy of approximately 89%.

ACKNOWLEDGMENT

The author(s) wish to acknowledge the kind support of Dr. Muhammad Sirajo Aliyu for providing the useful guidance in the research work.

REFERENCES

[1] S. R. Steinbaum, "Cardiovascular (Heart) Diseases: Types and Treatments," 2019. https://www.webmd.com/heart-disease/guide/diseases-cardiovascular (accessed Jun. 26, 2021).

[2] WHO, "WHO and Nigerian Government move to curb cardiovascular diseases | WHO | Regional Office for Africa," 2019. https://www.afro.who.int/news/who-and-nigerian-government-move-curb-cardiovascular-diseases (accessed Jun. 13, 2021).

[3] K. M. Almustafa, "Prediction of heart disease and classifiers' sensitivity analysis," *BMC Bioinformatics*, vol. 21, no. 1, pp. 1–18, 2020, doi: 10.1186/s12859-020-03626-y.

[4] 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, doi: 10.1155/2021/6663455.

[5] B. Usha Sri, "Effective Heart Disease Prediction Model Through Voting Technique," *Int. J. Eng. Technol. Manag. Sci.*, vol. 4, no. 5, pp. 10–13, 2020, doi: 10.46647/ijetms.2020.v04i05.003.

[6] L. Dandona *et al.*, "Nations within a nation: variations in epidemiological transition across the states of India, 1990–2016 in the Global Burden of Disease Study," *Lancet*, vol. 390, no. 10111, pp. 2437–2460, 2017, doi: 10.1016/S0140-6736(17)32804-0.

[7] M. Muibideen and R. Prasad, "A Fast Algorithm for Heart Disease Prediction using Bayesian Network Model," pp. 1–11, 2020, [Online]. Available: http://arxiv.org/abs/2012.09429.

[8] V. Jothi Prakash and N. K. Karthikeyan, "Enhanced Evolutionary Feature Selection and Ensemble Method for Cardiovascular Disease Prediction," *Interdiscip. Sci. Comput. Life Sci.*, no. 0123456789, 2021, doi: 10.1007/s12539-021-00430-x.

[9] J. Emakhu, S. Shrestha, and S. Arslanturk, "Prediction system for heart disease based on ensemble classifiers," *Proc. Int. Conf. Ind. Eng. Oper. Manag.*, no. August, pp. 2337–2347, 2020.

[10] A. Dutta, T. Batabyal, M. Basu, and S. T. Acton, "An efficient convolutional neural network for coronary heart disease prediction," *Expert Syst. Appl.*, vol. 159, 2020, doi: 10.1016/j.eswa.2020.113408.

[11] A. Shekhar, "What Is Feature Engineering for Machine Learning?," *Medium*, 2018. https://medium.com/mindorks/what-is-feature-engineering-for-machine-learning-d8ba3158d97a (accessed Oct. 02, 2021).

[12] R. N. Thomas and R. Gupta, "An efficient feature subset selection approach for machine learning," *Multimed. Tools Appl.*, vol. 80, no. 8, pp. 12737–12830, 2021, doi: 10.1007/s11042-020-10011-7.

[13] W. Ian, F. Eibe, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques*. 2011.

[14] J. L. Lustgarten, V. Gopalakrishnan, H. Grover, and S. Visweswaran, "Improving Classification Performance with Discretization on Biomedical Datasets," 2008.

[15] M. Mvurya, "Investigating Prediction Modelling of Academic Performance for Students in Rural Schools in Kenya," *J. Chem. Inf. Model.*, vol. 53, no. July, pp. 1–82, 2014, [Online]. Available: http://library.wur.nl/WebQuery/wurpubs/fulltext/353506.

[16] C. Arun Kumar, M. P. Sooraj, and S. Ramakrishnan, "A Comparative Performance Evaluation of Supervised Feature Selection Algorithms on Microarray Datasets," *Procedia Comput. Sci.*, vol. 115, pp. 209–217, 2017, doi: 10.1016/j.procs.2017.09.127.

[17] W. BestFirst, "BestFirst," *WekaDocs*, 2020. https://weka.sourceforge.io/doc.dev/weka/attributeSelection/BestFirst.html (accessed Mar. 19,

2022).

[18] E. Weka, "ExhaustiveSearch," *WekaDocs*, 2020. https://weka.sourceforge.io/doc.stable/weka/attributeSelection/ExhaustiveSearch.html (accessed Mar. 19, 2022).

[19] A. G. Karegowda, "Feature Subset Selection Problem using Wrapper Approach in Supervised Learning," no. August 2014, 2010, doi: 10.5120/169-295.

[20] O. Aytug, "A fuzzy-rough nearest neighbor classifier combined with consistency-based subset evaluation and instance selection for automated diagnosis of breast cancer," *Elservier*, vol. 42, pp. 6844–6852, 2015, doi: 10.1016/j.eswa.2015.05.006.

[21] B. Jason, "Naive Bayes for machine Learning," *Machine Learning Algorithms*. 2014, Accessed: Oct. 01, 2021. [Online]. Available: https://machinelearningmastery.com/naive-bayes-for-machine-learning/.

[22] J. Cheng, R. Greiner, J. Kelly, D. Bell, and W. Liu, "Learning Bayesian networks from data: An information-theory based approach," *Artif. Intell.*, vol. 137, no. 1–2, pp. 43–90, 2002, doi: 10.1016/S0004-3702(02)00191-1.