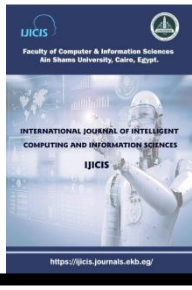
IJICIS, Vol.25, No.1, 29-40
DOI: 10.21608/ijicis.2025.363989.1377

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



UTILZING MACHINE LEARNING FOR HEART DIEASESE PREDICTION

Amna Badeen*

Department of Graduate Studies,
Faculty Information Technology, Al-Quds Open University,
Salfit, Palestine
amnabadeen@gmail.com

Thara' Salma

Department of Graduate Studies,
Faculty Information Technology, Al-Quds Open University,
Salfit, Palestine
tharaasalameh93@gmail.com

Nadia Abu AL-rob

Department of Graduate Studies,
Faculty Information Technology, Al-Quds Open University,
Jenin, Palestine
Nadiaabualrub5@gmail.com

Mohamed Dweib

Department of Technology & Applied Sciences,
Faculty of CIS, Al-Quds Open University,
Bethlehem, Palestine
mdweib@qou.edu

Received 2025-02-26; Revised 2025-02-26; Accepted 2025-03-12

Abstract: *In the past few years, the world has witnessed a significant increase in the prevalence of heart disease, which threatens the safety of people's lives, and it has become one of the most common diseases today. Therefore, it is necessary to find the best way to predict the disease in advance, because it may help people's lives.This paper aims to develop a system for predicting heart diseases that enables them to predict the probability of a person suffering from a heart disease in order to prevent him from it, based on the patient's medical record.We used different machine learning algorithms available in WEKA 3.8.1. Such as logistic regression, Naïve bayes, and decision tree J48 for classification of patients with heart disease. It includes 3 steps: First select data for 18 clinical features from a kaggle-like site BMI, Smoking, AlcoholDrinking, Stroke, PhysicalHealth...etc. Second, initialize the data, Thirdly, the development of the tree algorithm, Naive bayes, and logistic regression to predict heart disease based on clinical features.The accuracy of the logistic regression model was 92.1%. High compared to other algorithms that were used, as it was able to predict the possibility of heart disease in an individual. The heart disease prediction system provides health care to save human life, using appropriate medications.*

Keywords: *Heart disease, Prediction, Artificial Intelligence, Machine Learning Algorithms, Logistic Regression, Naive Bayes, decision tree J48.*

1. Introduction

***Corresponding Author**: Amna Badeen

Department of Graduate Studies, Faculty Information Technology, Al-Quds Open University, Salfit, Palestine

Email address: amnabadeen@gmail.com

Heart disease is one of the most serious challenges facing the world. According to the World Health Organization, it is the leading cause of death globally, causing 32% of deaths worldwide, and the death toll is 17.9 million deaths for the year 2021 [20]. Heart disease can sometimes go undetected until a person exhibits the signs or symptoms of a heart attack, heart failure, or arrhythmia, which can ultimately result in death [13]. Therefore, it's important to create a system that can predict heart diseases before they become complicated. This research presents the possibility of predicting heart disease using artificial intelligence algorithms by collecting data about the patient and some vital signs in the body related to the heart. This could help doctors diagnose heart disease in patients more quickly, increasing the likelihood that they can save their patients' lives. However, more research is required to use this knowledge to lower the incidence of heart disease. Recent medical research has identified risk factors that may contribute to the development of heart disease. Based on the health information of 319,795 adults obtained from the Kaggle website, the risk factors for heart disease were identified. These risk factors include body mass index, smoking, drinking alcohol, stroke, physical and mental health, difficulty walking, gender, age group, race, diabetes, physical activity, general health, sleep quality, asthma, kidney disease, and skin cancer [14]. We first chose the attributes, employed Weka's data mining tool, and then conducted the trials, where 80% of the data is used for training and the remaining 20% is used for test. Results of heart disease prediction experiments based on the three algorithms: Logistic Regression, Naïve Bayes, and Decision Trees J48. The most effective algorithm is (LR), which gives us an accuracy of 92.1% and identifies individuals as having a high or low risk of developing heart disease. It has been demonstrated that the use of data mining techniques in the healthcare sector produces faster and more accurate disease prediction [9]. Diagnostics is a difficult and crucial task that must be completed precisely and effectively [11]. The doctor's experience is frequently used to make the diagnosis. As a result, patients' therapies have expensive medical expenses and unfavorable outcomes. Consequently, a system for automated medical diagnosis would be quite beneficial.

2. Literature Review

Looking at many previous researches, a lot of work has been done to predict heart disease using machine learning datasets. Different levels of accuracy have been achieved using different data mining techniques, which are explained below.

Marwah et al, 2024 [6] discusses the significance of feature selection and data preprocessing in improving ML model performance. It emphasizes the potential for machine learning to enhance risk assessment and treatment recommendations by analyzing historical patient data.
Nadikatla et al, 2023 [10] implemented six major ML algorithms, including K-nearest neighbors and gradient boosting, on established heart disease datasets. By employing hyper parameter tuning and ensemble methods, the researchers aimed to maximize prediction accuracy, indicating a trend towards more sophisticated model training techniques.

Chang et al, 2022 [2] developed research aimed at detecting heart disease using the Python programming language. They found that VLRAKN had the highest accuracy compared to algorithms such as Naive Bayes, Random Forest Classification, Logistic Regression, and Decision Trees. The models were compared in terms of precision, specificity, f-measure, accuracy and sensitivity.

Shafique et al, 2015 [16] used Decision tree, Artificial neural network (ANN), Naive Bayes algorithms for experiments. Naive Bayes shows the highest accuracy 82.914%, while the Decision tree scores the lowest accuracy 77.219%, and Artificial neural network is nearest to it with 82.077% accuracy.

Rajdhan et al, 2020 [15] predict the chances of Heart Disease and classifies patient's risk level by implementing different data mining techniques such as Naive Bayes, Decision Tree, Logistic Regression and Random Forest. The trial results verify that Random Forest algorithm has achieved the highest accuracy of 90.16% compared to other ML algorithms implemented.

Pandey et al, 2020 [12] made an efficient heart disease prediction by using various algorithms to discover the heart diseases through five classification model as Naïve Bayes, SMO (Support Vector Machine), IBK (K-nearest Neighbor), J48(C4.5 Decision Tree), PART (Projective Adaptive Rejunance Theory). After making comparison it has been observed that classifier J48 provides good result for Heart diseases. The accuracy of J48 is 97.11% while the percentage of incorrectly classified instances is 3% (21 instances).

In the research of Gultepe & Rashed, 2021 [3] tried to predict heart disease from different attributes in heart-c.arff dataset which obtained from UCI against several data classification techniques using Weka software. After applying Bagging with J48, they successfully reached to an accuracy of 81.31%.

In these research Almustafa, 2020 [1] did a comparative analysis of different classifiers was done for the classification of the Heart Disease dataset. They used about 10 algorithms for the classification of the HD dataset produced very promising results. Where the Naïve Bayes algorithm excels with accuracy of classification of 99.7073.

3. Materials and Methods

This section provides information about the data and techniques used in the training phase of the algorithms and classification model used for prediction, feature selection and performance metrics used in this paper.

3.1 Dataset

Using an open source machine learning tool called Weka, it has tools for data pre-processing, classification, regression, clustering and correlation rules and visualization [17,18] instance in the data set is named HeartDisease (target). The HeartDisease attribute is labeled HeartDisease=Yes respondent had heart disease Heart disease= No respondent had no heart disease. Table 1 presents all the details about the dataset.

3.2 Data Mining Models

Preparation and pretreatment

Once the data is available, it needs some pre-processing data, it is the step before applying the data mining algorithm, it transforms the original data into a suitable shape to be used by a particular mining algorithm, and the significance of the output depends on the pre-processing of the data. Data pre-processing includes different tasks as data cleaning, data transformation, and feature selection, so running the model on the selected features makes prediction easy for the model and also saves time, so is a necessary step for effective and real- world data mining to increase the accuracy of the mining [19] .
The data goes through several stages in the mining process, as it first begins with data entry in the filtering and mining forms, and then it goes through the second stage, which is the processing process. The third

is the analysis process and it is analyzed by the algorithm. The algorithm calculates a set of summary statistics that describe the data, and identifies the rules and patterns within the data. The mining model contains a wealth of information about the existing data and patterns through analysis and finally passes the verification process as shown in figure 1.

Table 1 Description of the Heart Disease dataset.

#No	Attribute	Description	Type
0	HeartDisease	Respondents that have ever reported having coronary heart disease (CHD) or myocardial infarction (MI)(yes or no)	Nominal
1	BMI	Body Mass Index (BMI)	Numeric
2	Smoking	Have you smoked at least 100 cigarettes in your entire life? [Note: 5 packs = 100 cigarettes] (yes or no)	Nominal
3	AlcoholDrinking	Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week (yas or no)	Nominal
4	Stroke	(Ever told) (you had) a stroke? (yes or no)	Nominal
5	PhysicalHealth	Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30	Numeric
6	MentalHealth	Thinking about your mental health, for how many days during the past 30 days was your mental health not good?	Numeric
7	DiffWalking	Do you have serious difficulty walking or climbing stairs? (yes or no)	Nominal
8	Sex	Are you male or female?	Nominal
9	AgeCategory	Fourteen-level age category (65-69 60-64 ,Other)	Nominal
10	Race	Imputed race/ethnicity value (White Hispanic, Other)	Nominal
11	Diabetic	(Ever told) (you had) diabetes? (yes or no)	Nominal
12	PhysicalActivity	Adults who reported doing physical activity or exercise during the past 30 days other than their regular job (yes or no)	Nominal
13	GenHealth	Would you say that in general your health is... (Very good,Good,Other)	Nominal
14	SleepTime	On average, how many hours of sleep do you get in a 24-hour period?	Numeric
15	Asthma	(Ever told) (you had) asthma? (yes or no)	Nominal
16	KidneyDisease	Not including kidney stones, bladder infection or incontinence, were you ever told you had kidney disease? (yes or no)	Nominal
17	SkinCancer	(Ever told) (you had) skin cancer? (yes or no)	Nominal

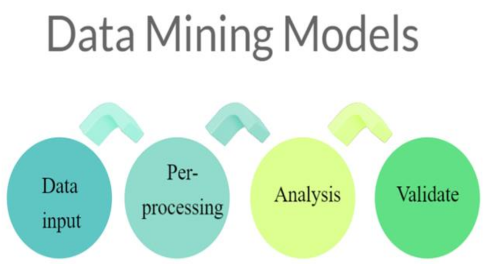


Figure 1: Data mining Model

3.3 Missing Values

There are no missing values in the data.

3.4 Filters Remove Folds

This filter takes a dataset and outputs a specified fold for cross validation. If you want the folds to be stratified use the supervised version. We have taken this step to reduce the number of data to save female processing time, so that the number of data becomes: 31980 weka. filters. unsupervised. instance. Remove Folds.

3.5 Feature Selection

Given the extensive number of features in the data, many of them may not contribute significantly to the analysis. Therefore, it is essential to eliminate these unimportant features from the dataset to enhance the accuracy of the algorithm. We employed supervised feature selection methods: filtering, casing, and intrinsicity. In our case, we adopted the filtering approach. During the exploration phase, the filtering approach was implemented, starting with the selection of an appropriate filter for the data. This filter then carried out the analysis process, identifying the features most closely associated with the target variable. Subsequently, the selected features were applied to the learning algorithms, and the results, along with the performance of each algorithm, are illustrated in figure 2.

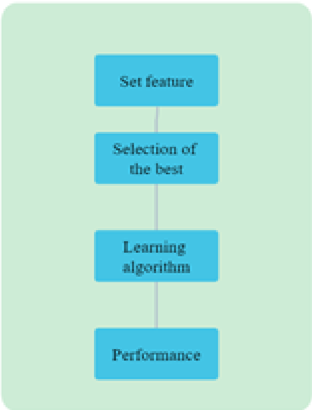


Figure 2: filtering approaches

The five filter approaches are applied to the dataset are CfsSubsetEval(CSE), GainRatio Attribute evaluation (GA), Information Gain Attribute evaluation (IG), Correlation Attribute (CA), and ReliefFAttributeEval. These five approaches are explained below:

- Correlation-based Feature Selection-CfsSubsetEval (CSE)

Evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. The importance of the traits is measured on the basis of the predictive ability of the traits and the degree of frequency to a subgroup that has a close association with the target group (heart disease). Where it was applied to the data, it showed six characteristics, which are as follows: Stroke, DiffWalking, AgeCategory, Diabetic, GenHealth, KidneyDisease.weka.attributeSelection.CfsSubsetEval.

- GainRatio Attribute evaluation (GA)

Evaluates the value of an attribute by measuring the gain ratio with respect to the class GainR (Class, Attribute) = (H(Class) - H (Class | Attribute)) / H(Attribute) Where H represents the Entropy It measures the importance of the characteristics in relation to the target group on the basis of the earning percentage. It was applied to the data set in the first order. Six characteristics are taken into account, namely:Stroke, Kidney Disease, DiffWalking, Diabetic, GenHealth, PhysicalHealth: weka. attributeSelection.GainRatioAttributeEval.

- Information Gain Attribute evaluation (IG)

Evaluates the worth of an attribute by measuring the information gain with respect to the class. InfoGain(Class,Attribute) = H(Class) - H(Class | Attribute). It was applied to the data set in the first order of six characteristics that are taken into account, which are: AgeCategory, GenHealth, DiffWalking, Diabetic, Stroke, PhysicalHealth: weka.attributeSelection. InfoGainAttributeEval.

- CorrelationAttributeEval (CA)

Evaluates the value of an attribute by measuring the correlation (Pearson's) between it and the class. Nominal attributes are considered on a value by value basis by treating each value as an indicator. An overall correlation for a nominal attribute is arrived at via a weighted average. It was applied to the data set in the first order of six characteristics that are taken into account, which are: DiffWalking, Stroke, Physical Health, Diabetic, Kidney Disease, Smoking: weka.attributeSelection.CorrelationAttributeEval.

- ReliefFAttributeEval (RA)

Evaluates the value of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class. Can operate on both discrete and continuous class data. The next step, based on the results, was to modify the data set and remove the unimportant features. The modified data in table 2 is to participate in the classification task, then show the result of the classification result before and after the features, and then compare. weka.attributeSelection.ReliefFAttributeEval. After the data was entered and configured by the Information Gain Attribute evaluation filter, the fields that have the greatest impact in the process of predicting heart disease were highlighted, and the field that has the largest percentage is stroke, then heart disease, then walking, and then diabetes, as shown in figure 3.

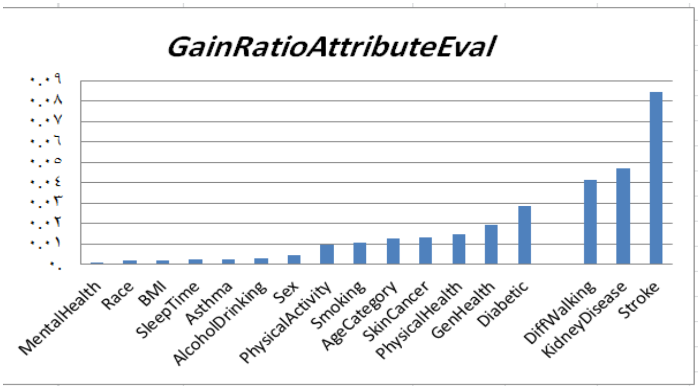


Figure 3: GainRatioAttributeEval

3.6 Feature Importance

After finding the best features from the data set that was, which were identified by the filtering technique as in Table 2 we obtained a specific set of variables and dropped the following variables BMI, MentalHealth, Sex, Race, PhysicalActivity, GenHealth, SleepTime, Asthma, SkinCancer, AlcoholDrinking.

After the five filters were applied to the used data, the recurrent and most important features associated with predicting heart disease were selected from each classification and the input variables that had the strongest relationship with the target variable were selected in order to reduce data, increase accuracy and reduce processing time. The following features were adopted in the process of building a model Prediction of heart disease, DiffWalking, Kidney Disease, Diabetic, GenHealth, PhysicalHealth, AgeCategory, Smoking, HeartDisease As shown in Table 2.

Table 2: The best features of the dataset that was selected by the Filter technique

Filters used	Returned ranking of attributes
CfsSubsetEval (CSE)	Stroke
GainRatioAttributeEval	DiffWalking
Information Gain Attribute evaluation (IG)	KidneyDisease
CorrelationAttributeEval (CA)	Diabetic
ReliefFAttributeEval (RA)	GenHealth
	PhysicalHealth
	AgeCategory
	Smoking
	HeartDisease

Figure 4 presents the incidence rate for each age group. The data indicates that the highest incidence of heart disease was observed in the age range of 65 to 69 years.

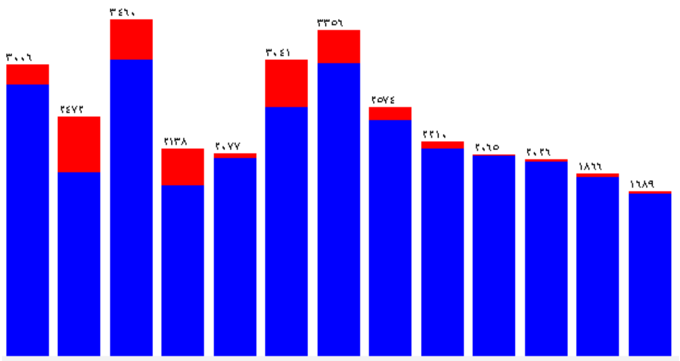


Figure 4: Age Category data for the prediction of heart disease

3.7 Measures of Evaluation:

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN} \quad \text{…....(1)}$$

TN = True Negative, FP = False Positive, FN = False Negative, TP = True. PositiveAccuracy represents the number of correctly classified data instances over the total number of data instances:

$$Precision = \frac{TP}{TP + FP} \quad \text{…....(2)}$$

TP = True Positive, FP = False Positive.

Precision should ideally be 1 (high) for a good classifier. Precision becomes 1 only when the numerator and denominator are equal i.e TP = TP +FP, this also means FP is zero. As FP increases the value of denominator becomes greater than the numerator and precision value decreases (which we don’t want).

$$Recall = \frac{TP}{TP + FN} \quad \text{……(3)}$$

TP = True Positive, FN = False Negative.

Recall should ideally be 1 (high) for a good classifier. Recall becomes 1 only when the numerator and denominator are equal i.e TP= TP +FN, this also means FN is zero. As FN increases the value of denominator becomes greater than the numerator and recall value decreases (which we don’t want).There are many different machine learning algorithms used in data mining models We test three different machine learning algorithms.
The following paragraphs provide a concise overview of these ML algorithms.

Decision Trees (DT): It is regarded as one of the most straightforward and widely used learning algorithms [4]. The decision tree has a structure akin to a flowchart, where each internal node signifies a test on a specific attribute, the branches denote the outcomes of those tests, and each leaf node corresponds to a class label. Utilized in statistics, data mining, and machine learning, it serves as a predictive model to infer conclusions about a set of factors and is particularly adept at managing missing classes, a feature unmatched by other tree-based models.

Logistic Regression (LR): It is a statistical model that predicts the likelihood of an event happening by expressing the event probabilities as a linear combination of one or more independent variables. Logistic regression serves as an estimate for the parameters of the logistic model, with these parameters determined through maximum likelihood estimation [8].

Naive Bayes Algorithm (NB): It is one of the supervised machine learning algorithms, characterized as a straightforward and efficient method that assumes the features are independent, meaning they contribute in a future context without any relationships or correlations among them [7].

After the process of filtering and initializing the data, the most important features associated with the process of predicting heart disease, which have the largest role and direct impact, were obtained. They are 9 out of 18 features, which are HeartDisease, Smoking, Stroke, PhysicalHealth, DiffWalking, Diabetic, GenHealth, KidneyDisease, and are shown in the table 3.

Table 3: modified Dataset after Feature selection.

Attribute	Description	Type
HeartDisease	Respondents that have ever reported having coronary heart disease (CHD) or myocardial infarction (MI)(yes or no)	Nom
Smoking	Have you smoked at least 100 cigarettes in your entire life? [Note: 5 packs = 100 cigarettes] (yes or no)	Nom
Stroke	(Ever told) (You had) a stroke? (yes or no)	Nom
PhysicalHealth	Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30	Num
DiffWalking	Do you have serious difficulty walking or climbing stairs? (yes or no)	Nom
AgeCategory	Fourteen-level age category (65-69 60-64 ,Other)	Nom
Diabetic	(Ever told) (You had) diabetes? (yes or no)	Nom
GenHealth	Would you say that in general your health is... (Very good, Good, Other)	Nom
KidneyDisease	Not including kidney stones, bladder infection or incontinence, were you ever told you had kidney disease? (yes or no)	Nom

4. Proposed Model

The working principle of the proposed model for the prediction of heart disease First, the data is entered, and then the filtering and initialization process for the data takes place, and then the processing process takes place, as the data was divided into two parts, 80% training data and 20% test data, where the data was trained on algorithms (the logistic regression algorithm and (Neva Bayes algorithm and decision tree algorithm) and when applying the test data in the model, the data was classified into a person with heart disease or normal. These steps were illustrated in Figure 5.

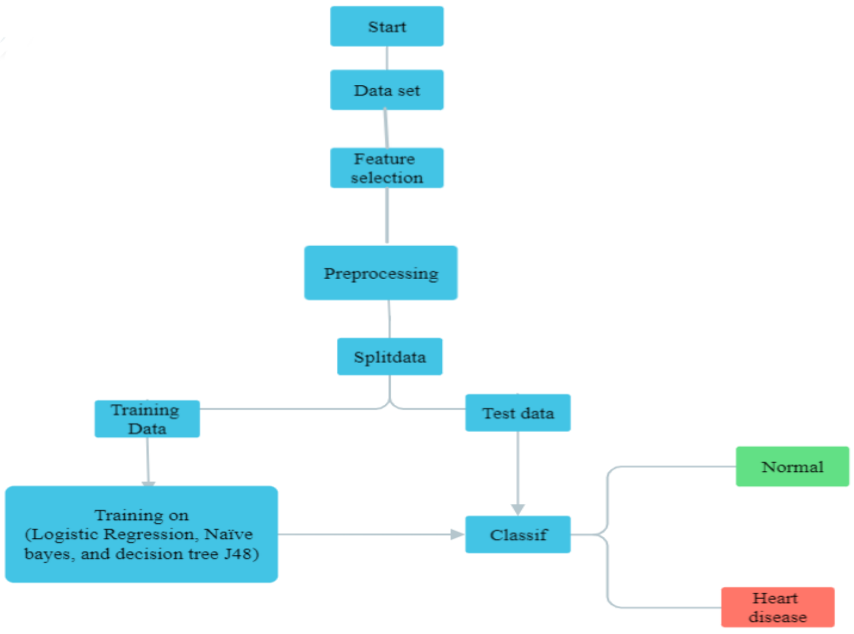


Figure 5: proposed model

5. Results and Discussion

This section outlines the prediction results to evaluate the performance of machine learning algorithms in relation to the methods for selecting features and attributes. The WEKA data mining tool was utilized, with default parameters applied for all classification algorithms. Experiments were conducted by splitting the dataset into 80% for training and the remaining 20% for testing. During training, the model is fed the input and output from the 80% training data, enabling it to learn exclusively from this portion. For prediction, the model processes the unseen 20% test data, generates predicted values, and these predictions are then compared to the actual outputs to assess accuracy. Five classifiers were used selection algorithms to select attributes before passing the data set, namely CfsSubsetEval, GainRatioAttributeEval, Information Gain Attribute evaluation, CorrelationAttributeEval, ReliefFAttributeEval. The outcomes of heart disease prediction experiments utilizing three algorithms—Logistic Regression, Naïve Bayes, and Decision Tree J48—are presented in table 4. This table displays classification results based on 9 carefully selected features, determined through five filtering algorithms that implemented feature reduction. The Decision Tree J48 algorithm initially achieved an accuracy of 91.9%, which improved to 92% after applying the accuracy scale. The Naïve Bayes algorithm demonstrated an initial accuracy of 87.6%, rising to 87.7% post-application of the accuracy scale. Lastly, the Logistic Regression algorithm had an accuracy of 91.6% before the implementation of the accuracy scale, which increased to 92.1% afterward.

Table 4: Accuracy before & after

Accuracy		
Classifier	Before	After
Decision tree J48	0.919	0.920
Naive Bayes	0.876	0.877
Logistic Regression	0.916	0.921

The highest model accuracy is attained with Logistic Regression, recording 91.6% before attribute selection and 92.1% after. Table 5 presents the outcomes of applying Recall and Precision measures. Initially, the Decision Tree J48 algorithm yielded results of (0.890, 0.165) for precision and recall, respectively, but after applying these measures, the results changed to (0.886, 0.112). In comparison, the Naive Bayes algorithm showed a precision and recall of (0.883, 0.307) before selecting the measures, which adjusted to (0.885, 0.287) afterward. For the Logistic Regression algorithm, the results were (0.889, 0.222) before and (0.213, 0.889) after applying the measures.

Table 5: Comparison based on Precision and Recall

Classifier	Before		After	
	Precision	Recall	Precision	Recall
Decision tree J48	0.165	0.890	0.112	0.886
Naive bayes	0.307	0.883	0.287	0.885
Logistic Regression	0.222	0.889	0.213	0.896

Following feature selection, it significantly impacts the prediction precision of the Logistic Regression classifier. Before feature selection, the Recall and Precision were 0.222 and 0.889, respectively. After feature selection, these values changed to Recall of 0.213 and Precision of 0.896. Figure 6 illustrates the accuracy results for each of the three algorithms both before and after applying the accuracy measure to the data.

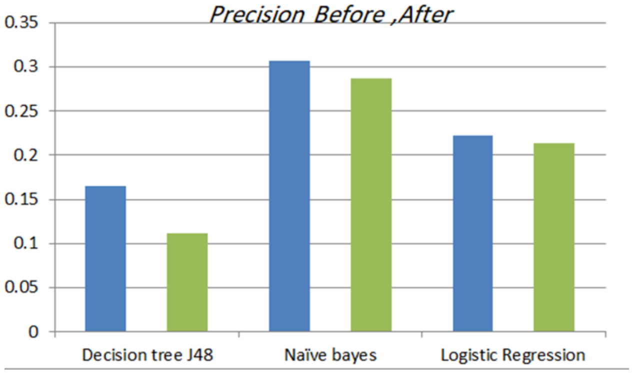


Figure 6: Comparison based on Precision before and after feature selection

Figure 7 shows a graph showing the accuracy results for each of the three algorithms before and after the process of applying Recall Precision measures to the data.

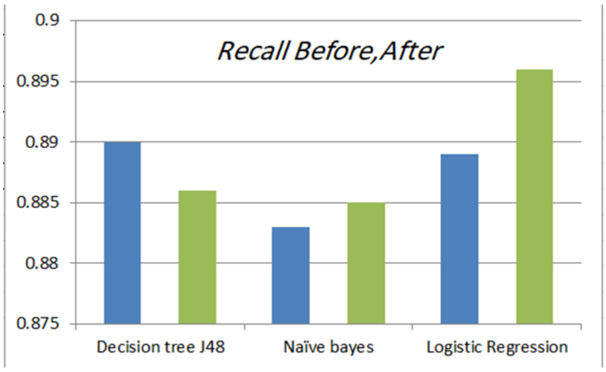


Figure 7: Comparison based on Recall before and after feature selection

6. Conclusions

Feature selection was employed as an initial step to reduce the problem's dimensionality and enhance accuracy. This study analyzed a dataset comprising 319,795 entries and 18 attributes. We utilized three algorithms—Naïve Bayes, Decision Tree J48, and Logistic Regression—to predict heart disease. Five filtering methods were applied for attribute selection: CfsSubsetEval, GainRatioAttributeEval, Information Gain Attribute Evaluation, CorrelationAttributeEval, and ReliefFAttributeEval. Classification was performed using Weka software, and the three algorithms were compared based on accuracy, recall, and precision. Logistic Regression yielded the best results in predicting heart disease. Leveraging machine learning techniques is more effective than human predictions for diagnosing and treating heart disease, enabling quicker and more cost-effective care for patients and doctors. We also compared results from various programs, including Orange and Rapid Miner, and encourage all Palestinian researchers to establish a comprehensive and accurate medical data repository for Palestine. This initiative aims to create a complete database that can predict numerous diseases in advance and facilitate better treatment outcomes compared to what is presented in this paper.

7. References

1. Almustafa, M. K. (2020): Prediction of heart disease and classifiers. Riyadh, Kingdom of Saudi Arabia: BCM Bioinformatics.

2. Chang, V., Bhavani, V. R., Xu, A. Q., & Hossainc, M. (2022): An artificial intelligence model for heart disease detection using machine learning algorithms. Healthcare Analytics.
3. Gultepe, Y., & Rashed, S. (2019): The Use of Data Mining Techniques in Heart Disease Prediction. International Journal of Computer Science and Mobile Computing.
4. Kamiński, B., Jakubczyk, M., Szufel, P. (2017): A framework for sensitivity analysis of decision trees. Central European Journal of Operations Research. 26 (1): 135–159. doi: 10.1007/s10100-017-0479-6. PMC 5767274. PMID 29375266.
5. Malik, Z. L., Fatema, M., Pooja, N., & Ankita, G. (2021): Heart Disease Prediction using Artificial Intelligence. Mumbai: International Journal of Engineering Research & Technology (IJERT).
6. Marwah Abdulrazzaq Naser, Aso Ahmed Majeed, Muntadher Alsabah, Taha Raad Al-Shaikhli, Kawa M. Kaky : A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges, Algorithms (2024), 17(2), 78; https://doi.org/10.3390/ a 17 02 0 078.
7. McCallum, Andrew: Graphical Models, Lecture2: Bayesian Network Representation (PDF). Archived (PDF) from the original on 2022-10-09. Retrieved 22 October (2019).
8. Miguel-Hurtado, O., Guest, R., Stevenage, S., Neil, G., & Black, S. (2016): Comparing Machine Learning Classifiers and Linear/Logistic Regression to Explore the Relationship between Hand Dimensions and Demographic Characteristics. PLOS ONE, 11(11), e0165521. doi: 10.1371 / journal.pone.0165521.
9. Mohan, S., Thirumalai, C., & Srivastava, G. (2019): Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques IEEEAccess.
10. Nadikatla Chandrasekhar, Samineni Peddakrishna: Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization. 2023, 11(4), 1210; https://doi.org/10.3390/pr11041210.
11. Ngare, K. N. (2019): Heart Disease Prediction System.
12. Pandey, S. K., Mishra, B., & Gautam, S. (2020): Knowledge Discovery in Heart Dataset using Classification Techniques. International Journal of Computer Science and Mobile Computing.
13. Phiri, M. E., Valan, A., & Selvam, G. (2022): Prediction of Heart Disease Using Machine Learning Techniques. Malawi: international journal of creative research thoughts (IJCRT).
14. Pytlak, K. (2020, May): Personal Key Indicators of Heart Disease. Retrieved from Kaggle: https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease? Resource = download.
15. Rajdhan , A., Sai, M., Agarwal, A., Ravi , D., & Ghuli, P. (2020): Heart Disease Prediction using Machine Learning. Bengaluru, India : International Journal of Engineering Research & Technology (IJERT).
16. Shafique, U., Majeed, F., Qaiser, H., Ul Mustafa, I. (2015): Data Mining in Healthcare for Heart Diseases. Gujrat, Pakistan: International Journal of Innovation and Applied Studies.
17. Tolles, Juliana; Meurer, William J: Logistic Regression Relating Patient Characteristics to Outcomes (2016). JAMA. 316 (5):533-4. doi:10.1001/jama.2016.7653. ISSN 00987484.OCLC6823603312 PMID27483067.
18. Waikato, T.U.o. WEKA documents, https://www.cs.waikato.ac.nz/ml/weka/.
19. website https://www.towardsdatascience.com/datapreprocessing, October (2021).
20. World health organization. (2021, June 11). Cardiovascular diseases (CVDs). Retrieved from World Health Organization: https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases (cvds).