International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8158
***Article Received:** 21 June 2023 **Revised:** 19 August 2023 **Accepted:** 05 September 2023*

Automated Heart Syndrome Forecast Model Exploiting Machine Learning Approaches

Parisha[1], Gaurav Kumar Srivastava[2], Santosh Kumar[3]
[1]Research Scholar: Dept. of Computer Science and Engineering
Babu Banarsi Das University
Lucknow, India
parisha369@gmail.com
[2]Assistant Professor: Dept. of Computer Science and Engineering
Babu Banarsi Das University
Lucknow, India
[3]Associate Professor: School of Computing Science & Engineering
Galgotias University
Greater Noida, India
Sant7783@hotmail.com

Abstract— Heart disease is a frequent condition that appears as a result of a poor diet and an irregular lifestyle. It is one of the most frequent diseases worldwide, with numerous reasons that damage the heart and have claimed countless lives in recent years. Due to the enormous number of risk factors for heart disease, it is critical to adopt a precise and dependable approach to provide an early diagnosis and correct prognosis. As a result, there is a broad potential for implementing various types of machine learning approaches for retrieving such critical data from the database. This study evaluates numerous machine learning algorithms for correctly predicting cardiac sickness and offers analytical findings, with an emphasis on various methodologies.

Keywords- Heart Infection, Machine Learning Techniques,, K-NN, Artificial Neural Network, Decision Trees.

I. INTRODUCTION

Around the previous decade, heart disease has been the leading cause of death all around the world. According to a study published by the World Health Organization, around 18-to-20 million people die each year as a result of heart disease. People from low- and middle-income countries are the most likely to die. Other habitual risk factors are smoking, alcohol overdoses, hypertension, and lack of physical activity.

Data mining is a significant data extraction approach from massive datasets in today's society. To predict cardiac disease, regression techniques, clustering techniques, association rules, and classification strategies using the Nave Bayes algorithm, decision trees, random forests, and the K-nearest neighbours algorithm are just a few of the ways used. A medical dataset was acquired from different hospitals for this inquiry. As a result, this study is putting trendy techniques and algorithms for predicting and detecting various cardiac disorders to the test.

II. MOTIVATION/BACKGROUND

Around the previous decade, heart disease has been the leading cause of death all around the world. According to a study published by the World Health Organization, around 18-to-20 million people die each year as a result of heart disease. People from low- and middle-income countries are the most likely to die. Other habitual risk factors are smoking, alcohol overdoses, hypertension, and lack of physical activity.

Data mining is a significant data extraction approach from massive datasets in today's society. To predict cardiac disease, regression techniques, clustering techniques, association rules, and classification strategies using the Nave Bayes algorithm, decision trees, random forests, and the K-nearest neighbours algorithm are just a few of the ways used. A medical dataset was acquired from different hospitals for this inquiry. As a result, this study is putting trendy techniques and algorithms for predicting and detecting various cardiac disorders to the test.

Several academics and scientists have previously done extensive work on cardiac problems in current settings. Let's start with some of their prior work on heart disease: Bhatla and Jyoti [1] sought to examine the various data mining approaches introduced in recent years for heart disease prediction. Dangare and Apte [2] investigated prediction methods for heart disease using a larger number of input attributes. Karthikeyan and Kanimozhi [3] suggested a Heart Disease Prediction System that uses a Deep Belief Network classification algorithm to estimate the user's odds of developing heart-related ailments. Masethe and Masethe [4]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8158
***Article Received:** 21 June 2023 **Revised:** 19 August 2023 **Accepted:** 05 September 2023*

predicted cardiac attacks with 99% accuracy. Data mining enables the health sector to predict patterns in the dataset. Rajkumar and SophiaReena [5] have supervised the data categorization, which is based on machine learning methods, resulting in accuracy and time spent to create the system. Kodati and Vivekanandam [6] offered a brief overview of an open source datamining tool for cardiac disease. As a source data, Venkatalakshmi and Shivsankar proposed using a 13 attribute structured clinical database from the UCI Machine Learning Repository. Decision trees and Naive Bayes were used, and their diagnostic performance was compared. Taneja [7] has created a cost-effective method for facilitating data base decision support systems by utilizing data mining technology. Patel et al [8] used WEKA to examine alternative Decision Tree classification algorithms in order to improve performance in heart disease diagnosis. Gomath and Priyaa [9] use data mining tools to analyze heart problems in male patients. Subbalakshmi et al. [10] created a Decision Support in Heart Disease Prediction System (DSHDPS) utilizing the Nave Bayes data mining modeling technique. Jabbar et al. [11] proposed an efficient associative classification system for heart disease prediction utilizing a genetic approach. Singh and Jindal [12] created a model that predicts heart disease with great accuracy. Kaur and Singh [13] reviewed several articles in which one or more data mining techniques were employed to predict cardiac disease. Hazra et al. [14] sought to synthesize some of the most recent research on predicting cardiac disease using data mining tools, as well as to examine the various methods combinations of mining algorithms used and conclude which technique(s) are effective and efficient.

A. Dataset

This research aims to analyze the likelihood of emergent cardiac sickness as revealed by the computer algorithm in order to assist physicians and inpatients in the clinical setting. In this work, we used a variety of machine learning methods to acquire datasets in order to accomplish the desired result, and the current work also revealed how specific features are important for improving prediction accuracy. We received a patient dataset from several hospitals that contained over 300 data attributes and 13 distinct variables, such as blood_pressure, the type of chest discomfort, and ECG results, all of which are shown in Table 2.1.1. The causes of heart disease were obtained using four different types of machine learning algorithms to obtain the best accurate model possible: decision tree, Nave Bayes, K-Nearest Neighbor, and Random forest.

Table 2.1.1. Characteristics of Heart Illness

S. No.	Characteristic	Illustrative icon	Facts
1	Oldness	Oldness	How old a Patient is, in years
2	Gender	Gender	0=Feminine, 1=Masculine
3	Coffer discomfort	Cp	a. Typical angina b. Atypical angina c. Non-angina d. Asymptomatic
4	Relaxation Plasma Heaviness	Tresrbps	Latent Systolic plasma gravity at time of admittance in infirmary
5	Blood serum Lipid	Chol	Fluid fat in mg/dl
6	Abstaining Plasma Darling	Fbs	Abstaining plasma honey>120 mg/dl (0-untrue, 1-factual)
7	Breather Cardiograph	Resteeg	0-Typical 1-intense ST-T surge indiscretion; 2- hypertrophy of the left ventricle
8	Max Heart Rate	Thalch	Maximum Heart Rate attained
9	Angina is exacerbated by exercise.	Exang	Angina induced by exercise (0-no; 1-yes)
10	ST Despair	Oldpeak	ST sadness caused by exercise in comparison to rest
11	Grade	Slope	Peak workout ST segment slope 1-upsloping 2-flat 3- downward sloping
12	No. of containers	Ca	The number of major vessels (0-3) fluoroscopically coloredy
13	Mediterranean anemia	Thal	Types of Defects 3- Normal 6-repaired flaws 7 reversible flaws
14	Num (Class Attribute)	Class	Heart disease status diagnosis 0 = no danger 1-low risk 2-possible risk 3-high danger 4-extremely high risk

B. Pre-Processing of Dataset

In fig.2.2.1, an activity diagram is designed in which new entries are input and these entries are sent forward to check whether the value is missing or not, if the value is missing, it is sent back to the input point where the missing value is corrected and sent to the decision box. If the value is acceptable, it is delivered to the dataset analyzer, where it is trained and tested. As the dataset passes through the dataset

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8158
***Article Received:** 21 June 2023 **Revised:** 19 August 2023 **Accepted:** 05 September 2023*

analyzer, machine learning techniques are applied to it to determine the accuracy of the dataset, and the process is terminated after the accuracy is determined.

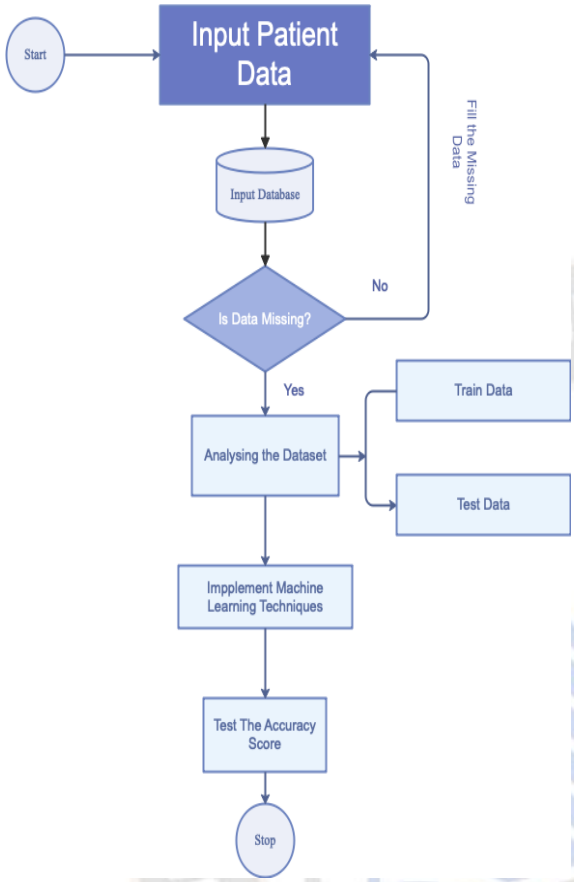

Figure 2.2.1. Activity Chart for Pre-Processing Dataset

III. RESEARCH METHODOLOGY USED

A. Machine Learning Algorithms Used

a. Decision Tree:

It employs a categorization strategy in conjunction with a numerical dataset. This type of structure (Tree-Like) is used here to make decisions more easily and as a widely used technique for directing clinical datasets. The decision tree has three nodes, each of which serves as the foundation for the decision tree model's analysis.

- Root node: controls the functionality of all other nodes that are connected with that.
- • Interior node: manages many characteristics
- • Leaf node: displays the results of each test.

This program divides the input dataset into two equivalent sets. Following an analysis of the entropy of each attribute, the data is divided into predictors with the highest information gain or the lowest entropy. As a result, entropy and gain are denoted as follows:

$$Entropy(\mathrm{S}) = \sum_{i=1}^{c} -Pi \log_2 Pi,$$

$$\text{Gain } (S, A) = \text{Entropy}(S) - \sum_{v \in \text{ Values}(A)} \frac{|Sv|}{|S|} \text{Entropy } (Sv)$$

After analyzing the dataset, this algorithm represents it in a tree-like style and returns the results in a more straightforward and accurate way. When only one attribute is tested in this algorithm, an over classified situation may emerge.

b. K-Nearest Neighbour (K-NN)

It is a supervised categorization approach in which objects are classified according on their distance from their neighbors. In this method, the Euclidean distance is utilized to calculate how far a property is from its neighbors. This approach classifies the dataset based on object similarity. It may fill in the missing values before applying all of the prediction algorithms in different ways to improve accuracy. This method is versatile and can be applied to search, regression, and classification. Although K-NN is the most basic strategy, its accuracy is limited by noisy and irrelevant data.

c. Random Forest Algorithm

In this approach, a forest is composed of multiple trees, each of which emits a class expectation, and the class with the most votes becomes the model's forecast. The random forest classifier becomes more accurate as the number of trees increases. The three most common ways are as follows:

- Forest RI (random input selection);
- Forest RC (random blend);
- Forest RI/RC combination

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8158
***Article Received:** 21 June 2023 **Revised:** 19 August 2023 **Accepted:** 05 September 2023*

Novelists andYear	Strategies	Attributes	Attributes Utilized	Fidelity
Asha Rajkumar (2010) et al.	Naïve Bayes	It is simple to evaluate, takes only a little amount of training data, assumes feature conditional independence, and is widely used when predicting a high number of classes.	14	53.33%
	Decision Trees (J48)	Take care of missing values and outliers; over-fitting is the most significant.		53%
	KNN	Simple to use, performs well on simple diagnosis issues, and is non-parametric (there are no predefined assumptions).		46.67%
M. A. Jabbar etal. (2011)	CBARBSN	Using both supervised and unsupervised techniques, quick processing time is used to recognize typical item patterns and extract characteristics.	14	NM*
Chaitrali S. Dangare et al. (2012)			13	
	Decision Trees (J48)	The highest noticeable characteristic is over-fitting.		97.66%
	Neural Networks	Generalization of input, non-linear data processing, high fault tolerance, and self-repair when a network node(s) fails.		99.25%
AbhishekTaneja (2013)	J48 UnPruned	Attend to missing values and outliers.	8	96.52%
	J48 Pruned	The most notable trait is over-fitting.		96.96%
	ANN	Non-linear data processing, the ability to generalize input, and a high failure tolerance.		95.85%
B.Venkatalak	Decision Tree	Over-fitting is simple and easy.	13	85.013 %

IV. MAJOR FINDINGS AND DISSCUSSION

The purpose of this study is to forecast a patient's risk of acquiring heart disease. In this application, classification techniques such as Naive Bayes, decision trees, random forests, and K-nearest neighbors are used. Weka was employed in several research that used various categorization approaches. The investigation was carried out using an Intel Corei7 8th generation PC with a 4.1 GHz 8750H CPU and 16 GB of RAM. The categorized data set was used to generate a training and testing set. The data is pre-processed using supervised classification algorithms like Nave Bayes and decision trees.

Table 4.1.1: Truth Table about Heart Illness Prediction Using Various Method

Scribes	Approach	Fidelity
Otoom eat al [12]	Naïve Bayes	86.5%
	Functional Tree	83.5%
	Support Vector Machine	86.5%
Chaurasia et al [14]	J48	85.65%
	Bagging	86.03%
	Support Vector Machine	96.01%
Seema et al [16]	Naïve Bayes	97.01%
	Decision Tree	98.05%
	Support Vector Machine	83.05%
Kumar Dwivedi et al [17]	Naïve Bayes	84.05%

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8158
***Article Received:** 21 June 2023 **Revised:** 19 August 2023 **Accepted:** 05 September 2023*

	K-NN	82.05%
	Classification Tree	78.50%
	Support Vector Machine	83.00%
	ANN	84.90%
	Logistic Regression	86.01%
The proposed model	Naïve Bayes	90.01%
	K-NN	92.50%
	Decision tree	81.26%
	Random Forest	88.50%

K-nearest Neighbors and Random Forest are used to get the accuracy score. The Python programming language was used to record the Fidelity number outputs of several classification techniques for training and testing the dataset. The % accuracy scores for various approaches are shown in Table 4.2. Table 4.1 compares the proposed model's accuracy rating for heart disease prediction with that of other authors.

Table 4.1.2: Accuracy of Grouping Approaches

K-Nearest	Decision Tree	Random Forest
83.05	78.01	88.21
95.05	84.26	86.04

V. CONCLUSIONS

We noticed that numerous data mining and machine learning approaches are utilized to forecast heart disease based on an evaluation of many current research articles on the prediction of heart disease using various data mining and machine learning methodologies and algorithms. numerous trials make use of numerous datasets of people suffering from cardiac disease. The bulk of experiments make use of data from multiple hospital databases. The dataset includes over 200 entries with 14 essential attributes, some of which have missing values.

According to the study, neural networks with 15 characteristics performed with 99.99% accuracy in one trial, whereas neural networks with 8 attributes functioned with 75.44% accuracy. In the majority of experiments with varied numbers of attributes, Naive Bayes provides high accuracy (89%). For example, the accuracy of decision lists (J48) increases to 99.53%, which is a very good performance. As a result, the accuracy of the various strategies is determined by the number of attributes employed and the implementation tool used. We extract the following findings from this work, which should be explored in future research for high accuracy and more accurate heart disease diagnosis using intelligent prediction systems.

ACKNOWLEDGMENT

Authors are grateful to Vice-Chancellor, Babu Banarsi das university Lucknow for providing the excellent facility in the computing lab of B. B. D. University, Lucknow, India. Thanks are also due to University Grant Commission, India for financial support to the University.

REFERENCES

[1] Nidhi Bhatla, and Kiran Jyoti, “An Analysis of Heart Disease Prediction using Different Data Mining Techniques”, International Journal of Engineering Research & Technology (IJERT), Vol. 1, Oct.2012.

[2] Chaitrali S.Dangare, and Sulabha S.Apte, “Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques”, International Journal of Computer Applications (0975 – 888), Vol. 47, No.10, June.2102.

[3] Dr. T. Karthikeyan, and V.A.Kanimozhi, “Deep Learning Approach for Prediction of Heart Disease Using Data mining Classification Algorithm Deep Belief Network” , International Journal of Advanced Research in Science, Engineering and Technology, Vol. 4, Issue 1, January 2017.

[4] Rajendran, P. S. ., & Kartheeswari , K. R. . (2023). Feature-Based Machine Intelligent Mapping of Cancer Beating Molecules. International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 266–277. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2652.

[5] Hlaudi Daniel Masethe, and Mosima Anna Masethe, “Prediction of Heart Disease Using Classification Algorithms”, in Proceedings of the World Congress on Engineering and Computer Science 2014 Vol. II WCECS 2014, 22-24 Oct. 2014, San Francisco, USA.

[6] Asha Rajkumar, and Mrs. G. SophiaReena, “Diagnosis of Heart Disease Using Data Mining Algorithm”, Global Journal of Computer Science and Technology, Vol. 10, pp. 38-43, Sept. 2010.

[7] Sarangam Kodati, and Dr. R Vivekanandam, “A Comparative Study on Open Source Data Mining Tool for Heart Disease” , International Journal of Innovations & Advancement in Computer Science, Vol. 7, Issue 3, March-2018.

[8] Prof. Virendra Umale. (2020). Design and Analysis of Low Power Dual Edge Triggered Mechanism Flip-Flop Employing Power Gating Methodology. International Journal of New Practices in Management and Engineering, 6(01), 26 - 31. https://doi.org/10.17762/ijnpme.v6i01.53.

[9] B.Venkatalakshmi, and M.V Shivsankar, “Heart Disease Diagnosis Using Predictive Data mining”, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Special Issue 3, March-2014.

[10] Abhishek Taneja, “ Heart Disease Prediction System Using Data Mining Techniques”, Oriental Journal Of Computer Science and Technology, Vol. 6, pp. 457-466, Dec.2013.

[11] Jaymin Patel, Prof.Tejal Upadhyay, and Dr. Samir Patel, “ Heart Disease Prediction Using Machine Learning and Data

Mining Technique", IJCSC, Vol. 7, No. 1, pp.129-137, September-2015.

[12] K.Gomath, Dr. Shanmugapriyaa, "Heart Disease Prediction Using Data Mining Classification", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol.4, Issue 2, February-2016.

[13] Joseph Miller, Peter Thomas, Maria Hernandez, Juan González, Carlos Rodríguez. Exploring Ensemble Learning in Decision Science Applications. Kuwait Journal of Machine Learning, 2(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/206.

[14] G.Subbalakshmi, K. Ramesh, and M.C. Rao, "Decision Support in Heart Disease Prediction System using Naive Bayes", Indian Journal of Computer Science and Engineering (IJCSE), Vol. 2, No. 2, Apr-May 2011.

[15] MA.Jabbar, Dr. Priti Chandra, and B.L. Deekshatulu, "Cluster Based Association Rule Mining For Heart Attack Prediction", Journal of Theoretical and Applied Information Technology, Vol. 32 No.2, October-2011.

[16] Navdeep Singh and Sonika Jindal, " Heart Disease Prediction System using Hybrid Technique of Data Mining Algorithms", International Journal of Advance Research, Ideas and Innovations in Technology, Vol.4, Issue 2, 2018.

[17] Ahammad, D. S. H. ., & Yathiraju, D. . (2021). Maternity Risk Prediction Using IOT Module with Wearable Sensor and Deep Learning Based Feature Extraction and Classification Technique. Research Journal of Computer Systems and Engineering, 2(1), 40:45. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/19.

[18] Beant Kaur, and Williamjeet Singh, "Review on Heart Disease Prediction System using Data Mining Techniques", International Journal on Recent and Innovation Trends in Computing and Communication, Vol.2 Issue 10, October-2014.

[19] Animesh Hazra, S.Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Advances in Computational Sciences and Technology, Vol. 10, No.7, July-2017.