International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *13 November 2023*

Prediction and Classification of Heart Disease Using Different Classifier Algorithms

Madhurima Mondal [1], Dr. Narendra Chaudhari [2], Dr. Amjan Shaik [3]

[1]Research Scholar, Department of Computer Science & Engineering, Mansarovar Global University, Sehore, M.P., India.

[2]Research Guide, Department of Computer Science & Engineering, Mansarovar Global University, Sehore, M.P., India

[3] Professor of Comp. Sc. and Engg. & Dean-R&D Cell, St.Peter's Engineering College (SPEC), Maisammaguda, Medchal, Hyderabad , Telangana, India

Abstract :

One of the leading causes of death all over the world is cardiovascular disease, making early detection of the condition crucial. Predicting and diagnosing cardiac illness is made easier with the use of computer-aided technologies. Data mining is a method used on huge databases to discover previously unseen connections or correlations using a hybrid approach to statistical analysis, machine learning, and database technology. Furthermore, the growth of diverse applications in the thriving healthcare industry makes medical data mining an incredibly significant study subject. This study's overarching objective is to create a cardiovascular disease prediction system that makes use of key risk indicators and employs many classifier methods, including Naive Bayes, Support Vector Machine, and K-Nearest Neighbors. This study employs the heart illness dataset housed in the UCI Machine Repository in an effort to better diagnose cardiovascular conditions. Based on the findings, it appears that the suggested unique optimized algorithm can serve as the basis for a reliable healthcare monitoring system capable of detecting and preventing cardiac disease at an early stage.

Keywords: Heart disease, Healthcare, Machine learning, Algorithm, Simulation

I. INTRODUCTION

The heart is such a vital organ that its proper functioning is crucial to our survival. The human brain, kidneys, and other organs are all interconnected, therefore improper heart function can have dire consequences. Heart disease is a medical condition that hinders the pumping ability of the heart. Several factors raise the danger of developing heart disease. These days, heart disease is the leading cause of mortality worldwide. According to the WHO, cardiovascular disease is the leading cause of mortality globally, accounting for an estimated 12 million deaths annually. Data mining techniques allow for precise disease forecasting. The IHDPS (intelligent heart disease prediction system) is able to mine a heart disease database for previously undiscovered insights. As a result, it can help healthcare analysts and practitioners make smart clinical judgments when diagnosing cardiac disease, something that is impossible for standard decision support systems to do. This study provides a table-based review of many data mining approaches that have proven useful to medical analysts and practitioners in making precise diagnoses of heart disease.

In many parts of the world, heart disease ranks high among the leading causes of mortality. It decreases a person's earning potential and raises healthcare bills. In the previous decade, cardiovascular disorders have cost India a staggering amount of money. As a result, the ability to reliably anticipate cardiovascular illnesses is crucial.

Data mining is a future-proof technique since it is hidden in the data and has never been discovered before. In a nutshell, it's the method used to gain insight from a variety of data sources. Data mining combines methods from statistical analysis, machine learning, and database management to discover hidden patterns and connections in massive data sets. Knowledge gained via study has practical implications in the field of medicine. Rapid advances in computing and information technology have sparked a surge of interest in medical and health records, propelling the discipline of health informatics into the spotlight. There are several common misunderstandings and unanswered questions in the field of medical data mining.

The following table provides a summary comparison of historical and projected data mining trends.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *13 November 2023*

Table 1: Various Data Mining Trends from Past to The Future

Data mining Trends	Algorithms/ Techniques Employed	Data formats	Computing Resources
Past	Statistical, Machine Learning Techniques	Numerical data and structured data stored in traditional databases	Evolution of 4G PL and various related techniques
Present	Statistical, Machine Learning, Artificial Intelligence, Pattern Reorganization Techniques	Heterogeneous data formats include structured, semi structured and unstructured data	High speed networks, High end storage devices and Parallel, Distributed computing etc…
Future	Soft Computing techniques like Fuzzy logic, Neural Networks and Genetic Programming	Complex data objects include high dimensional, high speed data streams, sequence, noise in the time series, graph, Multi instance objects etc.	Multi-agent technologies and Cloud Computing

II. REVIEWS OF RELATED STUDIES

A. Saboor et al. (2022), There is now a great deal of variation in the medical sciences as a result of technological and computational advancements, especially in the diagnosis of human cardiac disorders. One of the leading causes of death in humans today, it drastically cuts life expectancy. If human heart disease is correctly and swiftly detected, it is possible to avoid heart failure in its early stages, which will boost the patient's chance of survival. The diagnostic accuracy of heart disease using manual approaches is impaired by bias and inter-examiner variability. When it comes to distinguishing between those with and without cardiac disease, machine learning algorithms are reliable and accurate tools. According to the proposed research, we used several different machine learning algorithms to detect and forecast human heart disease, and we analyzed their results on the heart disease dataset using several different metrics, such as classification accuracy, F-measure, sensitivity, and specificity. Nine machine learning models (AB, LR, ET, MNB, CART, SVM, LDA, RF, and XGB) were applied to the final dataset both before and after hyperparameter tweaking to reach this goal. We additionally validate their performance on the concordant heart disease dataset by applying custom preprocessing, dataset normalization, and hyperparameter tweaking procedures. The machine learning algorithms were trained and validated using the standard K-fold cross-validation technique. Finally, the experimental results demonstrated that improving the prediction classifiers' performance was as simple as standardizing the data and modifying the machine learning classifiers' hyperparameters.

G.M. Muneeswari et al., (2022) Diagnosing and forecasting cardiac disease has always been a tough and time-consuming job for professionals. High-quality heart-related rebuilding treatments and activities are available at many medical clinics and specialty centers. Therefore, it is likely beneficial to individuals everywhere to keep a look out for early signs of heart disease, since this will allow them to begin treatment before any symptoms appear. Heart disease has skyrocketed in recent decades, and the leading contributors are poor lifestyle choices like excessive drinking and smoking and a lack of meaningful employment. The healthcare business has found that machine learning is useful for making choices and projections from the massive amounts of data collected over time. This heart disease prognosis was developed using a number of supervised machines learning methods, including artificial neural networks (ANN), decision trees (DT), random forests (RF), and Naive Bayes (NB). In addition, the results from various approaches are summarized. The results of this investigation allow for the early diagnosis of heart disease. The best algorithm will be selected once the four candidates' accuracy scores are compared.

Shetgaonkar, Pratiksha & Aswale, Shailendra (2021) The human heart is a very essential organ. The heart's proper operation is absolutely crucial to our survival. Today, it remains one of the leading killers worldwide. Even now, heart disease is among the leading causes of death in the Western world. It's commonly cited as the leading cause of death in the world. Detecting cardiac disease early is often challenging for doctors. There are currently many valuable hidden facts and information in the health sector that might prove to be extremely beneficial in making predicted

judgments in the medical profession. Data mining is a methodology for analyzing large datasets with the help of advanced AI algorithms to extract meaningful and actionable insights. This article explores the potential of three of these AI-based methods—the Decision Tree, the Naive Bayes, and the Neural Network—for predicting cardiovascular illness. All of these approaches will be ranked according to their performance on a variety of one-of-a-kind & parameters after having been optimized for higher precision. After that, we'll evaluate the methods' results based on how well they perform across a range of criteria. The most reliable method may then be used to forecast who among men and women would develop coronary heart disease. Medical professionals may utilize this method to make accurate prognoses, allowing patients to receive treatment as soon as possible.

R. Fadnavis et al., (2021) The healthcare industry makes available a vast trove of data. However, not enough analytic tools exist to mine data for unexpected but potentially beneficial patterns. There have been many successful applications of data mining, including the identification of previously unknown relationships and trends, the detection of abnormalities in data, and the facilitation of decision making. This research shows how data mining categorization methods may be used to make predictions about cardiovascular disease. Methods such as Nave Bayes and Decision tree are used to foresee and warn of cardiac illness in individuals, and their efficacy is compared. The Cleveland dataset, which includes 14 characteristics, was used.

Kirmani, Mudasir. (2017) Diseases affecting the heart, lymphatic system, and circulatory system are together referred to as cardiovascular disease. The World Health Organization (WHO) has shown that cardiovascular illnesses are among the leading causes of death and disability worldwide. Tobacco use, an unhealthy diet leading to obesity, lack of exercise, becoming older, substance abuse, and excessive alcohol use are only few of the major dietary and lifestyle factors that contribute to the development of cardiovascular disease. Risk factors for cardiovascular disease include but are not limited to: hypertension, diabetes, hyperlipidemia, stress, and other illnesses. Different methods have been used to make predictions about the frequency of cardiovascular disorders in general and heart disease in specific. Data mining is a computer-based prediction technology that may be used in the diagnosis and treatment of cardiovascular illnesses. It is possible to improve cardiac disease prediction and diagnosis using data mining-based methods. Methods such as decision Trees, Naive Bayes, Neural Networks, Support Vector Machines, Fuzzy Rules, Genetic Algorithms, and Ant Colony Optimization were among those investigated. Based on this data, it's clear that there isn't a single machine learning algorithm that can accurately diagnose and predict the course of CVD. The research also considers the types of behavior, selection criteria, and total number of elements that should be used for optimal prediction.

A.R. Mokashi et al., (2015) There is a massive quantity of information created by hospitals. There is a dearth of knowledge despite the abundance of data in the healthcare system. The lack of an efficient analytical approach to unearth previously unknown connections and pieces of information is to blame. For these reasons, data mining is used to collect and analyze relevant data. Decision trees, Naive Bayes, K-Nearest Neighbors, K-Means, and the BP algorithm are just few of the many data mining techniques used. However, the neural network is one of the most important data-gathering tools. Artificial neural networks (ANNs) are used here because of their intelligence. Disease forecasting is sped up and improved by using data mining techniques. The vast majority of neural network research findings have been made public.

Methaila, Aditya et al., (2014) Data mining's widespread adoption may be attributed to its prominence in businesses including e-commerce, advertising, and retail. Healthcare is one of the emerging industries. In spite of the "information rich" nature of the healthcare business, not all of the data are mined to reveal previously unseen patterns and facilitate sound decision making.The potential of uncovering previously unseen connections and patterns is often wasted. The problem can be improved with the use of cutting-edge data mining modeling tools. The purpose of this study is to use the Decision Tree, Naive Bayes, and Neural Network Classification Modeling Techniques from the field of data mining, as well as the weighted association Apriori algorithm and the MAFIA algorithm, to the problem of predicting the occurrence of heart disease. Predicting a patient's risk of developing heart disease based on factors including age, sex, blood pressure, and glucose levels.

III. PROPOSED METHODOLOGY

Some of the characteristics in a dataset may not be helpful, leading to undesirable outcomes. The primary motivation of this study is to enhance categorization and feature selection using a combination approach, ultimately leading to more accurate diagnoses of heart disease. Using a meta-heuristic approach, this research optimizes the identification of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *13 November 2023*

relevant characteristics in cardiac illness by employing an imperialist competitive algorithm. When compared to genetic and other optimization algorithms, this one can give a better solution for feature selection. The dataset was split into a training set and a testing set after initial processing. With 80% in the training set and 20% in the test set. Once features have been extracted, they may be fed into a classification model, such as a K-nearest neighbor (KNN) algorithm, a Naive Bayes Classifier, or a Support Vector Machine. Therefore, the outcome of cardiac illness diagnostics and its many features can be improved by combining these four approaches. In other words, we're looking for ways to enhance the precision with which cardiac disease is diagnosed through categorization. There has never been an attempt to combine the K-nearest neighbor (KNN) method, the Naive Bayes Classifier, and the Support Vector Machine classifier into one. In contrast to existing methods, the simulation results show that the suggested technique improves performance by reducing the number of features and raising the classification accuracy.

IV. COMPARATIVE RESULTS OF SIMULATION

Evaluation parameters

Accuracy, precision, and sensitivity are determined using the formulae shown below.

Accuracy: Accuracy measures how well data is categorized:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision (P): Precision refers to the proportion of positive predictions that were accurately labeled.

$$P = \frac{TP}{Total\ positive\ classified}$$

Sensitivity: Sensitivity refers to the proportion of correctly identified positive classes.

$$S = \frac{TP}{Total\ positive}$$

Simulation results for KNN

Confusion matrices for K-Nearest Neighbor methods are displayed in Fig. 1. KNN model effectiveness with target class and output class values are described in the matrix.

Here, TP = 3, TN = 9, FP = 1, FN = 2

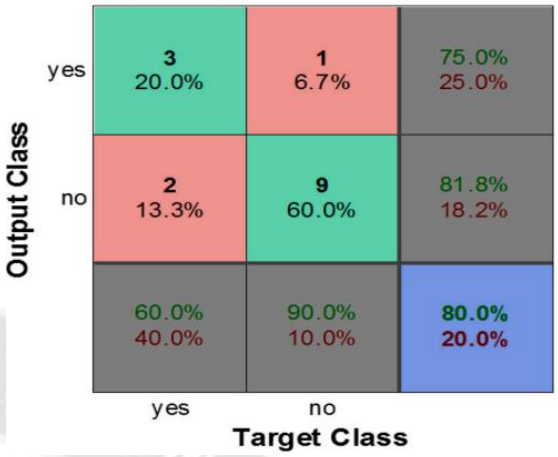

Figure 1: Confusion matrix for KNN based approach

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{3 + 9}{3 + 9 + 1 + 2} = 80\%$$

$$Precision = \frac{TP}{TP + FP} = \frac{3}{3 + 1} = 75\%$$

$$Sensitivity = \frac{TP}{TP + FP} = \frac{3}{3 + 2} = 60\%$$

Simulation results for Naïve Bayes

The confusion matrix for the Naive Bayes method is displayed in Fig. 2. The matrix containing the goal class and output class values reflects how well the NB model performed.

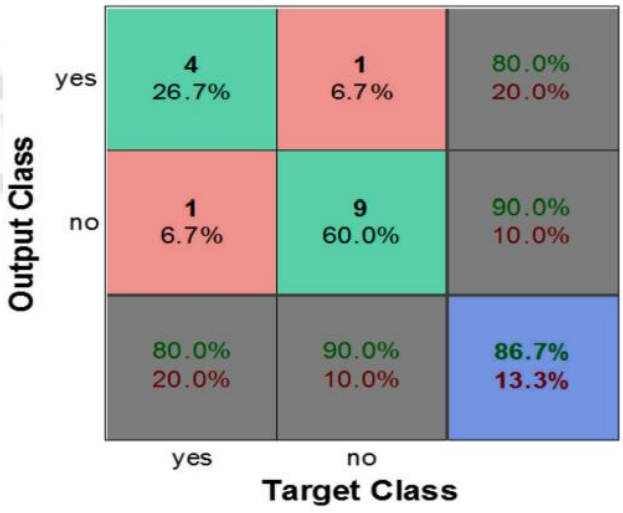

Figure 2: Confusion matrix for Naïve Bayes based

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *13 November 2023*

approach

Here, TP = 4, TN = 9, FP = 1, FN = 4

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{4 + 9}{4 + 9 + 1 + 1} = 86.7\%$$

$$Precision = \frac{TP}{TP + FP} = \frac{4}{4 + 1} = 80\%$$

$$Sensitivity = \frac{TP}{TP + FP} = \frac{4}{4 + 1} = 80\%$$

Simulation results for SVM

For SVM, the confusion matrix is depicted in Fig. 3.

Figure 3: Confusion matrix for SVM based approach

Here, TP = 4, TN = 8 FP = 2, FN = 1

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{4 + 8}{4 + 8 + 2 + 1} = 80\%$$

$$Precision = \frac{TP}{TP + FP} = \frac{4}{4 + 2} = 66.7\%$$

$$Sensitivity = \frac{TP}{TP + FP} = \frac{4}{4 + 1} = 80\%$$

Comparison of various classification methods

MATLAB R2018a was used to create the algorithm for classifying data. The research examined four different categorization strategies to see which one produced the most accurate and error-free predictions pertaining to cardiovascular disorders. The Accuracy, accuracy, and sensitivity of each approach are displayed in the respective confusion matrices. Table 2 displays the compared outcomes of each procedure for easy comparison:

Table 2: Comparative results for various classification methods

Proposed Method	**Accuracy**	**Precision**	**Sensitivity**
KNN	**80%**	**75%**	**60%**
Naïve Bayes	86.7%	80%	80%
SVM	80%	66.7%	80%

V. CONCLUSION

Heart disease is always lethal because of its very nature. This illness adds a layer of complication that can have fatal consequences, such as a heart attack or even death. Data mining classification ensures a smaller issue, shorter learning time, and a more straightforward taught model. In most cases, this reduction makes this model more comprehensible. The prediction accuracy is enhanced, and the classifier may be comprehended and trained without resorting to over-learning. The classification and weight values for the Neural Network in this study are determined using KNN and Nave Bayes algorithms independently. Researchers are employing data mining techniques in the diagnosis of heart disease since the annual death rate of heart disease patients is growing over the world and there is an abundance of data to mine. There has been some success in using data mining to aid in the diagnosis of cardiac disease, but less effort has been put into using the same methods to help patients find an effective therapy.

REFERENCES:

1. Abdul Saboor, Muhammad Usman, Sikandar Ali, Ali Samad, Muhmmad Faisal Abrar, Najeeb Ullah. A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms, Mobile Information Systems, Article ID 1410169, pp. 1-9, 2022.
2. Ganaga Muneeswari M., Soniya V., Aishwaryalakshmi R. K., Abisha D., “Transfer learning-based machine learning models for heart disease prediction in an earlier

stage", International Journal of Advance Research, Ideas and Innovations in Technology, Vol. 8, Issue 1, pp. 570-575, 2022.

3. Shetgaonkar, Pratiksha & Aswale, Shailendra. (2021). Heart Disease Prediction using Data Mining Techniques.
4. Fadnavis, R & Dhore, K & Gupta, D & Waghmare, J & Kosankar, D. (2021). Heart disease prediction using data mining. Journal of Physics: Conference Series. 1913. 012099. 10.1088/1742-6596/1913/1/012099.
5. Kirmani, Mudasir. (2017). Cardiovascular Disease Prediction using Data Mining Techniques. Oriental journal of computer science and technology. 10. 520-528. 10.13005/ojcst/10.02.38.
6. Ankita R. Mokashi, Madhuri N. Tambe, Pooja T. Walke, Review on Heart Disease Prediction using Ann and Classifier, International Journal of Engineering Research & Technology, Vol. 4, Issue 9, pp. 966-968, 2015.
7. G. Purusothaman and P. Krishnakumari, "A Survey of Data Mining Techniques on Risk Prediction: Heart Disease", Indian Journal of Science and Technology, vol. 8, no. 12, (2015).
8. S. J. Gnanasoundhari, G. Visalatchi and M. Balamurugan, "A Survey on Heart Disease PredictionSystem Using Data Mining Techniques", International Journal of Computer Science and Mobile Application, vol. 2, no. 2 (2014), pp. 72-77
9. Methaila, Aditya & Kansal, Prince & Arya, Himanshu & Kumar, Pankaj. (2014). Early Heart Disease Prediction Using Data Mining Techniques. Computer Science & Information Technology. 4. 53-59. 10.5121/csit.2014.4807.
10. D. Ratnam, P. HimaBindu, V. MallikSai, S. P. Rama Devi and P. RaghavendraRao, "Computer-Based Clinical Decision Support System for Prediction of Heart Diseases Using Naïve BayesAlgorithm", International Journal of Computer Science and Information Technologies, vol. 5, no. 2 (2014) pp.2384-2388.
11. K. Lokanayaki and A. Malathi, "Exploring on Various Prediction Model in Data Mining Techniques for Disease Diagnosis", International Journal of Computer Applications, vol. 77, no. 5, (2013), pp. 0975 – 8887.
12. Nidhi Singh and Divakar Singh, "Performance Evaluation of K-Means and Hierarchal Clustering in Terms of Accuracy and Running Time", Ph.D Dissertation, Department of Computer Science and Engineering, Barkatullah University Institute of Technology, 2012.