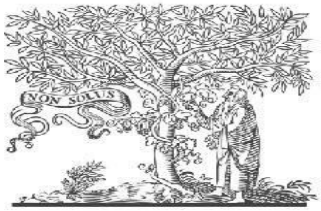
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 07th Sept 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 09

10.48047/IJIEMR/V12/ISSUE 09/03

Title **A HYBRID APPROACH FOR HEART PROBLEMS PREDICTION USING MACHINE LEARNING TECHNIQUES**

Volume 12, ISSUE 09, Pages: 16-25

Paper Authors **ABBAREDDI ANJALI, R ALTHAF, V SUBHASINI**

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

A HYBRID APPROACH FOR HEART PROBLEMS PREDICTION USING MACHINE LEARNING TECHNIQUES

ABBAREDDI ANJALI[1], R ALTHAF[2], V SUBHASINI[3]

[1]M.Tech Student, Dept of CSE, MJR College of Engineering & Technology, Piler, AP, India.
[2]Assistant Professor, Dept of CSE, MJR College of Engineering & Technology, Piler, AP, India.
[3]Assistant Professor, & HOD of CSE, MJR College of Engineering & Technology, Piler, AP, India.

ABSTRACT: Nowadays, heart disease cases are increasing rapidly, and it is imperative and concerning to predict such diseases in advance. This diagnosis is a difficult task that should be performed precisely and efficiently. The use of machine learning techniques in the field of medical diagnosis has become increasingly popular due to their ability to analyze large datasets and extract hidden patterns. Reliably predicting future cardiovascular disease is therefore an important public health priority. Large quantities of information can be used to produce forecasts and predictions using machine learning (ML). Using the techniques described here, it is possible to predict if a person has coronary heart illness and provide them with records or a diagnostic, letting them know the risk. The analysis of complex and large medical datasets can be automated by using Machine Learning algorithms and techniques. As a result, we will be removing irrelevant and unnecessary functions from the dataset in order to enhance the model's overall performance. A reliable weather forecast has become increasingly important in today's rapidly changing world.

Keywords: Analysis of Predictions, Heart Disease, Machine Learning.

A large number of deaths caused by cardiovascular disease (CVD) is a group of diseases related to the heart that has occurred worldwide few years ago and has become the utmost prevalent disease worldwide, It's not just in India. To effectively treat such diseases, a reliable, accurate, and feasible diagnosis system is needed. An array of medical datasets has been analyzed using machine learning algorithms to automate the analysis of large and complex data sets. Recently, several machine-learning techniques have been employed by researchers to aid health care professionals in diagnosing heart-

related disorders. In the Human body, the heart occupies the second most important position after the brain and blood is pumped into the heart and distributed to all body organs. In the medical field, it is important to predict heart disease occurrences. The use of data analytics enables medical centers to make accurate predictions from more information. Machine learning is a widely used concept all over the world. By providing physicians with faster access to diagnose sufferers, The healthcare enterprise will be enhanced through it. Depending on the forecast, the healthcare sector may be able to grow faster and more efficiently, saving time and money [1]. Almost four out of every five CVD deaths result from heart attacks and strokes, and one in three are occurring before 70 years of age [3]. The figure 1 shows different factors that influence cardiac disease risk.

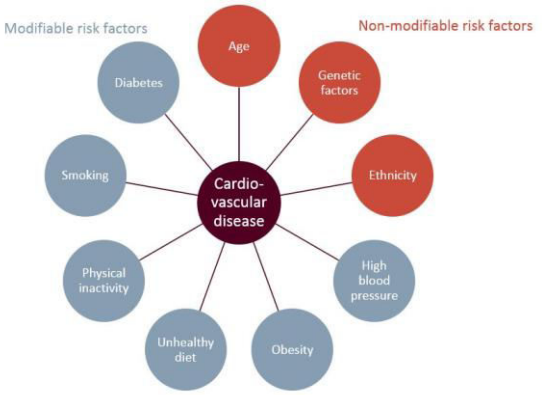


Fig. 1. Risk Factors of Cardiac Disease.

A harmful diet, inactivity, tobacco usage, and difficult alcohol feasting are the utmost imperative behavioral hazard factors for coronary heart illness and stroke [4]. Additionally, individuals can experience unnecessary blood pressure, extreme blood glucose, extreme blood lipids, and overweight or heaviness problems due to behaviors that pose a threat to their health. "Intermediate threat elements" found in number one care settings can indicate an increased risk of coronary heart attacks, strokes, and coronary heart failure, along with other health consequences [5]. Figure 2 illustrates the classification of cardiac diseases based on clinical conditions. These clinical conditions include heart failure, myocardial infarction, arrhythmias, valvular heart disease, and congenital heart defects, among others. These categories help healthcare professionals in the diagnosis and management of heart diseases.

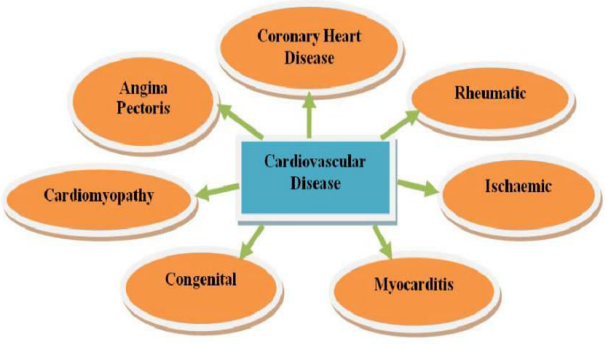


Fig. 2. Types of Cardiac Disease

These data-driven techniques,

which utilize advanced technology and algorithms, have the potential to revolutionize the field of medicine by assisting doctors in accurately diagnosing coronary heart disease and providing appropriate treatment plans. The introduction of machine learning has revolutionized various industries and sectors. Its ability to learn and improve from experience has made it a valuable tool for tasks that were previously thought to be impossible or too time-consuming. For example, in the field of weather forecasting, machine learning has greatly enhanced the accuracy and reliability of predictions.

Machine learning techniques have proven to be effective in predicting and diagnosing various diseases such as heart disease. Researchers have used machine learning to analyze medical data and compare data mining tools in order to classify heart disease using datasets such as the Cleveland dataset from UCI [6].

Computer-based decision support systems have the potential to significantly improve clinical decision-making and patient outcomes by leveraging innovative technologies includes large big data analytics and machine-learning algorithms. These hospital information systems allow for efficient management of healthcare data, ensuring that patient information is securely stored and easily accessible when needed. By incorporating artificial learning into the decision-making process, the potential for uncovering and utilizing this hidden information increases significantly. The objective of this paper is to advance an effective heart sickness prediction scheme with machine learning techniques. To achieve this objective, the authors conducted a comprehensive review of existing literature on machine-learning algorithms used for heart ailment estimation. They assessed the benefits and drawbacks of these algorithms and identified gaps in the current literature.

RELATED WORK

In [7], the utilization of data mining methods and machine learning is supported as a means to predict heart disease. Using data mining technology to uncover concealed patterns is the most essential factor. When comparing to LMT, the J48 procedures, which rely on UCI data, demonstrate the highest level of accuracy.

A research study [8] introduced the development of a prediction machine that utilizes a patient's clinical information to diagnose cardiac disorders. The machine considers various input characteristics in order to make accurate

predictions. The dataset underwent thorough examination, including information cleansing and integration. The cardiac machine has specific objectives regarding prediction accuracy. It is important to note that there is no previous knowledge available in the datasets concerning this particular software. Considering the vast amount of data sets, it is essential for the tool used to have the capability to scale and perform efficiently.

Authors in [9] suggested that precise decision-making is of utmost importance to clinical practitioners due to its high accuracy rates of 86.3% in testing and 87.3% in teaching.

The study conducted in [10] supports the findings regarding cardiovascular disease. The detection of this disease was made possible by employing data mining techniques discussed in this paper. This research can be valuable to healthcare professionals as it assesses existing methods used to extract information from the dataset. The objective can be accomplished by developing a medical decision tree based on consumption patterns.

Authors in [11] proposed a method called A Neural Fuzzy System (NFS) to analyze different cardiac conditions. The tests presented in this publication were based on research conducted on treating cardiac diseases. The primary goal for the author is to improve the efficiency of an affordable intelligent machine and enhance the performance of the current equipment. Data mining techniques are extensively utilized in this paper to boost the accurateness of predicting heart disease. Ultimately, the study showcases the significant predictive abilities of neural networks and SVM in expecting cardiac illnesses. Furthermore, the use of record-mining technologies does not show promising results in estimating coronary thrombosis heart disease.

In [12], a technique was proposed by the researchers to combine data mining performances through the Map-Reduce algorithm. As a result, in a test set comprising 45 instances, this approach demonstrated increased precision compared to a conventional fuzzy artificial neural-network. The employment of a self-motivated representation and linear scaling played a significant role in improving the outcomes achieved by this technique.

In [14], SVM and Nave Bayes models were employed for the prediction of cardiovascular disease. When evaluating their performance based on three distinct metrics, SVM demonstrated a higher level of accuracy compared to

Nave Bayes.

PROPOSED CLASSIFICATION MODEL

The Random Forest, Decision Tree, Logistic Regression, Support Vector Machine, KNN, and XGBoost are the six classification models included in the suggested hybrid model. Evaluating performance can be done by combining and comparing these six models. The classification model presented in Figure 3 illustrates how heart disease can be predicted using this approach.

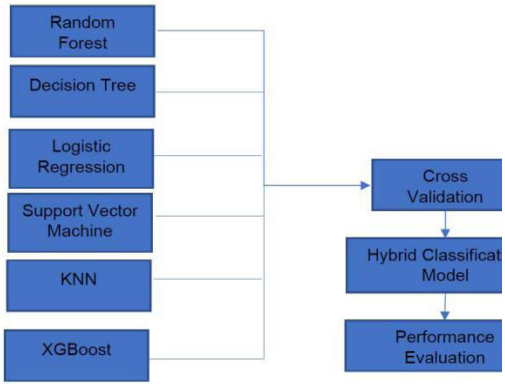


Fig. 3. Proposed Hybrid Model

This research employs a range of machine learning methods to forecast heart disease, such as the Naive Bayes classifier, logistic regression, random forest, support vector machine, decision tree classifier, and KNN. Python programming language is utilized for the implementation. Enhancing accuracy in the predictions is achieved through preprocessing the dataset and eliminating irrelevant information.

s

DATASET

This research utilizes the UCI Cleveland dataset [13], which has been extensively used for research and analysis. We utilize it for guessing heart disease. The UCI heart disease dataset consists of 307 patient chronicles data and includes 13 features. Our target label represents two modules: heart affected person or typical cases. Detailed information about the dataset matrix can be found in Table 1.

Table 1. Heart disease data description.

Serial no.	Feature name	Code	Description
1	Age	AGE	The patient's age in years.
2	Sex	SEX	The patient's sex: male = 1, female = 0
3	cp	CPT	Chest pain type: 0 = typical angina,1 = atypical angina, 2 = nonanginal pain, 3 = asymptomatic
4	trestbps	RBP	Resting blood pressure (in mm)
5	chol	CM	The patient's cholesterol measurement in mg/dl
6	fbs	FBS	The patient's fasting blood sugar > 120 mg/dl. 1 = true, 0 = false
7	restecg	REC	Resting electrocardiographic results: 0 = nothing to note, 1 = having ST-T wave abnormality, 2 = possible or definite left ventricular hypertrophy
8	Thalach	MHR	Maximum heart rate achieved
9	exang	EIA	Exercise-induced angina: 1 = yes, 0 = no
10	Oldpeak	OP	ST depression induced by exercise relative to rest checks the stress of the heart during exercise. The weak heart will stress more.
11	Slope	PES	The slope of the peak exercise ST segment: 0 = up sloping, 1= flat sloping, 2 = down sloping
12	ca	NMV	Number of primary vessels (0-3) colored by fluoroscopy.
13	thal	TS	Thallium stress result: 1, 3 = normal, 6 = fixed defect, 7 = reversible defect

DATA PREPROCESSING

In this research, the collected data underwent preprocessing. The Cleveland dataset contains 4 inaccurate accounts on NMV and two inaccurate accounts on TS. These erroneous entries have been substituted with optimal values. Subsequently, the standard scaler is employed to guarantee that each feature has an average of 0 and a variance of 1, and to standardize all the features accordingly.

FEATURE SELECTION

In machine learning, feature selection is crucial because the dataset may contain many irrelevant features that affect algorithm accuracy. By selecting features, the algorithms can be improved and their performance can be reduced. The most important features were ranked based on their relevance using different feature ranking techniques. Several factors will be considered in this study the importance of features using three well-known feature selection algorithms.

RANDOM FOREST

Predictions are made using an ensemble of decision trees using Random Forest, a supervised machine learning algorithm. The classification algorithm performs better in classification tasks compared to regression tasks. Additionally, it's been frequently utilized in various domains such as object recognition, pattern classification, and image processing to solve problems like classification, regression, and fusion. Individually decision tree is built using a subgroup of the original dataset and a subset of the available features, and the final output is determined by aggregating the predictions of these individual decision trees. The random forest algorithm is based on the principle that by growing a large number of decision trees, the algorithm can achieve convergence to the correct decision or clustering result. This approach is known as the ensemble method, which combines the predictions of multiple decision trees to enhance the overall accuracy and robustness of the model. It effectively handles extensive datasets characterized by significant dimensionality.

SUPPORT VECTOR MACHINES (SVMS)

The classification and regression capabilities of support vector machines extend to both linear and non-linear scenarios. A support vector machine is a powerful and widely used supervised classifier that is capable of accurately classifying data into different categories or classes. In the context of SVM classification, it is common practice to have two separate datasets: a training set and a test set. In the ideal situation, the machine learning algorithm would be able to perfectly classify and separate the different classes based on their inherent linear boundaries. According to the source, the Otsu thresholding method can be used to split a region that consists of two similar text lines. This observation highlights the challenge of creating

balanced data splits in datasets containing multiple labels for the same track. The best line is selected as the "Separating Line" based on specific criteria and algorithms such as Otsu's thresholding method.

The utilization of more advanced classifiers, such as Support Vector Machines, may be advantageous due to their ability to minimize both empirical risk and confidence. Using more sophisticated classifiers, such as support vector machines, can help minimize the number of errors made in various applications. The popularity of artificial neural networks in recent research can be attributed to their good overall empirical performance.

SOFTWARE USED

Python: To collect data, a web scraper programmed in Python was utilized due to its versatility, agility, and proven efficacy in web scraping tasks. Python's syntax reduces the number of lines of code programmers must write to express concepts. The Python programming language is ideal for scraping web content and extracting data, as it has a large collection of libraries and a vibrant community for coding help. It is also simple to read, understand, and write Python code, which is one of the key reasons why we use Python for web scraping.

MS EXCEL: Windows and Mac OS X users can use Microsoft Excel to create spreadsheets. Why choose MS Excel over another similar type of software that supports almost any file extension and has a variety of functionality? MS Excel features calculations, graphing tools, pivot tables, and macros. It also allows you to develop functions to use on spreadsheets using the VBA language, which is different from what you would do in Excel (Charts, Calculations, etc.).

RESULT AND DISCUSSION

This investigation focuses on the development of a diagnostic scheme using machine-learning (ML) techniques to predict the incidence of heart diseases in patients. The metadata is split into sets for train and test purposes. The researchers utilized a range of machine learning models such as Support Vector Machine, Logistic Regression, K-Nearest Neighbours using UCI data. The use of python allowed for the training and evaluation of machine learning models, as well as the division of the dataset into train and test data. The performance of the algorithms was assessed through a comparison of their accuracy scores, which can be in Figure 4 for each algorithm.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

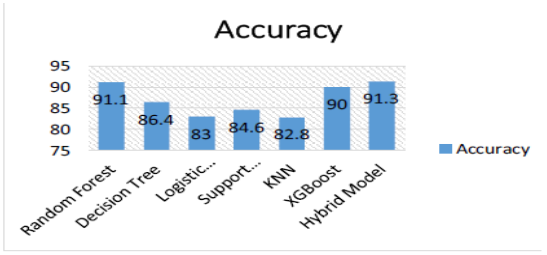


Fig. 4. Compare the hybrid model based on the accuracy

The study conducted an extensive analysis of various classification models, including Random Forest, Decision Tree, Logistic Regression, Support Vector Machine, KNN, and XGBoost and outperformed the traditional classification models in terms of recollection, accuracy, and F1-score.

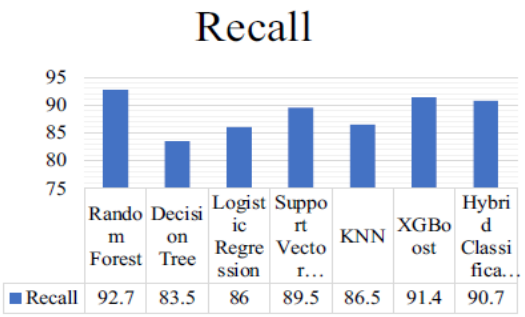


Fig. 5. Presentation measured by Recollection

The above picture clearly indicates that the combined classification approach outperforms other classification approach in terms of recall. The results presented in Figure 6 indicate that the random forest classifier outperforms other classifiers in correctly detecting the positive class.

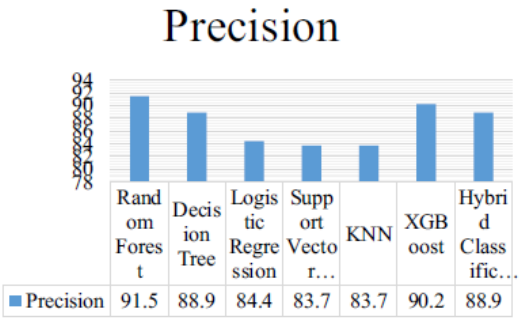


Fig. 6. Presentation measured by Accuracy

The hybrid classification model outperforms the other classification models in terms of precision as shown in Figure 7. The random forest algorithm consistently outperforms other machine learning techniques, demonstrating its superiority in predicting positive outcomes while disregarding negatives ones.

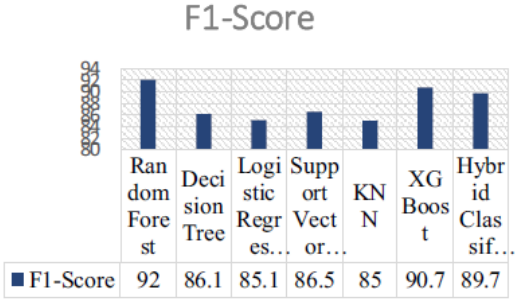


Fig. 7. Presentation measured by F1 Score

The figure 7 provides a visual representation of the performance

comparison between the combined classification approach and other classification approaches based on the F1-score metric.

CONCLUSION

This research aims to investigate and evaluate various data mining methods in order to categorize the utmost operative approaches for predicting cardiac diseases. The persistence of this research is to investigate and propose efficient and effective prediction techniques that utilize a reduced set of attributes and tests. A pre-processed and altered dataset was used in the model, which introduced potential biases and compromised the validity of the results. According to the study, the hybrid classification method attained the maximum exactness score of 91.3%. with an accuracy of 90.3%, the K-Nearest Neighbour algorithm exhibited the humblest presentation among the classification methods evaluated. In future, the potential for improving the performance of heart disease prediction lies in the utilization and optimization of neural network models.

[1] F. Desai et al., "HealthCloud: A system for monitoring health status of heart patients using

machine learning and cloud computing," INTERNET OF THINGS, vol. 17, Mar. 2022, doi: 10.1016/j.iot.2021.100485.

[2] Purushottam, K. Saxena, and R. Sharma, "Efficient Heart Disease Prediction System," Procedia Comput. Sci., vol. 85, pp. 962–969, 2016, doi:10.1016/j.procs.2016.05.288.

[3] L. Saba et al., "Global perspective on carotid intimamedia thickness and plaque: should the current measurement guidelines be revisited?," Int. Angiol., vol. 38, no. 6, pp. 451–465, Dec. 2019, doi: 10.23736/S0392-9590.19.04267-6.

[4] S. Jain and K. Chandrasekaran, "An Ensemble of Deep Recurrent Neural Networks for Detecting IoT Cyber Attacks Using Network Traffic," in Security and Privacy Issues in Sensor Networks and IoT, vol.7, no. 9, IGI Global, 2020, pp. 28–64. doi:10.1109/JIOT.2020.2996425.

[5] R. Sujatha, J. M. Chatterjee, N. Z. Jhanjhi, T. A.Tabbakh, and Z. A. Almusaylim, "Heart Failure Patient Survival Analysis with Multi Kernel Support Vector Machine," Intell. Autom. SOFT Comput., vol. 32, no. 1, pp. 115–129, 2022, doi:10.32604/iasc.2022.019133.

[6] V. S. Varale and K. S. Thakre, "Prediction of Heart Disease using Machine Learning Algorithm," Biosci. Biotechnol. Res. Commun., vol. 13, no. 14, SI, pp. 287–290, 2020, doi: 10.21786/bbrc/13.14/67.

[7] U. e Laila, K. Mahboob, A. W. Khan, F. Khan, and W. Taekeun, "An Ensemble Approach to Predict Early-Stage Diabetes Risk Using Machine Learning: An Empirical Study," SENSORS, vol. 22, no. 14, Jul. 2022, doi: 10.3390/s22145247.

[8] K. Karboub and M. Tabaa, "A Machine Learning Based Discharge

Prediction of Cardiovascular Diseases Patients in Intensive Care Units,"HEALTHCARE, vol. 10, no. 6, Jun. 2022, doi: 10.3390/healthcare10060966.

[9] H. Naz and S. Ahuja, "Deep learning approach for diabetes prediction using PIMA Indian dataset," J. Diabetes Metab. Disord., vol. 19, no. 1, pp. 391–403,Jun. 2020, doi: 10.1007/s40200-020-00520-5.

[10] M. D. A. Hossen et al., "Supervised Machine Learning-Based Cardiovascular Disease Analysis and Prediction," Math. Probl. Eng., vol. 2021, Dec. 2021, doi: 10.1155/2021/1792201.

[11] E. Giacoumidis, A. Matin, J. Wei, N. J. Doran, L. P. Barry, and X. Wang, "Blind Nonlinearity Equalization by Machine-Learning-Based Clustering for Single- and Multichannel Coherent Optical OFDM," J. Light. Technol., vol. 36, no. 3, pp. 721–727.

AUTHORS PROFILE

1. **ABBAREDDI ANJALI** Pursuing M.Tech at MJR College of Engineering & Technology, Department of Computer Science & Engineering, Piler, Andhra Pradesh.

   anjaliabbareddy@gmail.com

2. **R ALTHAF** Working as an Assistant Professor in MJR College of Engineering & Technology, Department of CSE, Piler, Andhra Pradesh.
3. **V SUBHASINI** Working as an Assistant Professor & HOD in MJR College of Engineering & Technology, Department of CSE, Piler, Andhra Pradesh.