

Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques

Shivam Verma¹, Yash Manshani², Mr. Ramesh Kumar Gupta³

¹B. Tech Student, Department of Information Technology, Shri Ramswaroop Memorial College Of Engineering and Management, Lucknow, Uttar Pradesh, India

²B. Tech Student, Department of Information Technology, Shri Ramswaroop Memorial College Of Engineering and Management, Lucknow, Uttar Pradesh, India

³Associate Professor, Department of Information Technology, Shri Ramswaroop Memorial College Of Engineering and Management, Lucknow, Uttar Pradesh, India

Abstract - Heart disease is one of the foremost significant causes of mortality in the world today. Prediction of cardiovascular disease could also be a critical challenge within the world of clinical data analysis. Machine Learning (ML) has occurred indicated to be effective in assisting in making judgments and projections From the massive quantity of

data delivered by the healthcare enterprise. We have also seen ML Techniques getting utilized in recent developments in different areas. Several types of research give only an impression into indicating heart disease with ML techniques. In this paper, we propose a completely unique method that aims at finding substantial details by applying machine learning techniques resulting in improving the exactness within the projection of cardiovascular disease. The prediction model is Instructed with various varieties of details and several known classification techniques. We generate an enhanced execution category with an the exactness level of 86% through the prediction Model for heart disorder with the hybrid random forest with a linear model

Key Words: Health Care, Heart'Disease, Machine Learning

• INTRODUCTION:

It is hard to specify heart disease because of numerous contributory danger components such as diabetes, High blood pressure, high cholesterol, abnormal pulse rate and many other factors. Several Methods in data mining and neural networks have been assigned to find out the stringency of heart Disease among humans. The stringency of the disorder is classified based on several techniques like K- Nearest Neighbor Algorithm, Decision Trees, Logistic Regression, and Support Vector Machine. The health of heart disorder is complicated and therefore, the disease must be Dealt with carefully. Not doing so may affect the heart or cause premature casualty. The viewpoint of Medical science and data mining are used for finding out various sorts of metabolic diseases. Data mining with sequence takes advantage of a crucial role in

the projection of heart disease and data Examination. The dataset with a radial basis function network (RBFN) is obtained for a category, where 70% of The data is borrowed for training and the staying 30% is used for classification. We also Introduce Computer-Aided Decision Support System (CADSS) in the field of medicine and Research. In earlier tasks, the usage of data mining methods in the healthcare business has been shown to take less time for the prediction of disease with more accurate results.

• LITERATURE SURVEY

There is ample related work in the fields directly related to this project report. ANN has been Introduced to supply the very best accuracy prediction within the medical field. The back Propagation multilayer perception (MLP) of ANN is employed to predict heart condition. The collected outcomes are compared with the results of existing models within an equivalent domain and located to be Improved. The database of heart disease victims obtained from the UCI laboratory is borrowed to Discover patterns with NN, DT, Support Vector machines SVM, and Naive Bayes. The findings are Correlated for execution and exactness with these algorithms. The proposed hybrid method Subsidies finding of 85% for conclusion, competing with the other existing methods. The Sequence without segmentation of Convolutional Neural Networks (CNN) is acquainted. Huge quantity of data base produced by the medical industry has not Been used effectively recently. The modern strategies related here reduce the expense and Enhance the projection of heart disease in a simple and beneficial way. The several different study Methods contemplated in this work for projection and classification of heart disease using ML and Deep learning (DL) techniques are highly accurate in establishing the efficacy of these methods. This study has formulated a model Intelligent Heart Disease Prediction System (IHDPS) using data mining methods, namely, Decision Trees, Naïve Bayes and Neural Network. Findings show that each method has its different Strengths in understanding the goals of the defined mining objectives.

- **MATERIALS And METHODS**

This work depicts about the requirements. It determines the appliance and programming prerequisite that maintenance needed for software to keeping in mind the end goal, to run the application properly. The Software Requirement Specification (SRS) is clarified in point of interest, which incorporates outline of this exposition and additionally the functional and non-practical necessity of this research

- **Technical Feasibility**

- Here it is considered with determining hardware and programming, this will effective fulfill the Client necessity the specialized requires of the framework should shift significantly yet may Incorporate

- The office to create yields in specified time.
- Reaction time under particular states.
- Capacity to deal with a particular segment of exchange at a specific pace.

- **Economic Feasibility**

A budgetary analysis is the frequently borrowed system for evaluating the feasibility of a launched Structure. Research more usually acknowledged as cost/favorable position examination. The method Is to center the focal points and trusts are typical casing a projected structure and a difference Them and charges. These points of interest surpass costs; as choice is engaged to diagram And realize the system will must be prepared if there is to have a probability of being embraced. There is inconsistent attempt that upgrades in exactness at call time of the system life cycle.

- **Technology used:**

Python : Python is a inferred, interact , object-oriented, and high-level Programming language. An interpreted language Python has a method ideology that underscores Code readability (notably borrowing whitespace incision to delimit code blocks rather than curly Brackets or keywords), and a syntax that permits programmers to convey theories in rarer lines Of code than might be used in languages such as C++or Java. It gives constructs that facilitate Clear programming on both minor and huge scales. Python interpreters are accessible for Several operating systems. Python, the passage enactment of

Python, is clear Source software and has a community-based modification model, as do almost all of its variant Implementations. C Pythons organized by the non-profit Python Software Foundation. Python Promotes a vibrant category network and voluntary recollection management. It benefits numerous programming paradigms, encompassing object-Oriented, essential functional and procedural, and has large and detailed standard Library

Machine Learning: Machine learning is an emerging subdivision of AI . Its primary focus is to style systems, allow them to find out and make predictions supported the experience. It trains machine learning algorithms employing a training dataset to make a model. The model uses the new input file to predict heart condition . Using machine learning, it detects hidden patterns within the input dataset to create models. It makes accurate predictions for new datasets. The database set is cleared and lost values are filled. The model uses the new input file to predict heart condition then tested for accuracy..

- **Input and Output Design:**

- **Input Design :**

The information method is the link between the evidence policy and the user. It encompasses The formulating specification and methods for data practice and people steps are crucial to place transaction data into accusable form for processing are often accomplished by inspecting the computer To read data from written or published article or it can happen by having people keying the data rapidly into the system. The design of input focuses on controlling the quantity of input required, Controlling the errors, avoiding delay avoiding extra steps and keeping the method simple. The Inputs designed in such away so that it provides security and cease of use with retaining the privacy getting correct information from the Computerized system. It is achieved by creating user-friendly screens for the info entry to handle large Volume of knowledge . The goal of designing input is to form data entry easier and to be freeform Errors. The data entry screen is meant in such how that each one the info manipulates are often Performed. It also provides record viewing facilities

- **Output :**

A quality output is one, which meets the wants of the end-user and presents the knowledge clearly. In addition network findings of processing are conveyed to the users and to Other network through outcomes. In outcome method it is committed how the evidence is to be Expelled for sudden need and also the hard copy output. It is the most significant and immediate Source data to the user.

Productive and creative output method expands the system's Connection to help user decision-making

Designing computer output should proceed in an organized, well thought out Manner; the proper outcome must be formulated while securing that every output element is meant in order that people will discover the network can use easily and effectively. When analysis design computer Output, they ought to Identify the precise output that's needed to satisfy the wants . Create article, report, or other layouts that include evidence produced by the system. The output sort of an data system should accomplish one or more.

5 Approach Methodology :-

This research aims to foresee the chances of getting heart condition as evidence of computerized prediction of heart condition that's helpful within the medical Feld for clinicians and patients. To accomplish the aim, we have discussed the utilization of varied machine learning algorithms on the info set And dataset analysis is mentioned during this research paper. This Paper additionally depicts which attributes contribute more Than the others to anticipation of higher precision. This may Spare the expense of various trials of a patient, as all the Attributes might not contribute such a considerable amount to Expect the result .

le accuracy.

- **Algorithms Used:**

- **Decision Tree :**

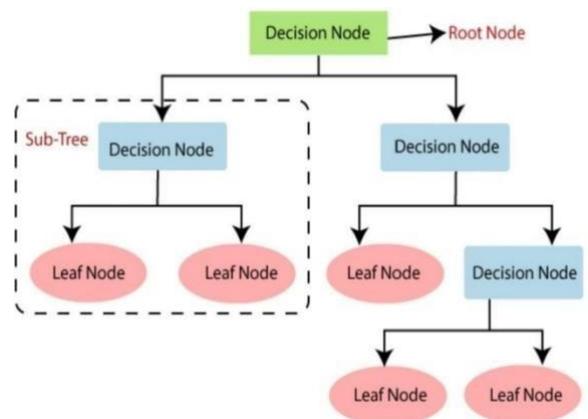
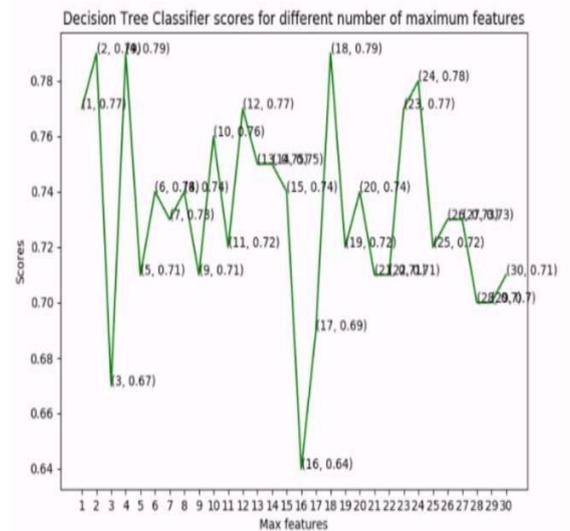
Decision tree may be a fundamental category and regression technique. Decision tree model features a tree Structure, which may describe the method of classification instances supported features. It is often Considered as a group of if-then rules, which can also be thought of as contingent probability Distributions defined in feature space and sophistication space. This algorithm divides the

- **Naïve Bayes' Classifier :**

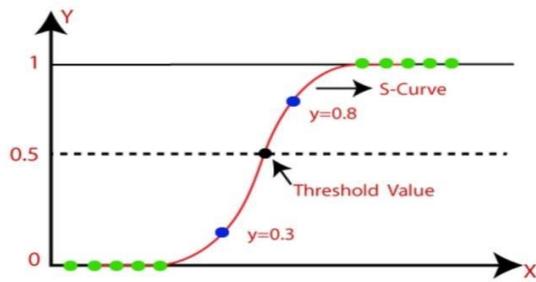
Naïve Bayes classifier is a supervised algorithm. It is an easy classification technique using Bayes theorem. It assumes Strong (Naïve) independence among attributes. Bayes theorem is a mathematical theory to get the percentage. The Predictors are neither associated with one another nor have correlation to at least one another. All the attributes alone participate to the percentage to maximise it. It is able to help With

data into two or more similar Sets based on the most significant pointers. The entropy of Each characteristic is evaluated and then the data are distributed, With predictors having ultimate information gain or mini-Mum entropy

$$\text{Entropy}(S) = \sum_{i=1}^c -P_i \log_2 P_i$$



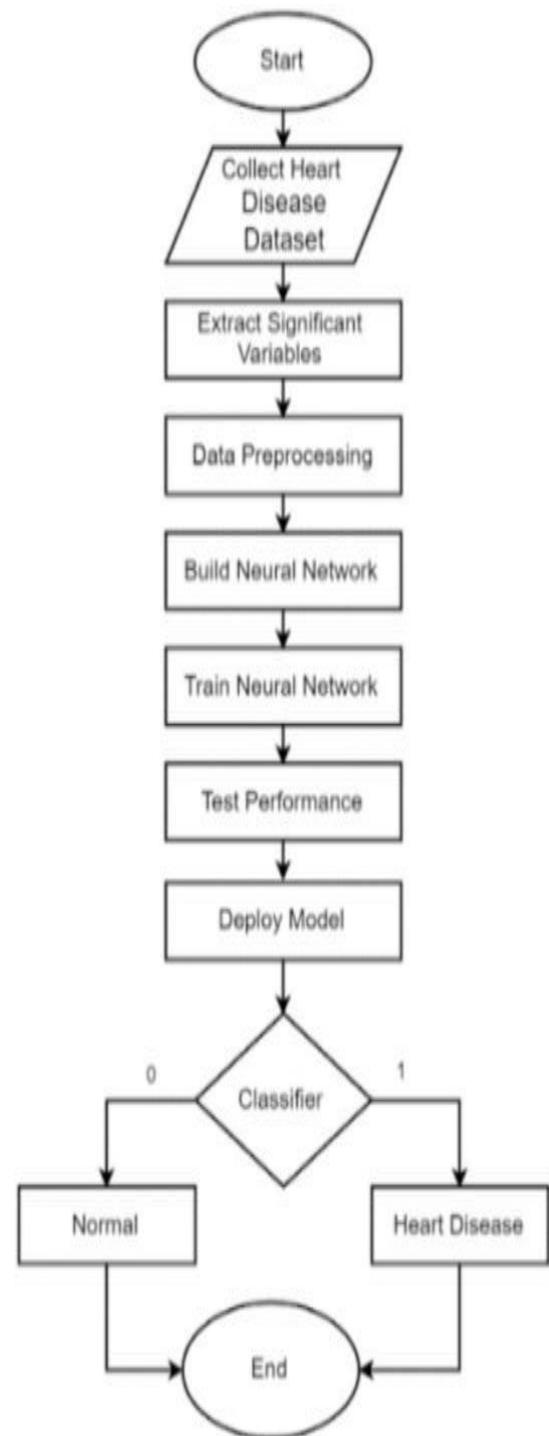
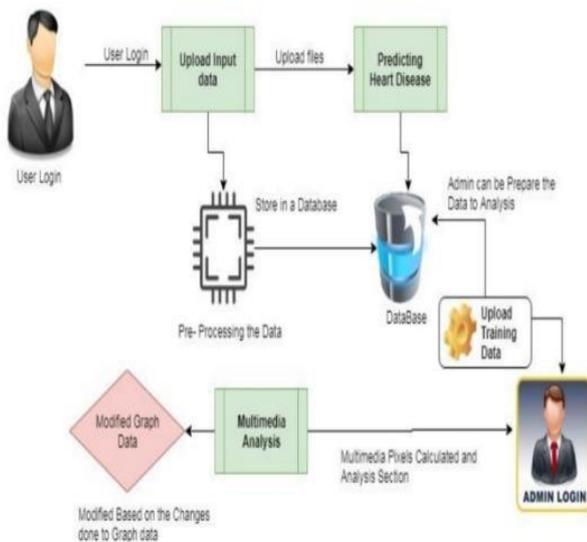
Naïve Bayes model and does not use Bayesian methods. Many complicated real-world circumstances use Naïve Bayes classifiers. An accuracy of 84.1584% has been achieved in Naïve Bayes with 10 most important predictors chosen using SVM-RFE.



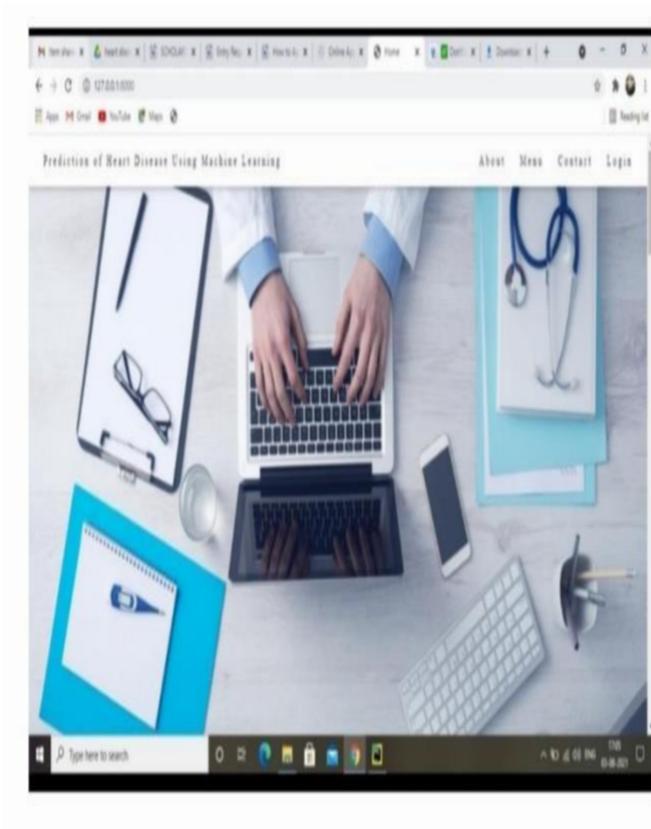
Data Pre- Processing :

The real-life evidence includes huge numbers with losing and noisy data. These data are pre-processed to overcome such issues and make predictions vigorously. Cleaning the collected data usually has noise and missing values. To get an accurate and effective result, the data need to be cleaned in terms of noise and missing values. To be filled up. Modification it shifts the configuration of the data from one form to another to make it more understandable. It pertains to Smoothing, normalization,

and accumulation tasks.



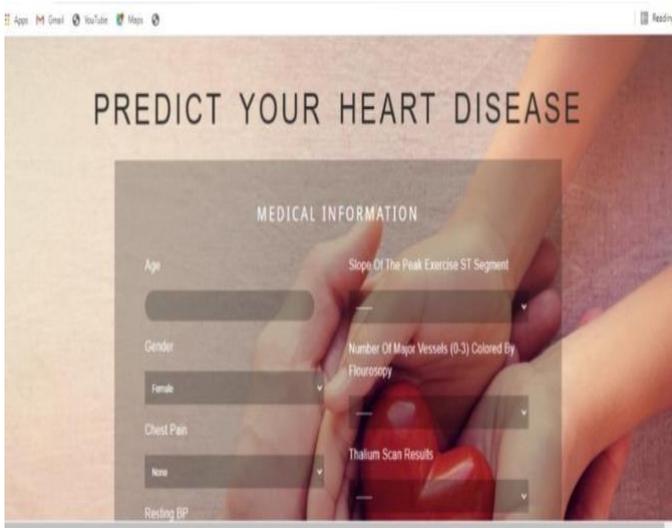
• Home Page Module :

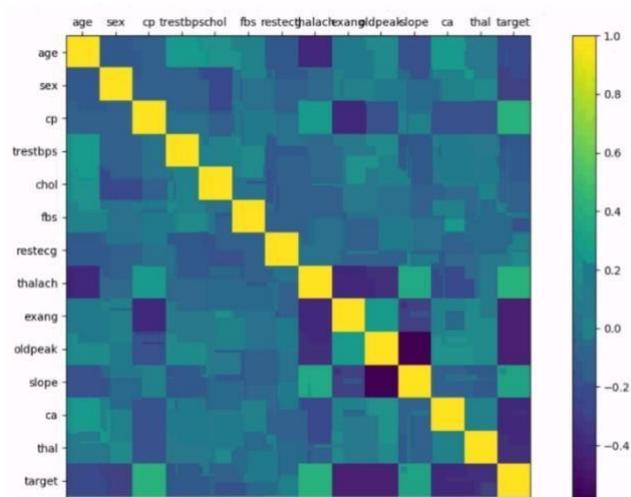


• 9 Data Set :

Age	Sex	CP	Trestbps	tbs	Restecg	Thalach	Exang	Oldpeak	Slope	ca	Thal	num
63	1	1	145	233	1	2	150	2.3	3	0	6	0
67	1	4	160	286	0	2	108	1.5	2	3	3	1
67	1	4	120	229	0	2	129	2.6	2	2	7	1
37	1	3	130	250	0	0	187	3.5	3	0	3	0
41	0	2	130	204	0	2	172	1.4	1	0	3	0
56	1	2	120	236	0	0	178	0.8	1	0	3	0
62	0	4	140	268	0	2	160	3.6	3	2	3	1
57	0	4	120	354	0	0	163	0.6	1	0	3	0

• Prediction Page Module





• **Result :**

Intend for this study is to indicate whether or not a patient Will develop heart disease. This study was done on supervised machine learning classification techniques using Naïve Bayes, decision tree, Logistic regression algorithm, K-Nearest Neighbor on UCI repository. Several operations using various Classifier algorithms were performed through the WEKA Tool. Research was performed on 8th generation Intel Corei7 Having an 8750H processor up to 4.1 GHz CPU and 16 GB Ram. Dataset was categorized and divide into a training set and A test set. Pre-processing of the data is done and supervised Classification techniques such as Naïve Bayes, decision tree The data were pre-processed and then used in the model. K-nearest neighbor, Naïve Bayes, and random forest are the algorithms indicating the nicest outcomes in this criterion.

• **Conclusion :**

Specifying the processing of natural healthcare data of heart information will help in the long term Saving of human lives and early detection of abnormalities in heart conditions. Machine learning Techniques were used in this work to filter data and supply a replacement and novel understanding Towards heart situation. Heart disease projection is demanding and very significant in the medical.

However, the mortality ratio can be drastically regulated if the disease is observed at the first Stages and preventive criteria are accepted as soon as feasible. Further extension of this study is very desirable to direct the investigations to real-world datasets rather than just theoretical Approaches and simulations. The proposed hybrid HRFLM strategy is utilise stirring the Characteristics of Random Forest (RF) and Linear Method (LM). HRFLM verified to be quite Accurate within the prediction of heart condition . The future course of this research can be performed With

diverse mixtures of machine learning techniques to better prediction techniques. Furthermore, New feature selection methods are often developed to urge a broader perception of the many Features to extend the performance of heart condition prediction.

• **References :**

[1] A. S. Abdullah and R. R. Rajalaxmi, "A data mining model for predicting the coronary heart Disease using random forest classifier," in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 22–25.

[2] Chauhan R, Bajaj P, Choudhary K, Gigras Y. Framework to Predict health diseases using attribute selection mechanism. In: 2015 2nd international conference on computing for sustainable global development (INDIACom). IEEE. p. 1880–84. [3] N. Al-milli, "Backpropagation neural network for prediction of heart disease," J. Theorappl.Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.

[4] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, "Analysis of neural networks based heart disease prediction system," in Proc. 11th Int. Conf. Hum. Syst. Interact. (HIS), Gdansk, Poland, Jul. 2018, pp. 233–239.

[5] P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using Weighted fuzzy rules," J. King Saud Univ.-Comput. Inf. Sci., vol. 24, no. 1, pp. 27–40, Jan. 2012. Doi: 10.1016/j.jksuci.2011.09.002.

[6] L. Baccour, "Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets," Expert Syst. Appl., vol. 99, pp. 115–125, Jun. 2018. Doi: 10.1016/j.eswa.2018.01.025.

[7] H. A. Esfahani and M. Ghazanfari, "Cardiovascular disease detection using a new ensemble Classifier," in Proc. IEEE 4th Int. Conf. Knowl.- Based Eng. 1011–1014