e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:11/November-2022 Impact Factor- 6.752 www.irjmets.com

HEART DISEASE PREDICTION AND DOCTOR RECOMMENDATION SYSTEM USING MACHINE LEARING

Juned Hakim Shaikh*1, Abhishek Ganesh Takale*2, Mrs. K.T. Mohite*3

*1,2Student, Department Of Computer Engineering, Sinhgad Academy Of Engineering, Pune, Maharashtra, India.

*3Professor, Department Of Computer Engineering, Sinhgad Academy Of Engineering, Pune, Maharashtra, India.

DOI: https://www.doi.org/10.56726/IRJMETS30965

ABSTRACT

Heart-related illnesses have emerged to be the most dangerous condition worldwide in recent years, accounting for a significant portion of deaths worldwide. not just in India, but everywhere. There is a demand for a trustworthy and a timely and accurate method of disease diagnosis and treatment for appropriate care. Automating the diagnostic process and therapy Machine Learning methodologies are using the medical dataset as a test case. As technology develops, Machine Learning techniques have been employed by researchers in the detection of cardiac problems to aid in healthcare industry. By displaying "True" or "False," if a person has that disease or not, it is also expected when matching to a specific heart disease. Future medical therapy could greatly benefit from the prediction of such a system.

Keywords: Disease Prediction, Machine Learning Classification, Doctor Recommendation, Python, SVM Algorithm.

I. INTRODUCTION

Heart disease is currently one of the leading causes of death, abnormal changes in cholesterol, blood pressure, Heart-rate, etc. Heart disease is one of the most prevalent conditions in the significant obstacles; there are many parameters and it requires intricacy to correctly forecast this condition. There is a huge need for study in the field of anticipating heart illness in people because, according to the most recent WHO survey, doctors could only properly predict 67% of heart disease. Researchers employ data mining techniques to assist healthcare practitioners in the identification of cardiac disease. A vast amount of clinical data is available, but where crucial information is concealed is rarely accessed and remains untapped. The support vector machine algorithm is being used to forecast the risk level of heart disease. Age, sex, the type of chest discomfort, resting blood pressure, cholesterol in mg/dl, fasting blood sugar, and other 14 variables are input into the algorithm. Once the risk level forecast is carried out, the user would be given advice on precautions, which would enable them to manage their amount of risk temporarily. The topic of heart-related diseases is covered in great detail in this paper. The user's comprehension of the numerous prediction-related parameters and their options for controlling those factors are covered in more detail later in the scope.

II. MOTIVATION

All across the world, different diseases impact people. Heart disease is today a serious concern and causes significant mortality in both men and women. WHO estimates that 17.9 million people worldwide die from heart disease each year, accounting for 31% of all facilities. For the prediction of heart diseases, there are machine learning tools and methodologies available, however there are yet no suitable models that can forecast the disease.

III. LITERATURE SURVEY

The available studies that are surveyed concentrate on the forecast of cardiovascular illness. The transformed data set is the system is trained using data obtained from a variety of sources and are subsequently put to use for forecasting. T. Princy and J. Thomas [2] use data mining techniques that are currently being used in research for the prediction of heart disease to provide specifics regarding various knowledge abstraction strategies.

In this study, algorithms are used to assess three types of data mining methods support vector machine Algorithm on sets of medical data. While Priyanga and Naveen [3] concentrate primarily on developing a decision support system for heart disease prediction using the Naive Bayes algorithm.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:11/November-2022 Impact Factor- 6.752 www.irjmets.com

The Cleveland dataset is used to construct a web application that collects user input and can access concealed information about heart disease from a historical database. M. Gandhi and S. Singh's [4] focus on data mining classification techniques for data discovery. For data classification and knowledge extraction, several data mining classification algorithms offer advantages and disadvantages. Richa Sharma, Purushottam, and K. Saxena[8] illustrated how to obtain the risk level from the heart disease database. The screening of clinical data for cardiac patients is contained in the input database. To increase the effectiveness of the mining process, the database is first pre-processed. The prediction is made using a decision tree. As reliable predictions are required in the medical field, the studies mostly concentrate on determining the effectiveness of various algorithms.

Authors Ahmed M. Alaa and Senthil Kumar Mohan experimented with a variety of variables and found that a random hybrid forest produced results with an accuracy of 88.7%. In order to make the decision support system simpler, the goal of this work is to enhance the savvy treatment. The diagnosis of heart illness by keeping an eye on a person's heartbeat is the subject of this extensive paper. You are allowed by the framework to specify your pulse's criteria. After establishing these limitations, a person can start monitoring their heart rate, and whenever it exceeds a certain threshold, they are warned that they are at risk of having a heart attack or coronary failure and are having a high pulse.

This study examines traditional supervised binary classification, where the dataset contains a number of attributes. Plasma glucose concentration is included in the dataset. Body mass index and blood pressure in mm Hg Years of age, etc. In order to identify those who have the condition, a number of factors are used, each with distinct characteristics. To solve the issue, we must analyze data, make any necessary adjustments, apply ML, train a model, examine the output of the trained model, and continue with several techniques until we get the most accurate result. When developing software or websites, it's important to understand the framework requirements and obtain the necessary data to communicate with customers and providers.

IV. HARDWARE AND SOFTWARE REQUIREMENTS

Table 1. Hardware Requirements

SR. No.	Hardware	Description
1.	Processor	Pentinum-IV
2.	Speed	1.1 GHz
3.	Hard Disk Space	40 GB
4	Device	HP Pavilion
5.	Others	Other required standard computer peripherals, such as keyboard and mouse.

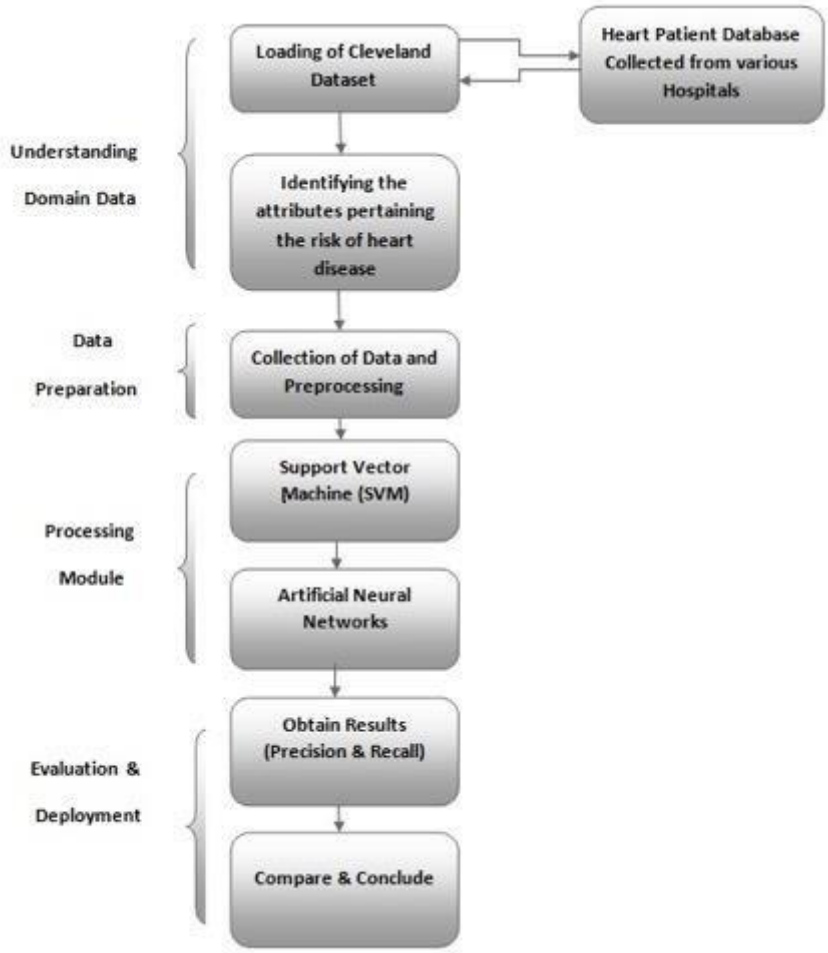
Table 2. Software Requirements

Sr. No.	Software	Description
1.	Operating System	Microsoft Windows 10
2.	IDE	Spyder
3.	Frontend	Tkinter
4.	Backend	Python
5.	Database	DBSQ Lite

V. ARCHITECTURE

In this section we mentioned about the system architecture. Fig represents the outline of system architecture. The nucleus modules of the proposed system consist of

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:11/November-2022 Impact Factor- 6.752 www.irjmets.com



VI. METHODOLOGY

Medical Dataset:

The Cleveland heart disease data set is retrieved from the UCI machine learning repository [5] for use in this paper. Despite the fact that there are 303 records with 76 attributes, we only used the 14 most relevant attributes. In the used data set, HD is present in 45.54% of the cases. Table 1 displays the 14 attributes that were used along with their descriptions.

Attributes	Description
Age	Age in years(29-77)
Sex	Male=1, Female=0
Fbs	Fasting blood sugar >120 mg/dl, Value 1=yes, Value 0=no
Cp	Chest pain types Value 1= typical angina, Value 2= atypical angina Value 3= non-angina, Value 4= asymptomatic
Trestbps	Resting blood pressure in mm Hg [94-200]
Chol	Serum cholesterol in mg/dl [126--564]
Restecg	Resting electrocardiographic , Value 0=normal, Value1= having ST-T wave abnormality

Data Pre-processing:

The most important first step in any project is data preparation. model of prediction It aids in the transformation of data into an comprehensible format to improve model efficiency Medical In general, data are

incomplete, lacking attribute values, and because it contains outliers or irrelevant data. The predicted attribute (num) in the original data set had five values: 0 indicated no HD, 1 to 4 indicated different levels of HD, and 5 indicated no HD. We are only interested in the presence or absence of HD in this study, not in the exact disease classification. As a result, the class attribute is reclassified as a binary value of 0 or 1, indicating the absence or presence of HD in the patients.

Data Classification:

Support Vector Machine (SVM) is a supervised learning model that is commonly used in classification. It is defined as finite-dimensional vector spaces, with each dimension describing a feature of a specific sample. A support vector machine seeks the best hyper-plane with the highest margin of separation between two classes. SVM has been shown to be an effective method in high-dimensional space problems due to its computational competence on large data sets.

VII. CONCLUSION

To make effective clinical decisions in the medical field plays an important role because there is a need for a dependable and accurate system. The paper investigates how machine learning and Support vector machine algorithm can be used to predict the heart disease by taking certain parameters into account that, cholesterol, heart rate, age, and gender are a few examples. Making use of these algorithm predicts the level of heart risk based on parameters. The disease between 0 and 1, where 0 represents no risk and 1 represents a high risk shows the presence of danger Based on the predicted risk level. Some precautionary measures are suggested to the user in order to reduce risk and stay fit.

VIII. REFERENCES

[1] T. Princy, J. Thomas, "Human Heart Disease Prediction System using Data Mining Techniques," International Conference on Circuit, Power and Computing Technologies [ICCPCT], 5090-1277, 2016.

[2] Priyanga, Naveen, "Web Analytics Support System for Prediction of Heart Disease Using Naive Bayes Weighted Approach (NBwa)," Asia Modelling Symposium (AMS), 2376-1172, 2017.

[3] M. Gandhi, S. Singh, "Predictions in Heart Disease using Techniques of Data Mining," 1st International Conference on Futuristic trend in Computational Analysis and Knowledge Management, 4799-8433, pp. 520-525, June 2015.

[4] Purushottam, K. Saxena, R. Sharma, "Efficient Heart Disease Prediction System using Decision Tree", 4799- 8890, International Conference on Computing, Communication and Automation (ICCCA), pp. 72-77, July 2015.

[5] Mrs. Shital. P. Chattar, Prathamesh Wadile, Bhushan Saraf, Omkar Kardile and Aditya Pawar." IOT Based Heart Disease Prediction." International Journal for Research in Applied Science & Engineering Technology (IJRASET) (2021):2321-9653.

[6] Megha Vyas, Priya Kunwar, Vinod Joshi, Mihir Maheshwari and Mr. Girish Ameta. "Software for Disease Prediction." International Journal for Research in Applied Science & Engineering Technology(IJRASET) :2321-9653,2020.

[7] Fayyad U.M, Piatetsky-Shapiro G, Smyth P., "Knowledge discovey of data mining : Towards a unifying framework in KDD", 1996 august , Vol 96, pp.82-88.

[8] Kurgan L.A, Musilek P ,"A survey of knowledge discovery and data mining process model," The knowledge engineering review 2006, 21(1), 1-24.

[9] Das R Turkoglu I, Sengur A 2009, "Effective diagnosis of heart disease through neural network ensembles", Expert system with applications 36(4), 7675-7680.

[10] Dewan A, Sharma M, " Prediction of heart disease using hybrid technique in data mining classification," In computing for sustainable global development (INDIA Com), 2015 2nd international conference on (pp. 704-706). IEEE March 2015.

[11] Chandna D, "Diagnosis of heart disease using data mining algorithm", International journal of computer science and information technologies, 5(2), 1678-1680.