Analysis of Coronary Diseases Using Machine Learning Models (ACDMLM)

V. Kalyan Viswanath[1], VPV Bharathi[2]
[1]*MTech Scholar, Raghu Engineering College, Visakhapatnam, India*
[2]*Asst. Professor, Raghu Engineering College, Visakhapatnam, India*

Abstract: The term "cardiovascular diseases" refers to a broad category of ailments that impact the heart. Approximately 20.5 million people die from these coronary diseases (CVDs) each year. In the last few decades, it has also been the leading cause of death globally. Right now, a precise and trustworthy method for obtaining an early disease diagnosis through task automation is required in order to carry out efficient management. To assist medical practitioners in diagnosing cardiac disease, numerous researchers employed a variety of data mining techniques. Nonetheless, fewer tests may be needed if data mining is used. A quick and efficient diagnostic method is essential to lowering the number of heart disease-related deaths. People can alter their lifestyles with the aid of early prognosis. If necessary, it also guarantees appropriate medical care. A quick and efficient diagnosis method is required to lower the number of heart disease-related deaths. By using several data mining approaches, including logistic regression, nearby K closest decision trees, and support vector machines, the suggested work forecasts the likelihood of cardiac illnesses. Consequently, a comparative analysis of the effectiveness of several machine learning algorithms is presented in this paper. This research describes a web-based system that uses basic parameters like smoking, diabetes, and cholesterol to estimate a person's risk of developing heart disease. A web-based approach to forecast the likelihood of developing coronary heart disease is established in this research. The test findings confirm that, in comparison to other applied machine learning methods, the Support Vector Machine reached a maximum accuracy of 86.76%.

Keywords: Machine Learning, Health care, Cardio Vascular Diseases, Classification, Logistic regression, K-nearest neighbors, Decision trees, Support vector machine.

I. INTRODUCTION

Our paper's primary goal is to anticipate heart conditions over the next ten years by utilizing a variety of data processing technologies. An essential organ of the physical body is the heart. It circulates blood to every region of our body. Cardiovascular Disease (CVD) is the leading cause of death globally, according to the World Health Organization.

The World Health Organization estimates that heart-related illnesses account for 17.7 million deaths year, or 31% of all fatalities worldwide. Heart-related illnesses are now the main cause of death in India as well. According to the 2016 Global Burden of Disease Report, which was published on September 15, 2017, heart diseases claimed the lives of 1.7 million Indians in 2016 [1]. In the US, a heart attack often results in death every 40 seconds. This makes it very evident that heart-related diseases require additional attention. This resulted in annual spending of more than $200 billion in the United States alone. This is likewise expected to rise rapidly in the years to come.

Stable angina, unstable angina, sudden cardiac death, and other conditions are typically included in the category of coronary heart disease. Chest pain, bloating, swollen legs, breathing problems, exhaustion, and an abnormal heartbeat rhythm are the primary signs of heart disease. Heart disease is caused by a number of variables, including smoking, poor diet, stress, age, and obesity [2].

A poor diet and excessive alcohol use also increase the risk of heart problems. In the majority of situations, it is difficult to tell if heart disorders can be detected early. As a result, the rise in computer-assisted detection also benefited in the analysis and comprehension of medical diagnosis.

Predicting these diseases accurately and efficiently is currently the main problem. Machine learning, which uses Python and its libraries for training and testing, is one such capability. The practice of extracting valuable information from a sizable collection of raw

data is known as data mining. Simply said, machine learning is the branch of data science that can effectively manage big datasets. Therefore, machine learning has a lot to offer the medical industry in terms of disease diagnosis and prediction.

This study compares a number of machine learnings algorithms, including support vector machines, decision trees, K-nearest neighbors, and logistic regression. The scope of future research and many advancements potential is also mentioned in the report.

One aspect of artificial intelligence is machine learning, which has the ability to learn on its own and from prior experiences, enabling it to make judgments and forecast outcomes. Using the dataset as input, we first train the classification algorithm. The model then learns from the data and can identify patterns in the dataset. It can then guess which class the data belongs to by testing with fresh data.

II. LITERATURE SURVEY

KetutAgung Enrico et.al [3] a system was proposed by him for heart disease prediction using KNN algorithm with an accuracy of 81.85%. Using KNN, with increase of number of parameters the performance decreases and it considers 90% of data for training which is computationally expensive.

Himanshustal [4] briefly discussed about large and small data set of heart diseases prediction. They shared that small data set take minimum time for training as well as testing and performed prediction using SVM and KNN algorithm. It also discussed about prediction of heart diseases and prove that some algorithms of machine learning do not perform better for accurateness predication.

Avinash Golande and et. al. [5] studies different ML algorithms that could be used for classification of heart disease. Research was carried out to study Decision Tree, KNN and K-Means algorithms that can be used for classification and their accuracy were compared. This research concludes that accuracy obtained by Decision Tree was highest further it was inferred that it can be made efficient by combination of different techniques and parameter tuning.

Lakshmana Rao et al,[6] Heart Disease Prediction in elements like diabetes, current smoker, high cholesterol, etc. contribute So, it is difficult to distinguish heart disease. Different systems in data mining have been utilized to discover the seriousness of heart disease among people. Machine learning makes rationale dependent on chronicled information.

Stephen F. Wenget.al, [7] studied application of various machine learning algorithms to improve the prediction in cardiovascular diseases. They showed that different Machine Learning algorithms are highly successful for improving the accuracy of prediction in CVD diseases, but it needs more patient records. So, the more the data, the better the results.

Marjia Sultana et.al, [8] traversed various datasets for heart disease illness and determined the usage of various Machine Learning algorithms with them. Obviously, datasets are to be preprocessed before applying any Machine Learning algorithms. They also suggest the various features that plays important for accuracy determination.

Prajakta Ghadge et al. [9] researched an intelligent heart attack prediction system. The main intention of this article was to find a system that uses big data and data mining modelling techniques. Therefore, the system can identify any hidden knowledge and pattern in the vast data.

AH Chen et al. [10] presented a heart disease prediction system which can help doctors predict the art disease status with the help of clinical data. The C language is used an artificial neural network for classification and prediction of heart disease. Programming languages like C and C# are used to develop system and with an accuracy of 80%.

III. PROPOSED METHODOLOGY

A web-based system is developed which consist of efficient method in AI to predict the possibility of a person to get coronary heart disease. The basic factors that both systems consider for prediction are gender, age, cholesterol, blood pressure, diabetes, smoking, family history and physical activity [11].

This paper describes that the proposed system provides a web-based interface to the user with

machine learning system to predict the chance of occurring Coronary Heart Disease. Initially, the webpage with form will appear where user can enter the details about the basic risk factors of coronary heart disease [11]. Using these details, system will predict the result based on the rules defined. Here, Fig 1 shows the Block diagram of Proposed System.

User can give details about the basic risk factors of CHD. These details are used to predict the result by checking it with the model created when the dataset is trained by using Machine Learning algorithm [11].

The following are some of the devices through which a user can enter his details.

Device 1: A mobile phone that has an active internet connection, can be used to enter the details of the patient i.e, Age, Gender, Blood Pressure, BPM etc. A user can go through the website and there's no need to login or register to it. He can just enter the details that are needed and click on submit to get the results. A user can go back and reset by clicking the Home button.

Device 2: A user can also enter his details from a laptop/desktop with internet connection. The process is same as in mobile phone and laptop as well. So, a user can enter the corresponding details by visiting the website.

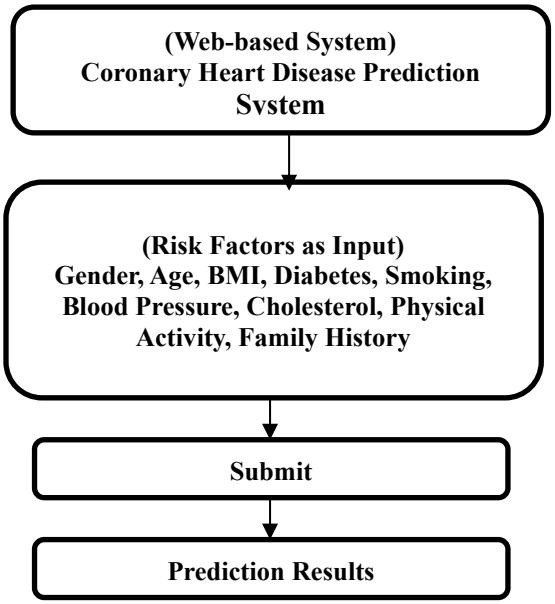

Fig 1: Block Diagram of Proposed System

Web-based Implementation:

Coronary Illness hazard prediction System is a web-based system created using web scripting language machine learning system to predict the possibility of occurring CHD. There is no need of login or sign up for user to check their status and user need not give any of their personal details. Any user can use the system directly [11].

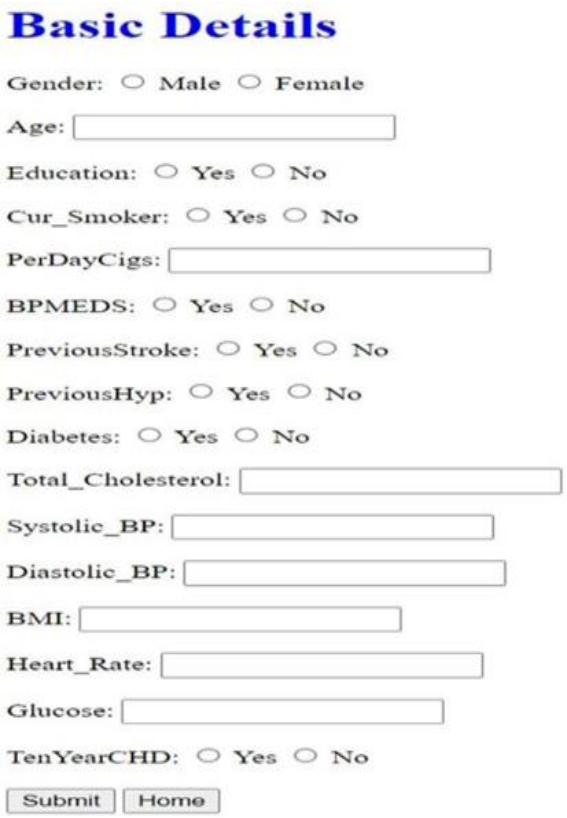

Figure 2: Input given to system

Initially, the user has to give all details about the basic risk factors that is asked for, in that page. Then they have to click the "Submit" button after giving the details that are required. When the click "Submit" button the server directs the details to the machine learning model that is trained and tested. Then the system checks the input value with the model of machine learning system and predicts the result which will be displayed back to the user. If they want to go back to main page, then they can click on 'Home' button.

Implementation of ML Algorithms for prediction of CVD

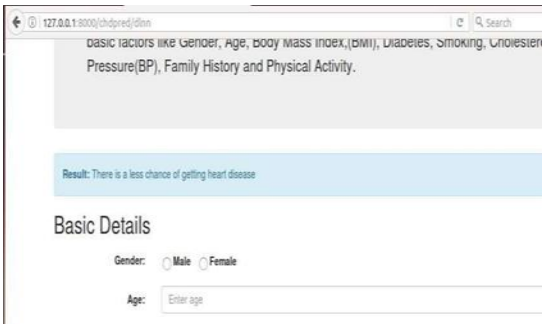

Figure 3: Web-based Result

Aim: To predict with the outcome of whether a patient ought to be determined to have a coronary illness in the next ten years. The goal of this classification is to predict whether the patient has 10-year risk coronary heart disease in the future. It includes over 15 attributes. A variable from each attribute is a potential risk factor.

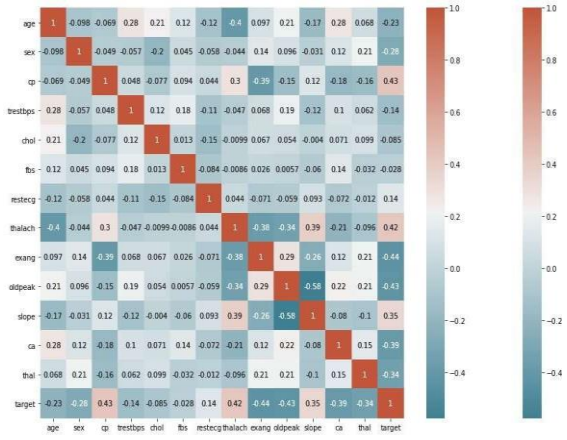

Figure 4: Information is emphatically or adversely associated with our indicator(target).

IV. PROPOSED DATASET

Attributes-

Gender: male or female (Nominal) Age: Patient's Age

Education: Is he/she currently pursuing education Cur_Smoker: whether patient is a current smoker or not PerDayCigs: cigarettes smoked per day

BPMEDS: blood pressure medication of the patient PreviousStroke: patient previous stroke information PreviousHyp: whether patient was hypertensive or not Diabetes: whether the patient had diabetes Total_Cholesterol: total cholesterol level Systolic_BP: systolic blood pressure

Diastolic_BP: diastolic blood pressure BMI: Body Mass Index

Heart_Rate: Heart Rate of the patient Glucose: glucose level of the patient

Target variable to predict:

Risk of coronary heart disease (10 years) – {1:'yes', 0:'no'} Here, Table 1 shows the rows and columns of the dataset.

Information Analysis: Correlation is nothing but an indication about the changes between two variables. The relationship of factors utilizing positive Correlation Matrix in Figure 2. Using this we can see there is a positive connection between diabetes and glucose levels in the body. This looks good because the more is the glucose in the body leading to diabetes.

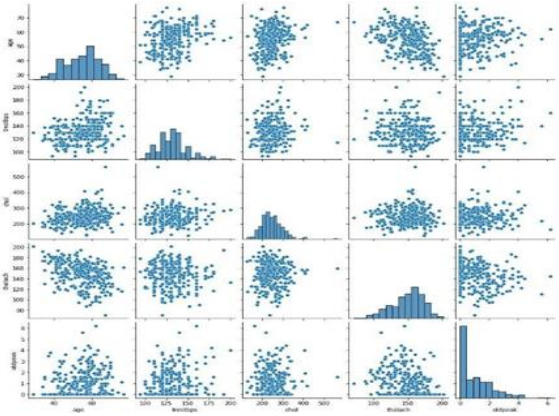

Figure 5: Plot we make a more modest pair plot with just the consistent factors, to jump further into the connections

Pair plots are likewise an extraordinary method to quickly see the relationships between all factors.

Modelling and training: Demonstrating different algorithms on the aimed information which gives the most important elevated precision. We will analyse the precision of comparative on different machine learning algorithms.

V. RESULTS AND DISCUSSIONS

Model 1: Logistic Regression

The target variable is selected and the probability of the variable occurrence is predicted by implementing

the logistic regression algorithm. Logistic Regression is a classification algorithm mostly used for binary classification problems. The data coded of the target variable is in binary nature (0 or 1) [2]. Here, table 1 shows the result.

	precision	recall	f1-score	support
0	0.77	0.67	0.71	30
1	0.71	0.81	0.76	31
accuracy			0.74	61
macro avg	0.74	0.74	0.74	61
Weighted avg	0.74	0.74	0.74	61

Table 1: Accuracy with Logistic Regression.

Logistic regression is a simple yet very effective classification algorithm so it is commonly used for many binary classification tasks.

MODEL 2: K-NN (K-NEAREST NEIGHBORS)

The K-Means splits the data into N groups. Each one is determined by a centroid, which is randomly generated artificial data, to represent the entire group [12]. The following visualization represents the centroids. Then each sample is allocated in the group of the nearest centre. K-means is vastly used for clustering in many data science applications, especially useful if you need to quickly discover insights from unlabeled data. We see how to use k-Means for customer segmentation.

	precision	recall	f1-score	support
0	0.78	0.7	0.74	30
1	0.74	0.81	0.77	31
accuracy			0.75	61
macro avg	0.76	0.75	0.75	61
Weighted avg	0.76	0.75	0.75	61

Table 2: Accuracy with K-NN

Model 3: Support Vector Machine

They are amazingly adaptable managed AI calculations which are utilized both for characterization and relapse. SVM models will categorize new text after being fed sets of named training data for each group. Here, table 3 shows the report of Support Vector Machine.

	precision	recall	f1-score	support
0	0.78	0.7	0.74	30
1	0.74	0.81	0.77	31
accuracy			0.75	61
macro avg	0.76	0.75	0.75	61
Weighted avg	0.76	0.75	0.75	61

Table 3: Accuracy with SVM

MODEL 4: XGBOOST

It is an algorithm which is implemented to make decisions for classifying the given data records (here patient's details). It supports all programming languages and is portable with remote storage, the following table 4 shows the accuracy obtained.

	precision	recall	f1-score	support
0	0.84	0.7	0.76	30
1	0.75	0.87	0.81	31
accuracy			0.79	61
macro avg	0.79	0.79	0.78	61
Weighted avg	0.79	0.79	0.79	61

Table 4: Accuracy with XGBoost

Highlight Importance

This section gives the importance factor which is causing the disease with the score comparison of the selected variables. The accuracy of the model mostly depends on the selection of the factors and the ranges of the values which a variable assigned with. The calculated result of each factor leads to selection and inclusion of the factor in the model if the result is low then factor is omitted if it is showing desirable results it is included. The correlated factors of the included factor are examined to consider them for including them in the decision.

CONCLUSION

The heart related disease detection, prediction and medication aspects are always challenging and crucial for patient. It is always risk if the heart patients are not

treated in time and if they are far from their home town or if the regular doctored is not available. The Support vector machine was the best performing model in terms of accuracy and the F1 score. Its high AUC shows that it has a high true positive rate and the SMOTE technique helped in improving the model's sensitivity by balancing the dataset, this is when compared to the performance metrics of other models on different notebooks on the same dataset. Furthermore, the advancements and availability of internet enables users to access irrespective of their economic and social factors.

This model not only increases the confidence of the patient but also enhances the doctor's ability to treat with utmost care. The treatment module of the model is completely based on the quality and availability of internet connection; thus, care should be taken to maintain the internet available.

REFERENCE

[1] Baban.U. Rindhe, Nikita Ahire, RupaliPatil, ShwetaGagare ManishaDarade , Heart Disease Prediction Using Machine Learning, International Journal of Advanced Research in Science, Communication and Technology, Volume 5, Issue 1, May 2021, ISSN (Online) 2581- 9429.

[2] B Padmajaa, Chintala Srinidhi b, Kotha Sindhuc, Kalali Vanajad , N M Deepikae , E Krishna RaoPatrof, Early and Accurate Prediction of Heart Disease Using Machine Learning Model, Turkish Journal of Computer and Mathematics Education, Vol.12 No.6 (2021), 4516- 4528.

[3] I Ketut Agung Enriko, Muhammad Suryanegara, Dinda Agnes Gunawan, "Heart-disease prediction system using k-Nearest neighbor algorithm with simplified patient's health parameters", 2016.

[4] Himanshu Sharma and M A Rizvi. (2017). "Prediction of Heart Disease using Machine Learning Algorithms: A Survey" International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 8, IJRITCC August 2017.

[5] Avinash Golande, Pavan Kumar T, Heart Disease Prediction Using Effective Machine Learning Techniques, International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019.

[6] A. Lakshmana rao, Y. Swathi, P. Sri Sai Sundareswar, Machine Learning Techniques for Heart Disease Prediction, International Journal Of Scientific & Technology Research Volume 8, Issue 11, November 2019.

[7] Stephen F. Weng, Jenna Reps, Joe Kai1, Jonathan M. Garibaldi, NadeemQureshi,an machine-learning improve cardiovascular risk prediction using routine clinical data? PLOS ONE https://doi.org/10.1371/journal.pone. 0174944 April 4, 2017.

[8] M. Sultana, A. Haider, and M. S. Uddin, "Analysis of data mining techniques for heart disease prediction," 2016 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol. ICEEICT 2016, 2017

[9] Ghadge, P., Girme, V., Kokane, K. and Deshmukh, P. (2016) Intelligent Heart Attack Prediction System Using Big Data. International Journal of Recent Research in Mathematics Computer Science and Information Technology, 2, 73-77.

[10] AH Chen, SY Huang, PS Hong, CH Cheng, and EJ Lin,2011, "HDPS: Heart Disease Prediction System", Computing in Cardiology, ISSN: 0276-6574, pp.557- 560.

[11] Harish Rajora, Narinder Singh Punn, Sanjay Kumar Sonbhadra, and SonaliAgarwal, Web based disease prediction and recommender system, Indian Institute of Information Technology Allahabad, India.

[12] A Chakravarthy, BS Panda, SK Nayak, .Review and Comparison for Alzheimer's Disease Detection with Machine Learning Techniques, International Neurourology Journal, 27 , 4, pp. 403-406, 2023.