A Machine Learning-Based Predictor of Cardiovascular Disease

RAJESH RAAVI[1], DODDA NAVEEN[2], N. REVANTH KUMAR[3], SAPNA R[4]
[1, 2, 3] *Department of Computer Science and Engineering, Presidency University, Bangalore, India.*
[4] *Assistant Professor, Department of Computer Science and Engineering, Presidency University, Bangalore, India.*

Abstract- ***Heart disease is the biggest cause of death worldwide. Various forms of cutting-edge technologies are used to treat heart conditions. The fact that many medical staff members lack the skills necessary to treat patients successfully is the most common issue in healthcare facilities. They consequently form their own opinions, which frequently result in disastrous effects. Early illness identification is essential since the prevalence of heart disease is rising at an alarming rate. The study's main goal is to identify which people, depending on a range of medical indications, are most prone to acquire heart disease. To predict and identify persons with heart disease, we used a variety of techniques, such as logistic regression, random forests, and k-nearest neighbour (KNN) algorithms. The suggested method can predict a person's risk of acquiring cardiovascular disease with accuracy. This technique for predicting heart illness enhances patient care, facilitates the diagnosis of the condition, and permits the contemporaneous exploration of massive amounts of data.***

Indexed Terms- ***K-nearest neighbour (KNN), Logistic Regression (LR), Cardiovascular Disease (CVD), Heart Disease Prediction System (HDPS)***

I. INTRODUCTION

Machine learning (ML), a division of computation-based intelligence, has made learning-based algorithms a reality for computers to create and implement. To predict and assess the correctness of the dataset, we employed a number of supervised and unsupervised learning techniques.

The term "cardiovascular diseases" refers to conditions that can affect both the heart and the circulatory system. Heart disease, which have been prevalent for a while, is among the deadliest diseases that are currently recognised. Cardiovascular disease (CVDs) is the leading cause of death globally, according to the WHO. Twenty one percent of those passing away happen before the age of seventy, and cardiovascular-related events account for almost four out of every five fatalities [1].An obstruction in the blood supply to either the brain or the heart results in a heart attack, which can have dire consequences. Stress, hypertension, hyperglycemia, and excessive cholesterol are risk factors for heart disease. With basic medical equipment, each of these factors may be quickly and easily measured at home.

Cardiomyopathy, cardiovascular disease, and coronary heart disease are the three subtypes of heart disease. The term "heart disease" refers to a wide range of ailments that impact the heart, blood arteries, and the movement of fluid throughout and out of the blood vessels. One of the primary causes of demise, disability, and a variety of other ailments is cardiovascular disease (CVD). Making a diagnosis of an illness is a crucial and challenging task in medicine.

II. RELATED WORK

The foundation for this work was a substantial body of research on the use of machine learning techniques to detect CVD. Numerous precise predictions of cardiac illness have been generated using ML systems. Using both traditional and cutting-edge machine learning approaches as well as deep learning methodologies, the Innovative Heart Disease Prediction System (IHDPS) model was able to reasonably predict a person's likelihood of getting

heart disease [2]. The primary focus was on fundamental parameters including age, sex, blood pressure, pulse, and glucose level. In contrast, modern models are more trustworthy, accurate, and efficient since they make use of neural networks and deep learning..According to test results, a unique neural network algorithm has an 81.8% categorization power for properly identifying the absence of coronary artery disease patients and a category power of 77% for identifying people with a history of coronary artery disease. Heart attacks are caused by a blockage in the blood flow to the brain or circulatory system, which can have catastrophic results. Heart disease is a risk factor for stress, high blood pressure, high cholesterol, and hyperglycemia. Each of these parameters may be rapidly and readily measured at home with simple medical equipment. The three distinct kinds of heart disease include coronary heart disease, cardiomyopathy, and cardiovascular disease. The phrase "heart illness" is used to refer to a wide variety of illnesses that impact the heart, blood arteries, and the movement of fluid throughout and out of the circulatory. Cardiovascular disease (CVD) is one of the leading causes of death, disability, and many kinds of other illnesses. In medicine, determining the cause of a disease is an important and difficult undertaking.

III. PROPOSED METHOD

The main goal of the suggested approach is to recognise the onset of cardiovascular disease in order to rapidly and precisely identify the issue. To forecast the development of heart disease, our system makes use of many health-related variables, data mining, and machine learning techniques like Logistic Regression, Naive Bayes, K Nearest Neighbour (KNN), Decision Tree, Artificial Neural Networks (ANN), and Random Forest. Five data mining techniques were used to do study on our coronary disease data collection. We use a variety of techniques to identify the classification algorithm that best predicts heart disease. There is the highest degree of accuracy available. Five tests were conducted out to compare the outputs of the Logistic Regression KNN, Neural Networks, Decision Trees, and Naive Bayes techniques.

We put five data mining methods to the test using our collection of information on heart disease. We investigate which categorization approach has the best level of accuracy and is most beneficial for detecting heart disease using a number of algorithms. We trained our model based on the data set collected from Kaggle [5]

IV. RESULT

The results of the experiment, as shown in Figure1, the accuracy of Logistic Regression is the highest (88%), followed by that of ANN and KNN, the two of which are 87%. Compared to Logistic Regression, KNN is slower. KNN can just produce labels, however Logistic Regression can evaluate the level of certainty in its prediction. The accuracy-focused Random Forest Classifier (RFC) performs best while properly fitted; on the other hand, it quickly develops weight. RFC networks of choice trees can employ random selections on each tree to precisely record more complex feature patterns with the highest degree of accuracy.

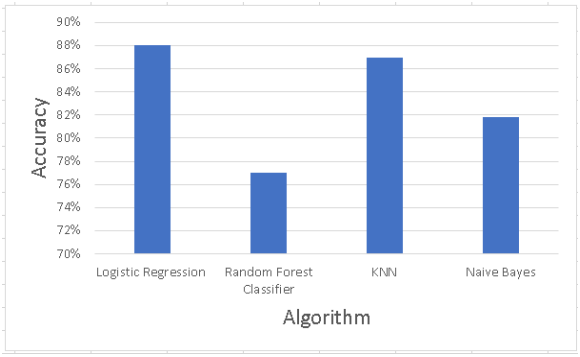

Figure.1

Using the worth of characteristics and tree graphs, RFC can also tell us the quantity that each feature influences classification prediction for simpler interpretation. Our results show how data mining may be applied to the healthcare sector to identify and forecast disease at its earliest stages.

CONCLUSION

The application of data mining technologies in medicine, specifically in the detection of heart disease, was the primary focus of our work. A cardiac condition might be fatal. Data mining methods were applied using algorithms such as Logistic Regression, KNN, Neural Networks, Decision Tree, Naive Bayes,

and Random Forest. The following rate measures were utilised to evaluate performance: accuracy, TN, FP, FN, and TP. Five forecasts of heart disease were made using the same set of data. For easier comparison and interpretation, the output for each developed method is supplied in tabular form.

FUTURE SCOPE

To increase classification precision Further research is required when implementing complex approaches, including bagging, machine learning support, table selection, etc. Apply the proposed system to the relevant area after evaluating the accuracy of the predictions made by each algorithm. In order to increase the implementation accuracy of the method, we may add additional characteristics as necessary.

It should be used by stakeholders as a specific tool to aid in decision-making. No parameter modifications were made in our implementation. The settings of the experiment can be altered to improve and alter subsequent runs.

With more information on heart disease and other data reduction techniques, future work will be possible.

The use of suitable, high-quality datasets can enhance cardiac disease prediction models.

REFERENCES

[1] Palaniappan, S., & Awang, R. (2008). Intelligent heart disease prediction system using data mining techniques. 2008 IEEE/ACS International Conference on Computer

[2] Dutta, A., Batabyal, T., Basu, M., & Acton, S. T. (2020). An efficient convolutional neural network for coronary heart disease prediction. Expert Systems with Applications, 159, 113408. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] Komal Kumar Napa, G.Sarika Sindhu, D.Krishna Prashanthi, A.Shaeen Sulthana, “Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifiers”, https://www.researchgate.net/publication /340885231_Analysis_and_Prediction_of _Cardio_Vascular_Disease_using_Machine_ Learning_Classifiers, April 2020.

[4] Muhammad Usama Riaz, SHAHID MEHMOOD AWAN, ABDUL GHAFFAR KHAN, “PREDICTION OF HEART DISEASE USING ARTIFICIAL NEURAL NETWORK”, October 2018

[5] DataSet - https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset