JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 10, ISSUE 07, 2023

MACHINE LEARNING FOR REAL-TIME HEART DISEASE PREDICTION

Dr. SUBBA REDDY BORRA [1], T.VAIBHAVI[2], T .SREEJA[3], V.GAYATHRI[4], SITALA SRIVALLI[5]

[1]Professor and Head, Department of Information Technology, Malla Reddy Engineering College for Women (UGC-Autonomous), Maisammaguda, Hyderabad, TS, India.

[2,3,4,5] UG Students, Department of Information Technology, Malla Reddy Engineering College for Women (UGC-Autonomous), Maisammaguda, Hyderabad, TS, India.

ABSTRACT:

In this paper, we used data analytics to perform research on heart disease. Since more data is becoming available, the subject of heart disease prediction is still relatively new. Several researchers have examined it using different methodologies and techniques. We used data analytics to locate and predict the sufferers of illnesses. We initially carried out a pre-processing step in which we identified which attributes were the most important based on the correlation matrix using three data analytics methodologies (Decision tree, Random forest, SVM, and KNN) on data sets of different sizes. This enabled us to assess the precision and stability of each method. The datasets are categorised using medical parameters. To examine such factors, our system employs a data mining classification technique. The datasets are analysed in Python using machine learning methods, with the best model demonstrating the highest level of accuracy for heart disease.

Keywords: KNN, SVM, Random forest, Decision tree.

I INTRODUCTION

Heart attacks are the leading cause of death among all deadly disorders. Medical professionals carry out many surveys on heart illnesses to learn more about heart patients, their symptoms, and the development of their condition. Heart attacks that can be fatal are commonplace today. Some signs offered a hint as to what was to come. Medical science has made excellent use of technological breakthroughs to raise the standard of healthcare. Accurate illness diagnosis and prognosis are now achievable because of technological breakthroughs. Machine learning might be a great option for you if you want to accurately anticipate cardiac illnesses. As a result, three algorithms will be employed. It is a technique that combines a random forest and a decision tree for logistic regression. Additionally, these three methods deliver results much more frequently and reliably. Predictions are becoming simpler to comprehend as a result of technological advancements. People today strive arduously to become affluent and famous while living luxurious lives all around the world. People neglect to take care of their health as a result of their hectic schedules. They now live in different ways and consume different meals as a result. Blood pressure, diabetes, and a

variety of other disorders are more prone to develop in young people who are under strain and stress in their lives. Any one of these causes can lead to the development of heart disease.

The heart disease dataset from the UCI repository was subjected to a number of classification and clustering approaches in this study using Python. The major objective is to target every possible combination of the traits against various algorithms. The optimal technique among all of the procedures is then determined to predict heart disease at an early stage. If the three algorithms—Decision tree, Random forest, and Logistic Regression—were used, it would be easier to identify and categorise the disease. A dataset is utilised to categorise and train the model. After the model was trained, the disease was predicted using the most precise and efficient method.

Problem Statement

Blood pressure, total cholesterol, LDL cholesterol, and HDL cholesterol were all found to be strongly linked with the condition. In 12 years, 383 men and 227 women were projected to have CHD. The accuracy of this categorical technique was shown to be comparable to CHD prediction when the continuous variables themselves were used. Since heart disease cannot be predicted with better accuracy and learning rate in its earlier stages, the current algorithms can only predict it with an accuracy of 93%.

II LITERATURE SURVEY

Numerous research have honestly said the increase of heart sickness scientific prognosis based totally upon tool getting to know variations with the goal of giving an HDPM with boosted performance. 2 overtly without issues to be had coronary heart problem datasets, especially Statlog and additionally Cleveland, have definitely been significantly carried out to assessment the performance of prediction versions amongst researchers. For Statlog dataset, a coronary heart trouble medical choice assist device primarily based upon mayhem firefly technique in addition to tough units-based high-quality reduction (CFARS-AR) have become created via Long et al. (2015). Through units had been implemented to reduce the form of abilities even as the sickness firefly components have end up implemented to categorize the infection. The installed version have turn out to be after that during comparison with severa extraordinary variations alongside NB, SVM and additionally ANN.

The aggregate of harsh devices-primarily based completely trends preference and BPNN (RS-BPNN) emerge as recommended via Nahato et al. (2015). With the selected characteristics, the counseled RS-BPNN performed an accuracy of about ninety.Four%. Dwivedi (2018) contrasted 6 artificial intelligence variations (ANN, SVM, LR, precise enough-nearest next-door neighbor (kNN), classification tree and additionally NB) with numerous performance metrics. The consequences determined that LR finished higher than the several distinctive fashions via challenge, about 85%, 89%, 81%, similarly to 80 five for the precision, stage of sensitivity, area of know-how, and furthermore, accuracy, especially.

Amin et al. (2019) accomplished comparison evaluation via way of spotting great traits and moreover, using synthetic intelligence designs (okay-NN, DT, NB, LR, SVM, Semantic Network (NN) in addition to a hybrid (poll with NB and LR)). The test results disclosed that the crossbreed version (ballot with NB and LR) with picked characteristics attained the satisfactory precision (87.Forty one %). The Cleveland cardiovascular sickness dataset has actually been typically made use of by scientists to deliver predictive designs.

Verma et al. (2016) created a hybrid prediction layout based upon correlation characteristic component (CFS), fragment swam optimization (PSO), K-method clustering and moreover, MLP. The consequence was that the recommended crossbreed model attained precision of as tons as 90.28%.

Haq et al. (2018) [6] done a comparative have a examine on a hybrid model based totally on several feature alternative techniques (remedy, minimal-redundancy maximal-relevance (mRMR), the very least outright shrinkage and alternative operator (LASSO)) in addition to artificial intelligence fashions (LR, kNN, ANN, SVM, DT, NB, and furthermore RF). Their studies look at the observe that the abilities decrease affects the general performance of the fashions. The research ended that a mix of Relief-based totally definitely truly feature opportunity further to LR-primarily based on complete device discovering set of regulations (MLA) offers higher accuracy (as much as 89%) in evaluation with different combos made use of within the studies.

Saqlain et al. (2019) advised a manner based on implied Fisher rating characteristic desire gadgets (MFSFSA) in addition to SVM elegance format. The chosen functions are based definitely upon the better Fisher score rather than the recommended score. After that, SVM made use of the chosen feature part to have a have a look at and calculate the MCC through a reputation method. The studies look at determined that the combination of FSFSA in addition to SVM generates precision, stage of sensitivity, in addition to specificity of as lots as eighty one.19%, seventy .Ninety %, similarly to 88.Sixty eight%, particular.

Latha and Jeeva (2019) proposed a crossbreed format with a majority poll of NB, BN, RF, as well as MLP. The advised version attained a precision of as an entire lot as eighty five. Forty 8%. Ali et al. (2019) [5] encouraged piled SVMs to beautify the clinical evaluation approach.

The initial SVM was implemented to put off the non-applicable attributes and furthermore, the second one to are looking forward to coronary coronary heart problem. The consequences found out that the advocated format attained far better trendy ordinary performance than exquisite versions and additionally, previous studies consequences. Mohan et al. (2019) supplied a crossbreed RF with a right away model (HRFLM) to beautify the performance of the HDPM. They positioned that the advised technique completed accuracy, accuracy, stage of sensitivity, f-diploma further to strong point of as a good deal as

88.Four%, ninety.1%, ninety .Eight%, ninety%, and 82.6%, mainly.

Recently, Gupta et al. (2020) set up a device information shape containing variable evaluation of mixed facts (FAMD) and RF-based totally completely MLA. The FAMD became made use of to find out the applicable talents and the RF to assume the state of affairs. The speculative effects showed that the advocated approach outshines other variations further to preceding research consequences thru carrying out the precision, degree of sensitivity, and specificity of as a lot as ninety 3.Forty four%, 89.28%, and furthermore, 96.Ninety six%, respectively.

EXISTING SYSTEM

Cardiovascular diseases (CVDs), which are heart-related infections, have become the most dangerous illness worldwide and in India during the past few years. They are the main cause of a staggering number of deaths worldwide. Therefore, a trustworthy, accurate, and attainable framework is required to analyse these diseases in time for appropriate treatment. To automate the analysis of vast and complex data, AI algorithms and methods have been used for many clinical datasets. Recently, many scientists have started using a few AI techniques to assist the healthcare sector and the experts in the investigation of heart-related ailments.

PROPOSED SYSTEM

The leading cause of death worldwide is heart disease. Medical professionals find it challenging to predict because doing so needs higher levels of predictive knowledge and skill. Although there is a knowledge gap, the healthcare sector is an information-rich environment. Healthcare systems have a lot of data online, but there is a shortage of efficient analysis tools to find hidden patterns in the data. An automated method will improve medical effectiveness while cutting costs and waiting times. Using data acquired from Kaggle, this software seeks to forecast the occurrence of a disease. By using data mining techniques on the dataset, the goal is to uncover hidden patterns and estimate the presence value on a scale. The amount of data needed to forecast cardiac disease is enormous and too difficult to analyse and analyse using traditional methods. Our goal is to identify a good method for heart disease prediction that is accurate and effective.

METHODOLOGY

Data Source

The UCI Machine Learning repository was used to obtain the dataset used in this analysis to forecast cardiac disease. The databases that make up UCI are utilised to implement machine learning algorithms. This dataset is an actual dataset. 300 data instances with the proper 14 clinical parameters make up the dataset. The clinical parameter of the dataset pertains to tests that are done for heart illness, such as blood pressure level, type of chest pain, electrocardiogram result, and others.

Description of Algorithms

In this section describe two main algorithms are used in this system namely

i) Decision tree Classification algorithm
ii) SVM Algorithm.
iii) K neighbors algorithm
iv) Random forest.

Finding out if the patient has cardiac disease is the goal of this endeavour.

The datasets' records are split into training and test sets. Data mining classification techniques such as decision trees and naive Bayes were used once the data had been preprocessed. The outcomes of the categorization models created using Python programming are displayed in this section. Both training datasets and test data sets are used to obtain the findings.

RESULTS EXPLANATION

In our project, we've developed a user interface (UI) with a database where users can sign up and log in to forecast their heart illness. If users don't want to sign up and log in, they may still predict their heart disease by clicking the rapidly predict button. The user must next enter the 13 attributes, such as their age, gender, cholesterol level, etc. Additionally, after entering the information, a model will be built based on the user's details, and the dataset will be trained. Test and Train are split into 25% and 75%, respectively. If the percentage is greater than 60%, then there is a risk of heart disease, and the user can access prevention and symptoms options on our user interface (UI) by selecting the build model and heart disease risk percentage. If a user logs in using the same login information as when they predicted earlier, they can access their prior records on our website. The user has the choice to get in touch with the user directly by phone or email if the model anticipated heart illness. The database contains all the user data, which is available whenever needed.

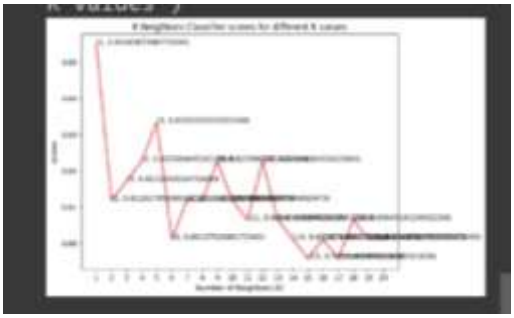

Fig.1. K neighbors algorithm.

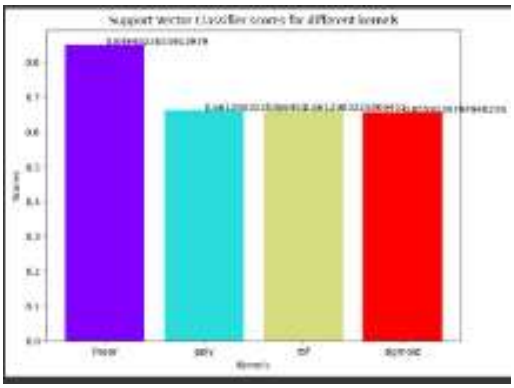

Fig.2. SVM classifier.

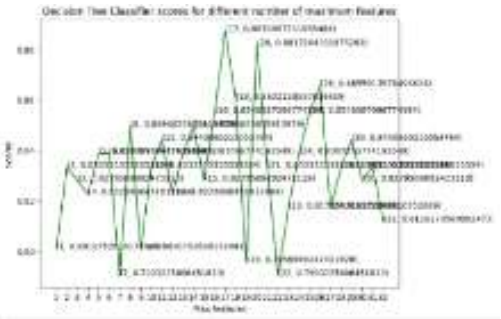

Fig.3. Decision tree algorithm.

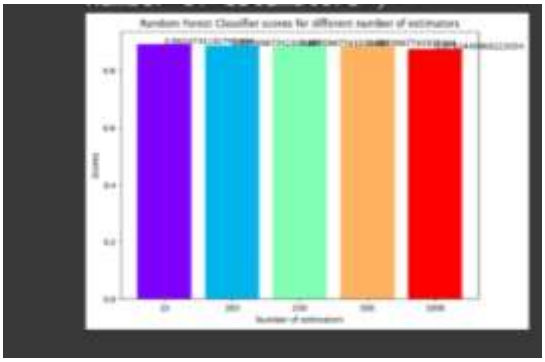

Fig.4. Random forest classifier.

Python is used to apply multiple classification and clustering techniques to the heart disease dataset from the UCI repository. Targeting every potential combination of the qualities against different algorithms is the main goal. The technique that works the best to

forecast heart disease at an early stage is then identified out of all the techniques.

CONCLUSION

Building a system that can effectively and reliably forecast heart disease is becoming more crucial as the number of deaths from heart disease rises. Finding the best effective ML algorithm for heart disease detection was the study's driving force. The dataset from the UCI machine learning repository is used in this work to examine the accuracy scores of K Neighbours, Logistic Regression, and Random Forest for predicting heart disease. The findings of this study show that the most effective algorithm for predicting heart disease, with an accuracy score of 89%, is the logistic regression algorithm. The dataset utilised for both training and testing purposes affects how accurate machine learning algorithms are. Other machine learning methods can be used to forecast cardiac disease. Additionally, binary classification issues like the prediction of heart disease can be successfully handled using logistic regression. Decision trees may not perform as well as random forests. The data set can also be used using ensemble methods and artificial neural networks. It is possible to compare and improve the outcomes.

REFERANCES

[1]. Rairikar, A., Kulkarni, V., Sabale, V., Kale, H., & Lamgunde, A. (2017, June). Heart disease prediction using data mining techniques. In 2017 International Conference on Intelligent Computing and Control (I2C2) (pp. 1-8). IEEE.

[2]. Gandhi, Monika, and Shailendra Narayan Singh. "Predictions in heart disease using techniques of data mining." In 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp. 520-525. IEEE, 2015.

[3]. Chen, A. H., Huang, S. Y., Hong, P. S., Cheng, C. H., & Lin, E. J. (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557-560). IEEE.

[4]. Aldallal, A., & Al-Moosa, A. A. A. (2018, September). Using Data Mining Techniques to Predict Diabetes and Heart Diseases. In 2018 4th International Conference on Frontiers of Signal Processing (ICFSP) (pp. 150-154). IEEE.

[5]. Sultana, Marjia, Afrin Haider, and Mohammad Shorif Uddin. "Analysis of data mining techniques for heart disease prediction." In 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pp. 1-5. IEEE, 2016.

[6].Al Essa, Ali Radhi, and Christian Bach. "Data Mining and Warehousing." American Society for Engineering Education (ASEE Zone 1) Journal (2014).

[7].Shetty, Deeraj, Kishor Rit, Sohail Shaikh, and Nikita Patil. "Diabetes disease prediction using data mining." In 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), pp. 1-5. IEEE, 2017.

[8]. Methaila, Aditya, Prince Kansal, Himanshu Arya, and Pankaj Kumar. "Early heart disease prediction using data mining techniques." Computer

Science & Information Technology Journal (2014): 53-59.

[9]. Dewan, Ankita, and Meghna Sharma. "Prediction of heart disease using a hybrid technique in data mining classification." In 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 704-706. IEEE, 2015.