International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-12, Special Issue-1, March-2024
https://doi.org/10.55524/CSISTW.2024.12.1.29
Computing for Sustainable Innovation: Shaping Tomorrow's World (CSISTW-2024) on 16th-17th February, 2024
Organized by B.N. College of Engineering and Technology, Lucknow, India

Heart Disease Prediction model using Machine Learning

Ashish Yadav [1], and Dr. Ram Niwas [2]

[1] M.Tech, Student Department Of Computer Science And Engineering, Niec Agra, U.P,
[2]Assistant Professor of Computer Science And Engineering, Niec Agra, U.P,

Correspondence should be addressed to Ashish Yadav: ashish854380@gmail.com

ABSTRACT- We develop a heart disease prediction model that can help medical professionals to predict the heart disease status of patients based on their daily data. Data analysis and data modelling are becoming very useful in the world of medical science. Today I am going to tell this in my research paper my research paper is heart related disease, which plays a very important role in the human body. The work of the heart is to pump pure blood to various parts of the body and provide flow of pure blood. If any part of the body stops working, it means that blood does not reach there. In this research paper, I will consider starting from problem statement till data modelling in which I will consider many processes like data collection, data pre-processing, data splitting into two part like training and testing, data model evaluation and use of KNN. By doing this we will create a data model whose accuracy will be 91.4%. We will create it using logistic regression and will also deploy it which will be a user friendly console in which when the user gives some input data, an output will come which will tell whether the user has heart diseases or not. In which we have taken 14 clinical features such as age, sex, Activity, Alcoholic, BMI etc., after that we have created an algorithm using data pre-processing and some techniques.

KEYWORDS-- Data analysis, Data modelling, KNN, Data pre-processing, Machine Learning.

I. INTRODUCTION

from heart disease. By using machine learning algorithm, we can detect the beginning of heart disease. Which kills millions of people? Heart has been playing a valuable role in the human body.. It is very safe in the left side of the body inside the lungs. Data analysis is becoming useful for taking decisions using business intelligence and it is playing a very important role in medical sciences. I am also exploring heart disease using business intelligence.[4] In this research paper, we develop a heart disease prediction system that can help medical professionals to predict the heart disease status of patients based on their daily data. In which we have taken 13 clinical features such as age, sex, Activity, Alcoholic, BMI etc, after that we have created an algorithm using data pre-processing and some techniques. There are many instruments available to predict heart disease but the big problem is that one is very costly and the other is the accurate result, currently according to WHO report the accuracy of prediction is only 67%.

II. LITERATURE REVIEW

a) A data mining model for predicting the coronary heart disease using random forest classifier Authors:

A. S. Abdullah and R. R. Rajalaxmi. Abstract:[1] Heart diseases are a very big health problem. Since accurate prediction of a person's disease condition is very important, medical science can be simplified by using such intelligent tools for disease diagnosis and treatment to reduce doctors' mistakes and financial losses. In this paper, Particle Swarm Optimization (PSO) algorithm, which is considered one of the most important and powerful evolutionary algorithms, is used to follow the rules for heart disease. First the random rules are stored and then they are followed based on the accuracy of the diagnosis with the help of PSO algorithm. Finally we compare our results to the evolution algorithm to see how good the result is.

b) Backpropagation Neural Network for Cardiovascular Disease Prediction Author:

N.Al-Milli, Abstract:[2] Recently, many software, tools and many algorithms have been applied by researchers to develop more and more effective medical decision support systems . Despite all this, new algorithms and new tools are being developed and introduced day by day. Diagnosis of cardiovascular disease is a very big issue and has been investigated by many researchers to improve the ability of health practitioners to develop good medical decision support systems. Neural networks are being used experimentally to predict the treatment of heart disease. This is where I glean as much knowledge as I can from examples or raw data on heart disease. Here, a rule-based clinical decision support system (CDSS) for cardiovascular disease prevention, which automatically derives information from the patient's daily data, is presented. The clinical life support system for problem prediction of cardiovascular patients involves two steps, (1) automated approach for generation of weighted fuzzy rules and decision tree rules, and, (2) developing a fuzzy rule-based decision support system. To do . In the first step, we use weighted fuzzy rules mining technique, feature selection and feature weighting method to obtain this rule.[5] Then, the fuzzy system is constructed according to the weighted fuzzy rules and some features chosen. Finally, we conduct experiments on the proposed system using datasets obtained from the UCI repository to further compare the performance of the system to neural network-based systems using accuracy, sensitivity and specificity.

III. PROPOSED SYSTEM

This project presents various properties related to heart disease, and Models based on supervised learning Logistic Regression, Decision Algorithms tree, random forest, support vector Machine, Gaussian Naive Bayes, Multinomial Naive Bayes and gradient Boosting classifier algorithm. uses it Existing dataset from Cleveland Database of UCI reserves in heart disease patients. The dataset contains 31000 examples and 74 Property. [3] Only 14 of these 74 characteristics The characteristics considered for testing are, It is important to certify the performance of Various algorithms. this research paper to imagine the possibility of Development of heart disease in patients.

IV. SYSTEM DESIGN

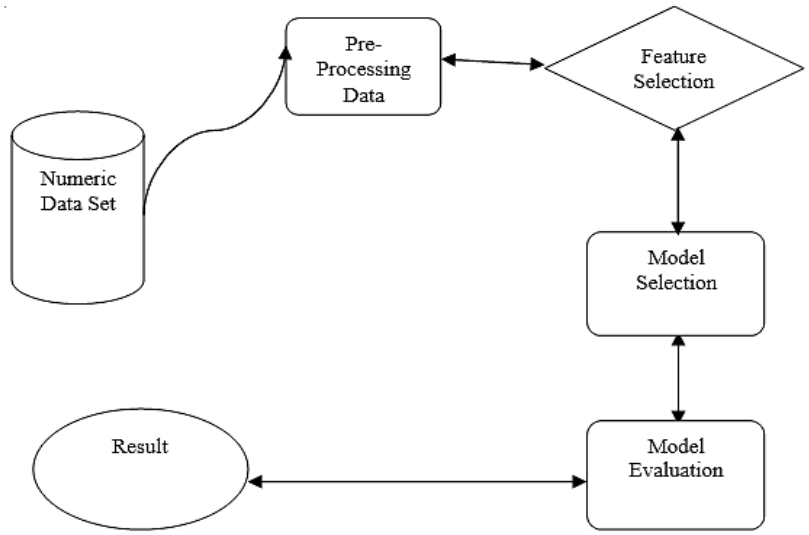

Figure 1: Architecture of system

V. MODULE DESCRIPTION

To develop an effective heart disease prediction model using data science techniques machine learning, I am following the module details given below. The description here provides an overview of the different components and many steps involved in the process.

a) Data collection:

Gathered a comprehensive dataset that included relevant characteristics related to heart disease, such as age, gender, blood pressure, cholesterol levels, family history, smoking habits, and medical test results (e.g., ECG readings) several daily routines to ensure that the dataset is representative and includes a sufficient number of positive and negative heart disease examples.

b) Data Pre-processing:

Some steps are followed in data pre-processing such as data cleaning, preparing the dataset for tanning the data so that a machine learning model can be prepared in which missing values are removed, normalization and standardizing and dividing it into categorical variables.

c) Feature selection:

Analyzed the dataset to identify the most informative features for predicting cardiovascular disease. Consider using techniques such as correlation analysis, feature importance ranking, or dimensionality reduction methods (for example, principal component analysis) to select the subset of features with the highest predictive power that will lead to accurate heart disease prediction.

d) Model Selection:

After exploring various machine learning algorithms suitable for classification tasks I selected models, specifically for predicting heart disease. Commonly used algorithms include logistic regression, decision trees, random forests, support vector machines, naive Bayes, and k-nearest neighbors. The performance of these models can be evaluated and compared using appropriate evaluation metrics and techniques such as cross-validation.

e) Model training:

Divided the pre-processed dataset into two parts, training and validation sets. Using the training set to train selected machine learning models on the data. During the training process, the pattern of the model is determined by its features and the target variable (heart disease or not).

f) Model evaluation:

I evaluated the performance of the trained model using the validation set. Used some techniques to measure the predictive abilities of the model and calculated evaluation metrics such as (accuracy, precision, recall, F1 score and area under the ROC curve). Compare the performance ith the different models and select the model with the highest accuracy and desirable trade-offs in terms of other evaluation metrics.

g) Hyper parameter Tuning:

Optimize Hyper parameters of a model chosen to enhance its performance. To explore different combinations of hyper parameter values and find the optimal configuration I used some techniques such as grid search, random search, or built the model using Bayesian optimization.

h) Model Validation:

I validated the final trained model on an independent test set to get an unbiased estimate of its performance. Which helps us assess how well the model generalizes to unseen data and helps provide confidence in its predictive ability.

i) Model Deployment

Once the model is built, validated and deemed effective, we need a user interface to deploy it to predict heart disease in real-world scenarios. The model can be run on a software application, web service, or platform integrated into a healthcare system to make predictions on the data.

j) Conclusion

The knowledge of predicted outcome of heart disease data will help in long term saving of human life and early detection of abnormalities in heart conditions so that the user can know whether he has heart related disease or not. We have used algorithms using machine learning to process the raw data and provide a new and innovative understanding towards heart disease. Prediction of cardiovascular disease is becoming challenging and very important in the medical field. However, if we detect any disease in its early stages and adopt preventive measures as soon as possible, then the mortality rate can be controlled to a great extent. The overall objective is to define various data mining techniques useful in effective cardiovascular disease prediction. I aim to make efficient and accurate predictions with a small number of features and tests. In the study here, I have considered only 14 essential characteristics. I applied four data mining classification techniques, which are Logistic Regression, Decision Tree Random Forest Support Vector Machine Gaussian Naive Bayes Multinomial Naive Bayes and gradient boosting classifier. The data is pre-processed and then used in the model. Logistic Regression, Gaussian Naive Bayes and Decision Tree algorithms are showing the best results in the model.[6] After I implemented these algorithms I got the highest accuracy in KNN. We can also further expand this research by incorporating other data mining techniques such as time series, clustering and association rules, support vector machines, and genetic algorithms. Considering the limitations of this study, there is a need to apply more complex and better combinations of models to achieve higher accuracy for early prediction of cardiovascular disease.

VI. CONCLUSION

In this research endeavour, we embarked on a comprehensive exploration of predictive modeling for early detection of cardiovascular diseases using machine learning. The study integrated electronic health records (EHRs), demographic details, and lifestyle factors to develop and evaluate various machine learning algorithms, including deep learning models (CNN, RNN), support vector machines (SVM), and ensemble methods (Random Forest, Gradient Boosting).

CONFLICT OF INTERREST

The authors declare that they have no conflict of interest.

REFERENCES

[1] N. Al-milli, "Backpropogation neural network for prediction of heart disease," J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131-135, 2013.

[2] 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. (HSI), Gdansk, Poland, Jul. 2018, pp. 233- 239.

[3] P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules," J. King Saud Univ -Comput. Int Sci., vol. 24, no. 1. PP 27-40. Jan 10.1016/j.jksuci. 2011.09.002. dor 2012

[4] 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.cswa 2018.01.025

[5] H. A. Esfahani and M. Ghazanfari, "Cardiovascular disease detection using a new ensemble classifier," in Proc. IEEE 4th Int. Conf. Knowl Based Eng Innov. (KBEI), Dec. 2017, pp. 1011-1014. [6] F. Dammak, L. Baccour, and A. M. Alimi, "The impact of criterion weights techniques in TOPSIS method of multi-criteria decision making in crisp and intuitionistic fuzzy domains," in Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE), vol. 9. Aug. 2015. pp. 1-8

[6] R. Das, L. Turkoglu, and A. Sengur, "Effective diagnosis of of heart disease through neural networks ensembles," Expert Syst. Appl., vol. 36, no. 4, pp. 7675- 7680, May 2009 10.1016/j.eswa.2008.09.013. doi

[7] M Durairaj and V. Revathi, Prediction of heart disease using back propagation MLP algorithm Int J Sev Technol. Res vol. 4, no. X. pp. 235 239, 2015

[8] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, "Prediction of heart disease using machine learning," in Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA), Mar. 2018, pp. 1275- 1278.

[9] B. S. S. Rathnayake and G. U. Ganegoda, "Heart diseases prediction with data mining and neural network techniques," in Proc. 3rd Int. Conf Converg. Technol. (12CT). Apr. 2018. pp. 1-6.

[10] N. K. S. Banu and S. Swamy. "Prediction of heart disease at early stage using data mining and big data analytics: A survey, in Proc. Int. Conf, Flect, Electron, Commun. Comput Optim Techn (ICEECCOT), Dec. 2016. pp. 256-261