Heart Disease Prediction Using Machine Learning

[1]Shreya Pattankude, [2]Aakanksha Vetekar, [3]Prof.Sirasappa Y Pattar

[1]M.Tech, [2]M.Tech, [3]M.Tech (PhD)
[1] Department of Medical Electronics,
[1]BMS College of Engineering, Bengaluru, India

***Abstract*—Day by day the cases of heart diseases are increasing at a rapid rate and it's very Important and concerning to predict any such diseases beforehand. This diagnosis is a difficult task i.e. it should be performed precisely and efficiently. The research paper mainly focuses on which patients is more likely to have a heart disease based on various medical attributes. We used machine learning algorithm such as logistic regression and to predict and classify the patient with heart disease. A quite Helpful approach was used to regulate how the model can be used to improve the accuracy of prediction of Heart Attack in any individual.**

Index Terms*— *Heart diseases, Machine learning, Accuracy, Logistic Regression, Prediction

I. INTRODUCTION

The Given heart disease prediction system enhances medical care and reduces the cost. This project gives us significant knowledge that can help us predict the patients with heart disease It is implemented on the .pynb format.Machine Learning is a very vast and diverse field and its scope and implementation is increasing day by day. Cardiovascular diseases are very common these days, they describe a range of conditions that could affect your heart. Day by day the cases of heart diseases are increasing at a rapid rate and it's very important and concerning to predict any such diseases beforehand. This project focuses on mainly data mining techniques like Logistic regression. The accuracy of our project is 85% for which is better than previous system where only one data mining technique is used. Logistic regression falls under the category of supervised learning. Only discrete values are used in logistic regression. By using this dataset, we predict whether the patient can have a heart disease or not. To predict this, we use 13 medical attributes of a patient and classify him if the patient is likely to have a heart disease. These medical attributes are trained under the algorithms: Logistic regression.

The objective of this project is to check whether the patient is likely to be diagnosed with any cardiovascular heart diseases based on their medical attributes such as gender, age, chest pain, fasting sugar level, etc. By using this dataset, we predict whether the patient can have a heart disease or not. To predict this, we use 14 medical attributes of a patient and classify him if the patient is likely to have a heart disease. These medical attributes are trained under three algorithms: Logistic regression.

II. METHODOLOGY

This paper shows the analysis of various machine learning algorithms , the algorithm that are used in this are logistic regression . The proposed methodology includes steps:

Data Collection and Preprocessing: The dataset consists of a total of 1328 attributes using 13 features available in the Kaggle website for the analysis.

Classification: The input dataset is split into 80% of the training dataset and the remaining 20% into the test dataset. Training dataset is the dataset which is used to train a model. Testing dataset is used to check the performance of the trained model. For the algorithms the performance is computed and analyzed based on different metrics used such as accuracy and logistic regression.

Logistic Regression: It is a classification algorithm mostly used for binary classification problems. In logistic regression instead of fitting a straight line or hyper plane, the logistic regression algorithm uses the logistic function to squeeze the output of a linear equation between 0 and 1. There are 13 independent variables which makes logistic regression good for classification.

Logistic regression is a one of the machine learning classification algorithm for analyzing a dataset in which there are one or more independent variables (IVs) that determine an outcome and also categorical dependent variable (DV). Linear regression uses output in continuous numeric whereas logistic regression transforms its output using the logistic sigmoid function to return a probability value which can then be mapped to two or more discrete classes. The logistic regression forms three types as below. a) Binary logistics regression (two possible outcomes in a DV) b) Multinomial logistics regression (three or more categories in DV without ordering) Ordinal logistics regression (three or more categories in DV with ordering).

Table 1: Various attributes used are listed below:

S. No	Observation	Description	Values
1.	Age	Age in Years	Continuous
2.	Sex	Sex of Subject	Male/Female
3.	CP	Chest Pain	Four Types
4.	Trestbps	Resting Blood Pressure	Continuous
5.	Chol	Serum Cholesterol	Continuous
6.	FBS	Fasting Blood Sugar	< ,or > 120 mg/dl
7.	Restecg	Resting Electrocardiograph	Five Values
8.	Thalach	Maximum Heart Rate Achieved	Continuous
9.	Exang	Exercise Induced Angina	Yes/No
10.	Oldpeak	ST Depression when Workout compared to the Amount of Rest Taken	Continuous
11.	Slope	Slope of Peak Exercise ST segment	up/ Flat /Down
12.	Ca	Gives the number of Major Vessels Coloured by Fluoroscopy	0-3
13.	Thal	Defect Type	Reversible/Fixed/Normal
14.	Num(Disorder)	Heart Disease	Not Present /Present in the Four Major types.

Figure1: shows methodology of this project

III. RESULTS

The output will give a prediction result if the person has a heart disease or not .Our project also tells us that Logistic Regression in the prediction of the patient diagnosed with a heart Disease. This proves that Logistic Regression is better in diagnosis of a heart disease. The accuracy on training data is 86% and accuracy on testing data is 83%. Use of more training data

IV. CONCLUSION

This project helps us predict the patients who are diagnosed with heart diseases by cleaning the dataset and applying logistic regression to get an accuracy of an average of 85% on our model . The accuracy on training data is 86% and accuracy on testing data is 83%. Use of more training data ensures the higher chances of the model to accurately predict whether the given person has a heart disease or not ensures the higher chances of the model to accurately predict whether the given person has a heart disease or not. There are a number of medical databases that we can work on as these Machine learning techniques are better and they can predict better than a human being which helps the patient as well as the doctors.

Data Pre-Processing: In order to build up a more accurate ML model, data preprocessing is required. Data pre-process is the process of cleaning the data. This includes identification of missing data, noisy data and inconsistent data. This is followed by classifying the data into training and test data and then applying accuracy and logistic regression algorithm to the data. Confusion matrix is also used to check reliability of the data.

Confusion Matrix Outcomes: This has been used to indicate the summary of prediction results including correct and incorrect on a classification problem. Further, this was used to not only errors but also types of errors. The segments of the confusion matrix indicate the following parameters. True Positives (TP): cases which are predicted yes (they have the disease), and they do have the disease. True Negatives (TN): cases which are predicted no, and they do not have the disease. False Positives (FP): cases which are predicted yes, but they do not actually have the disease (Type I error). False Negatives (FN): cases which are predicted no, but they actually do have the disease (Type II error). The following outcome indicates the confusion matrix of the database.

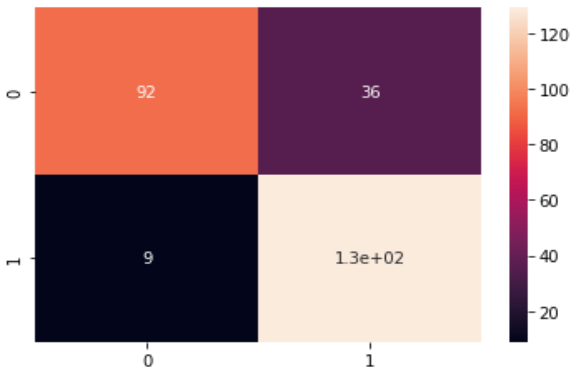

Figure 2: Shows confusion matrix

V. ACKNOWLEDGMENT

We take immense pleasure in thanking Sirasappa Y Pattar, Professor, Department of Medical Electronics, for inspiring us and giving us an opportunity to work on this project.

VI. REFERENCES

[1]]K.Sudhakar, Dr. M. Manimekalai "Study of Heart Disease Prediction using Data Mining", IJARCSSE 2016.
[2] NagannaChetty, Kunwar Singh Vaisla, NagammaPatil, "An Improved Method for Disease Prediction using Fuzzy Approach", ACCE 2015.
[3] VikasChaurasia, Saurabh Pal, "Early Prediction of Heart disease using Data mining Techniques", Caribbean journal of Science and Technology,2013
[4] ShusakuTsumoto," Problems with Mining Medical Data", 0-7695- 0792-1 I00@ 2000 IEEE.
[5] Y. Alp Aslandoganet. al.," Evidence Combination in Medical Data Mining", Proceedings of the international conference on Information Technology: Coding and Computing (ITCC'04) 0-7695-2108-8/04©2004 IEEE.
[6] Carlos Ordonez, "Improving Heart DiseasePrediction Using Constrained Association Rules," Seminar Presentation at University of Tokyo, 2004
[7] 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.
[8] T.Nagamani, S.Logeswari, B.Gomathy," Heart Disease Prediction using Data Mining with Mapreduce Algorithm International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019.
[9] Fahd Saleh Alotaibi," Implementation of Machine Learning Model to Predict Heart Failure Disease", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.
[10] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin,"Design And Implementation Heart Disease Prediction Using Naives Bayesian" International Conference on Trends in Electronics and Information(ICOEI 2019).
[11] Theresa Princy R,J. Thomas,'Human heart Disease Prediction System using Data Mining Techniques', International Conference on Circuit Power and Computing Technologies,Bangalore,2016.
[12] Nagaraj M Lutimath,Chethan C,Basavaraj S Pol.,'Prediction Of Heart Disease using Machine Learning', International journal Of Recent Technology and Engineering,8,(2S10), pp 474-477, 2019.s
[13] UCI, ―Heart Disease Data Set.[Online]. Available (Accessed on May1 2020): https://www.kaggle.com/ronitf/heart-disease-uci.
[14] Sayali Ambekar, Rashmi Phalnikar,"Disease Risk Prediction by Using Convolutional Neural Network",2018 Fourth International Conference on Computing Communication Control and Automation.
[15] C. B. Rjeily, G. Badr, E. Hassani, A. H., and E. Andres, ―Medical Data Mining for Heart Diseases and the Future of Sequential Mining in Medical Field,‖ in Machine Learning Paradigms, 2019, pp. 71–99.
[16] Jafar Alzubi, Anand Nayyar, Akshi Kumar. "Machine Learning from Theory to Algorithms: An Overview", Journal of Physics: Conference Series, 2018
[17] .A. Gavhane, G. Kokkula, I. Pandya and K. Devadkar, "Prediction of Heart Disease Using Machine Learning," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 1275-1278, doi: 10.1109/ICECA.2018.8474922.
[18] Denisse Castaneda; Aibhlin Esparza Mohammad Ghamari; Cinn Isler, Y.: Discrimination of systolic and diastolic dysfunctions using multi-layer perceptron in heart rate variability analysis. Comput. Biol. Med. 76, 113–119 (2016) 32.
[19] Shah, S.J., Katz, D.H., Selvaraj, S., Burke, M.A., Yancy, C.W., Gheorghiade, M., Bonow, R.O., Huang, C.C., Deo, R.C.: Phenomapping for novel classification of heart failure with preserved ejection fraction. Circulation 114 (2014) 33.
[20] Srinivas, K., Rao, G.R., Govardhan, A.: Analysis of attribute association in heart disease using data mining techniques. Int. J. Eng. Res. Appl. 1680–1683 (2012) 34.
[21] Nahar, J., Imam, T., Tickle, K.S., Chen, Y.P.P.: Association rule mining to detect factors which contribute to heart disease in males and females. Expert Syst. Appl. 40(4), 1086–1093 (2013) 35.
[22] Methaila, A., Kansal, P., Arya, H., Kumar, P.: Early heart disease prediction using data mining techniques. Comput. Sci. Inf. Technol. J. 53–59 (2014) 36.
[23] Sudhakar, K., Manimekalai, D.M.: Study of heart disease prediction using data mining. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 4(1) (2014) 37.
[24] Ilayaraja, M., Meyyappan, T.: Efficient data mining method to predict the risk of heart diseases through frequent itemsets. Procedia Comput. Sci. 70, 586–592 (2015) 38.
[25] Subramanian, S., Mohanapriya, S., Nagasandhiyalakshmi, B., Shanmugapriya, N.: Prediction of outbreak heart diseases using text mining. Discovery 1070–1077 (2016) 39.
[26] Yang, G., Ren, Y., Pan, Q., Ning, G., Gong, S., Cai, G., Zhang, Z., Li, L., Yan, J.: A heart failure diagnosis model based on support vector machine. In: 2010 3rd International Conference on Biomedical Engineering and Informatics (BMEI), vol. 3, pp. 1105–1108. IEEE (2010) 40.
[27] Bou Rjeily, C., Badr, G., El Hassani, A.H., Andres, E.: Sequence prediction algorithm for heart failure prediction. In: International Conference e-Health, pp. 109–116 (2017) 41.
[28] Bou Rjeily, C., Badr, G., El Hassani, A.H., Andres, E.: Predicting heart failure class using a sequence prediction algorithm.

In: 2017 International Conference on Advances in Biomedical Engineering, IEEE (2017) 42.
[29] Zaki, M.J.: Spade: an efficient algorithm for mining frequent sequences. Mach. Learn. 42(1), 31–60 (2001)