Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 5- Issue 2, Paper 36 August 2022

Cardiovascular Disease Prediction Using Machine Learning Methodology's

Pooja G T , Suvarna G Kanakaraddi
, Shantala Giraddi ,Ashok K Chikaraddi , Vishwanath.P.Baligar
Computer Science
K.L.E Institute of Technology
Hubli, India

Abstract- The human heart is one of the prominent and complicated organs, which holds the responsibility of pumping and purifying the blood. Due to present busy life cycle, humans get into stress in their daily life, which leads to Heart attack disease. Cardiovascular attack can occur at any age in life. Prediction of cardiovascular attack at the early stage and providing guidance to the patients can save the human life. Artificial intelligence and machine learning industry is a very fast-growing field, which can provide a promising solution in medical industry. To overcome the problem of heart disease, in this work, cardiovascular disease prediction using machine learning is proposed. This research is helpful in medical industry to analyze the patients with heart attack related issues. The proposed system is used to analyze cardiovascular attack disease with different types of supervised machine learning algorithms. UCI Repository Machine Learning dataset is used and in total 7 different models are used and compared in the analysis. The proposed algorithms for analysis are Naive Bayes, Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Neural Network and K-Nearest Neighbor. The algorithm was well trained and experiments were conducted on different models. The results of this proposed algorithm were able to analyze cardiovascular attack more effectively. Logistic Regression of accuracy 82.6% and Random Forest of accuracy 90.7% of success were achieved.

Keywords: Heart disease, Logistic Regression, Supervised Machine Learning, Random Forest, Support Vector Machine.

I] Introduction:

As per the survey of World Health Organisation, millions of people have faced and lost their lives due to heart attack disease. Early prediction helps the medical industry to handle disease in a promising life saving model. The main function of the heart is to pump and purify the blood and supply it to the various part of the human body. Cardiovascular attack disease can occur due to multiple types of medical disorders like uncontrolled stress and long-term diseases like diabetics, blood pressure, paralysis and many more. Due to today's modern world life style and bad habits of food consumptions also leads to cardiovascular disorders.

As perthe World Health Organisation, cardiovascular related diseases lead to death of 17.7 million lives every year, and it almost 31% of all global deaths. India is in top 5 positions in heart attack related death rates. Heart attacks have killed 1.7 million Indians in 2016, according to the 2016 Global Burden of Disease Report, released on September 15, 2017. Heart related diseases increase the spending on health care and also reduce the productivity of an individual. Estimates made by the World Health Organisation (WHO), suggest that India has lost up to $237 billion, from 2005-2015, due to heart related or cardiovascular diseases.

Artificial Intelligence (AI) and Machine Learning (ML) are the emerging technologies, which is going to lead automation world to the next generation technology. New and enhanced methods are applied in Neural Network and Deep Neural Network (DNN) to handle complex issues. Huge job and requirements are getting generated to processing these real

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 5- Issue 2, Paper 36 August 2022

time data, images and videos which has lead to a new technical field of Data Science or Data Analyst. In medical industry, all kind of CT, MRI, Ultrasound, ECG, EEG and many more scanned images data can be analyzed with Artificial Intelligence techniques.

II]Related Work

Many researchers have proposed multiple machine learning methods to diagnosis heart disease. J P Li [1] have recommended Prediction of the heart disease usingfeature selection algorithm FCMIM-SVM Techniques with an accuracy of 76%. The main objective is to predict the occurrence of heart disease for early automatic diagnosis of the disease within result in a short time.

M S Raja [2], have used Cleveland dataset having 303 data, with some data having missing attributes. Pre-processing of data is implemented by just removing the missing value from the dataset which are six in number and then from the remaining 297 data, they divided the data as training 70% and testing 30% before training Linear Regression algorithm.

Ali, Liaqat, et al, [5] propose a system containing two models based on linear Support Vector Machine (SVM). The first one is called L1 regularized and the second one is called L2 regularized. First model is used for removing unnecessary features by making coefficient of those features zero. The second model is used for prediction. Predication of disease is done in this part. To optimize both models they proposed a hybrid grid search algorithm. This algorithm optimizes two models based on metrics: accuracy, sensitivity, the Matthews's correlation coefficient, ROC chart and area under the curve.

III]Proposed Methodology:

The implementation of this work is done using UCI Machine Learning Repository Dataset. This is an open-source dataset which is downloaded from Kaggle website. This UCI dataset contains 303 samples of data with 14 input parameters and 1 output parameter. The output label is having two classes heart disease and no heart disease. The dataset is with different medical parameters such as age, sex blood sugar, heart rate, are some of the attributes are included to identify with the output class, if the person has heart attack disease or not. The list of the parameters is shown in the below Table.1.

S.No	Attribute	Value	Description
1	age	29 – 62	age in years
2	sex	0 – male, 1- female	gender
3	cp	1-typical angina; 2-atypical angina 3-non-anginal pain; 4-asymptomatic	chest pain type
4	trestbps	Numeric value(140mm/Hg)	resting blood pressure in mm/Hg
5	chol	Numeric value(289mg/dl)	serum cholesterol in mg/dl
6	fbs	1-true, 0-false	fasting blood pressure>120mg/dl
7	resteeg	0-normal, 1-having ST-T, 2-hypertrophy	resting electrocardiographic results
8	thalach	140,173	maximum heart rate achieved
9	exang	1-yes, 0-no	exercise induced angina
10	oldpeak	Numeric value	ST depression induced by exercise relative to rest
11	slope	1-upsloping, 2-flat, 3-downsloping	the slope of the peak exercise ST segment
12	ca	0-3 vessels	number of major vessels colored by flourosopy
13	thal	3-normal, 6-fixed defect, 7-reversable defect	thalassemia
14	num	0: < 50% diameter narrowing 1: > 50% diameter narrowing	diagnosis of heart disease (angiographic disease status)

Table.1. UCI Dataset

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 5- Issue 2, Paper 36 August 2022

The steps carried out in this system where the first method is to collect the details of the patients who suffered from the cardiovascular attack disease. After the complete observation and study of the data, only the required attributes are considered to carry out the pre-processingstage according to the algorithm used for implementation.Every supervisedmachine learning algorithm is implemented individually and the prediction of accuracy is done. The proposed system diagram is shown in the below Fig.1.

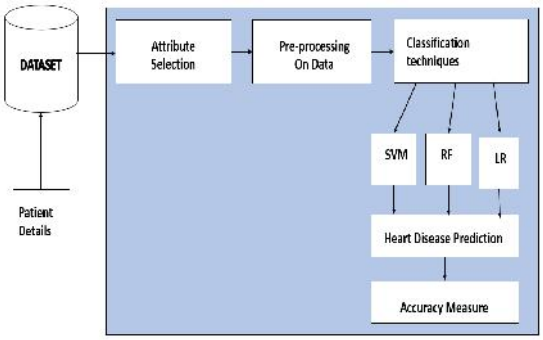

Fig.1. Proposed System

UCI Dataset is used to train the supervised machine learning model. The pre-processing stage involves data visualization. To analyze the data, pandas library is used. All the required libraries are installed using pip command. Jupyter notebook tool is used to execute the code. In jupyter tool, codes are written and executed in cells. While testing the code, individual cell can be executed which saves the developer time. Matplotlib and seaborn packages are used for graphical visualization. Matplotlib is best library package for data visualization. Graphical view makes the data reading in efficient way. Science kit learn (Sklearn) is the packages which played a major role in the analysis. Sklearn library was introduced in 2000, which is a library ofprogramming functions aimed primarily at projects based on machine learning algorithms.

The dataset is splited in ratio of 80:20 (train: test), to build the training model. The proposed methodology is mentioned in Fig.2.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 5- Issue 2, Paper 36 August 2022

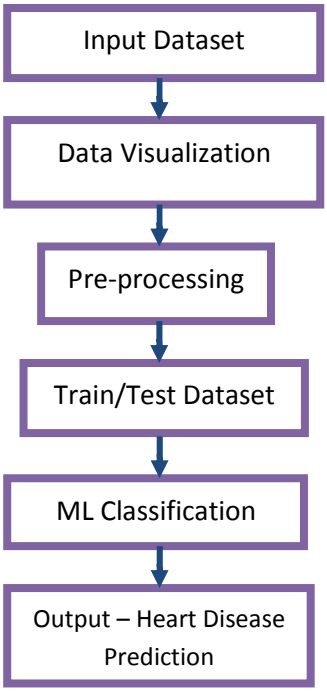

Fig.2. Methodology

IV]Results:

Different statistical approaches where analyzed in the implementation of pre-processing stage of UCI dataset. The correlation matrix of all the parameters was drawn using seaborn library packages. The training data is passed for all the selected supervised machine learning models. The below graph Fig.3, shows the representation of the bar graph which have comparison of all algorithms used for implementation.

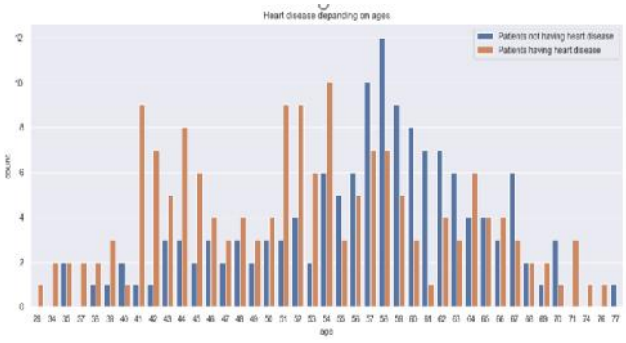

Fig.3. Heart attack analysis based on age attribute

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 5- Issue 2, Paper 36 August 2022

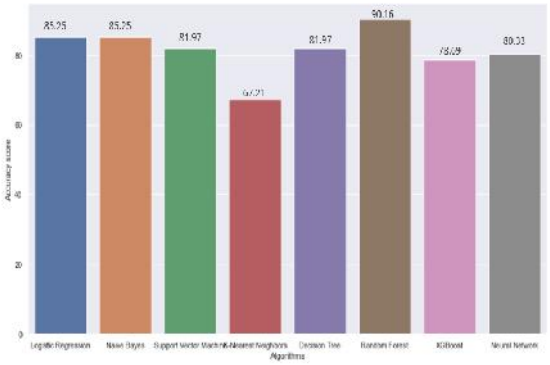

Fig.4. Algorithm Comparison Graph

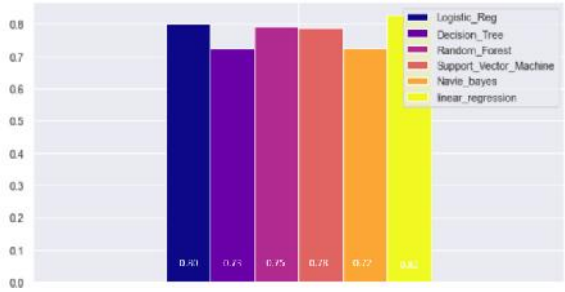

Fig.5. Algorithm Comparison Graph of coded Machine learning techniques.

The final best two models, Logistic Regression and Random Forest of the proposed method have given better validation accuracy for the training and testing data. F1 score, precession, recall were calculated with confusion matrix. The accuracy result of best two models is shown in below Fig.6.

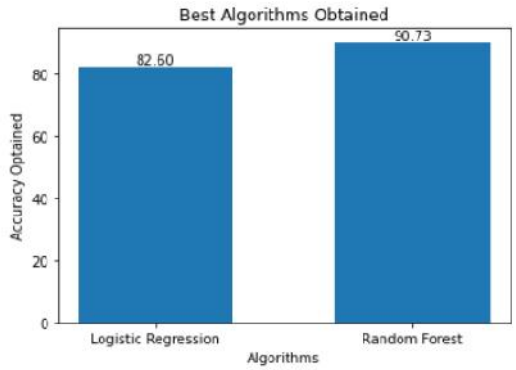

Fig.4. Accuracy Result of Best 2 models

V]Conclusion:

In this proposed work, an efficient method of multiple supervised machine learning algorithm were analysed using UCI Repository Machine Learning Dataset. The UCI dataset was used to train the listed algorithms likeNaive Bayes, Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Neural Network and K-Nearest Neighbour. With this study, the results to analyze cardiovascular attack more effectively analyzed withLogistic Regression of accuracy 82.6% and Random Forest of accuracy 90.7% were achieved.Based on the experimental results, our proposed model was able to outperform cardiovascular disease detection methods with respect to accuracy, precision, recall and F1 measure. As a future scope, this research work can be extended further with large dataset. The model developed can be used for patients to get early prediction of cardiovascular attack disease to save human life.

References:

[1] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," in IEEE Access, vol. 8, pp. 107562-107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[2] M. S. Raja, M. Anurag, C. P. Reddy and N. R. Sirisala, "Machine Learning Based Heart Disease Prediction System," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-5, doi: 10.1109/ICCCI50826.2021.9402653.

[3] I. Mirza, A. Mahapatra, D. Rego and K. Mascarenhas, "Human Heart Disease Prediction Using Data Mining Techniques," 2019 International Conference on Advances in Computing, Communication and Control (ICAC3), 2019, pp. 1-5, doi: 10.1109/ICAC347590.2019.9036836.

[4] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[5] R. Katarya and P. Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 302-305, doi: 10.1109/ICESC48915.2020.9155586.

[6] A. Gavhane, G. Kokkula, I. Pandya,"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.

[7] A. Kumari and A. K. Mehta, "A Novel Approach for Prediction of Heart Disease using Machine Learning Algorithms," 2021 Asian Conference on Innovation in Technology (ASIANCON), 2021, pp. 1-5, doi: 10.1109/ASIANCON51346.2021.9544544.

[8] M. Kavitha, G. Gnaneswar, R. Dinesh, "Heart Disease Prediction using Hybrid machine Learning Model," 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 1329-1333, doi: 10.1109/ICICT50816.2021.9358597.

[9] P. Motarwar, A. Suganya and M. Premalatha, "Cognitive Approach for Heart Disease Prediction using Machine Learning," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-5,doi: 10.1109

[10] S. Islam, N. Jahan and M. E. Khatun, "Cardiovascular Disease Forecast using Machine Learning Paradigms," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 487-490, doi: 10.1109/ICCMC48092.2020.ICCMC-00091.

[11] T. Y. Rashme, L. Islam, S. Jahan "Early Prediction of Cardiovascular Diseases Using Feature Selection and Machine Learning Techniques," 2021 6th International Conference on Communication and Electronics Systems (ICCES), 2021, pp. 1554-1559, doi: 10.1109/ICCES51350.2021.9489057.

[12] Prediction of Heart Disease Using Machine Learning by Aditi Gavhane,Gouthami Kokkula,Isha Pandya,Prof. Kailas Devadkar (PhD) ,Published: 2018 (ICECA)IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1

[13] McClellan, M., Brown, N., (2019) Urgent Challenges in Cardiovascular Disease: A Presidential Advisory From the American Heart Association. Circulation, 139: e44– e54.

[14] Pouriyeh, S., Vahid, S., G., Pietro, G. D.,. (2017) A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease. In 2017 IEEE symposium on computers and communications (ISCC). Heraklion. pp.204-207.

[15] Potdar, K., Pardawala, T. (2017) A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers. International Journal of Computer Applications, 175: 7–9.