

Predicting the risk of myocardial infarction using multi linear regression, neural network and logistic regression: A comparative Study

Ruchi Panwar¹, Amit kishor², Pankaj Pratap Singh³, Anirudh kumar tripathi⁴

¹ M.Tech Scholar CSE, department, S.I.T.E, Swami Vivekananda Subharti University, Meerut, India

^{2,3,4} Assistant Professor CSE department, S.I.T.E, Swami Vivekananda Subharti University, Meerut, India

ABSTRACT:

Myocardial infarction has become the main root of death in the world. In the field of medical science myocardial infarction detection is one of the growing area. Due to Heart disease one person kills in every 50 seconds-“According to American heart association” some of these death are occur due to the impaired blood supply and also even before the patient reach the hospital. Massive amount of patient’s related data is also maintained day by day. The stored data can also helped in detecting the chance of future disease and extraction of knowledge can be solved by the data mining technique. The main aim of this comparative study would be enable patients to become better informed about their condition and also motivate them to seek better care earlier in any situation for the detection of myocardial infarction. The data was collected from kaggle UCI repository, 10 attributes of clinical factors can be reported by the patients were studied. Researchers have already applied the multi linear regression, neural network and logistic regression individually but in our research work multi linear regression, neural network and logistic regression with accurate detection to detect the risk of myocardial infarction multi linear regression accuracy is 65.9% neural network accuracy is 77.9% and the best logistic regression model in the terms of performance accuracy 81.9%.

INTRODUCTION:

The heart is a main part of the human body. If the heart stop working in an efficient manner, our whole body can get affected badly and in most cases this disease can leads to death some of the main factors that contributes to the heart disease including family issues, high blood pressure, cholesterol level, age mostly above 35 no proper diet, decrease calcium rate, a stretched blood vessels of heart, lack of exercise etc. In today’s worlds the world health disease is the major challenging problem. The world health organization has approximated that 15 million death occur worldwide every year due to the myocardial infarction. To reduce the risk of heart disease prediction should be done diagnosis usually depend on sign, symptoms, physical exploration of the patient. Almost all the doctors have predicting heart disease by learning and experience. Even the diagnose of the disease is a difficult and arid or stuffy task in the medical field so the health care organization generates the large amount of data about the patients disease diagnose , electronic patients records and medical device etc. It is a key resource to processed and analyze for knowledge extraction that enable support for cost saving and decision making process. A number of difficulties arrive during diagnose such as less accurate result , less experience , time dependent performance and also accurate study of the patients clinical test data on the history of the individuals health. Various techniques have been applied by the researchers in various medical areas individually with accuracy. In this research work, accurate detection outcomes using multi linear regression technique, neural network and logistic regression technique for health care research. These techniques have comparison study and advantages and disadvantages. In the direction of the Paper aims to compare the accuracy performance of the multi linear regression, neural network and logistic regression and also achieve the accurate diagnosis. Data mining techniques involves a combination of statistical methods with machine learning algorithm. Data mining technique help the system in analyzing the symptoms and machine learning methods help in predicting the disease based on the analysis performed. From comparison study, we prove that which one is better for the myocardial infarction detection. The diagnosis of disease is difficult task in medical field hospitals and clinics collect a huge amount of patient’s data and

this data provide a basis for analysis of risk factor for myocardial infarction. The part and parcel objectives is to compare the study for detection of myocardial infarction using multi linear regression, neural network and logistic regression technique. From the comparison study we prove that the multi linear regression, neural network and logistic regression which is more better to detect the heart disease.

K.Subhadra ,Vikas B “.Performed Neural networks based Intelligent system for prediction heart disease”. Back propagation algorithm was applied to train the data and compare the perimeters iteratively.

K. Polaraju, D.Durgaprasad Performed the prediction of heart disease using multi linear regression model analyze has been performed to accurately predict the chance of heart disease.

Ayer and Chhatwal compared logistic regression and artificial neural network model in breast cancer risk estimation. Both the model have the potential to help physician w.r.t understanding cancer risk factor, risk estimation and diagnosis.

RELATED WORK:

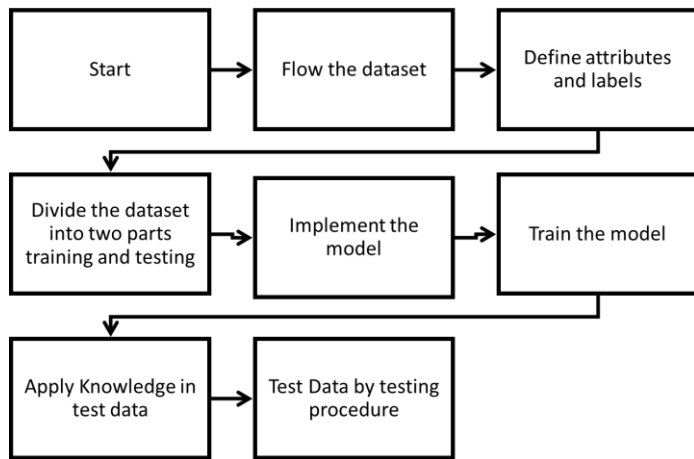
Extensive work in literature related with myocardial infarction detection using comparative analysis of neural network and logistic regression technique with better accuracy. Researchers have already developed the neural network and logistic regression technique accuracy individually for the heart disease detection. The researchers in the medical area, predict or diagnose the disease and also providing the efficient care for the patient by employing the neural network and logistic regression technique both. In this research paper we will describe + comparative study and analyze the data of myocardial infarction using both neural network and logistic regression technique with better accuracy. There are many approaches and algorithms have been used to predict the heart attack. Hai et.al work proposed neural based learning classifier system for classifying data mining tasks. They manage experiment on 13 different dataset from the university of California, Irvine repository and are artificial dataset. They exhibit that neural based learning classifier system performs fairly to supervised learning classifier system an five datasets remarkably good presentaton on six datasets and remarkably poor presentation on three datasets.

METHODOLOGY:

This methodology implemented the comparison concept of neural network and logistic regression technique are proved to be effective detection of myocardial detection with accuracy this concept is classify into 2 phases:In the first phase, following attributes are fed into as input and then trained with training data. In second phase, test with testing data.

The performance of the comparison criteria can be classify by following measures:

Make smart medical decision for the patient. Here this comparison study can analyze risk level using multi linear regression, neural network and logistic regression technique



There are many following objectives are as follows:

- 1) Doctor also follow the health recommendation for prevention of heart disease.
- 2) At high risk, patient require special attention
- 3) Accessibility and availability should be assured to all those who require both hospital and physician.
- 4) Medical and administrative should be made very clear for the patients of heart disease.
- 5) Check in and check out procedures should be patients friendly.

SOURCES:

Myocardial infarction disease dataset have been taken from a kaggle UCI machine learning repository . The dataset provides the Myocardial infarction disease patients information for various attributes. For this study, around 338 patient data with 10 attributes has been used for the prediction analysis. . Table 1 shows the information for all the attributes with their ranges.

Table1.1. Heart disease dataset with different attributes and range

S.No.	Attribute Name	Attribute Information	Range (Min-Max)
1	age	age in years	28-75
2	sex	represented as binary no. 1- Male 0- Female	0-1
3	CP	chest pain type value 1. Typical angina, 2. atypical angina, 3. non-anginal pain, 4. asymptotic	1-4
4	BP	resting blood pressure measure (in mm hg admission to	92-200

		hospital)	
5	Chol	serum cholesterol in mg/dl	85-603
6	FBS	fasting blood pressure > 120-mg/dl 1 – true 2 – false	0-1
7	ECG	Restecg- resting electrocardiographic results	0-2
8	HR	maximum heart rate achieved	69-190
9	Exer	exercise induced angina 1 – yes 2 – no	0-1
10	old peak	ST depression induced by exercise Relative to rest	0-5

The distribution of age for the selected data is shown in the figure. The data is considered for a wide range from 28 year to 75 year of the age. The probability of attack is more in the range 40 year to 60 year. Hence more data is considered in this range for better accuracy of prediction. The data includes from both the sex (Male and Female) as shown in the figure.

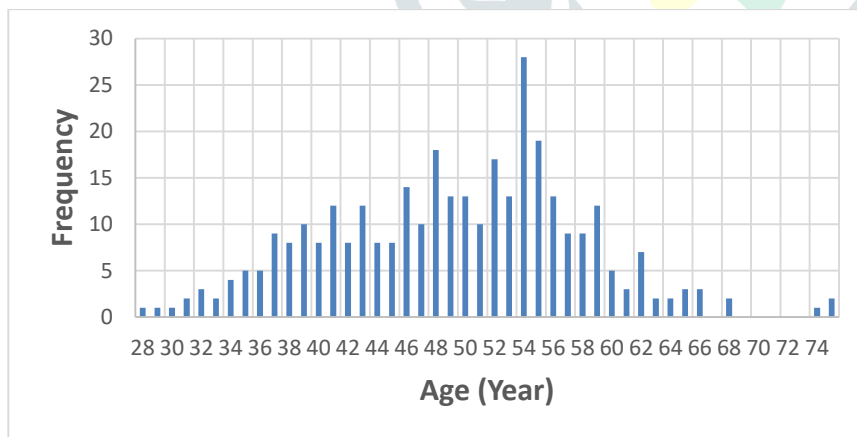


Figure 1

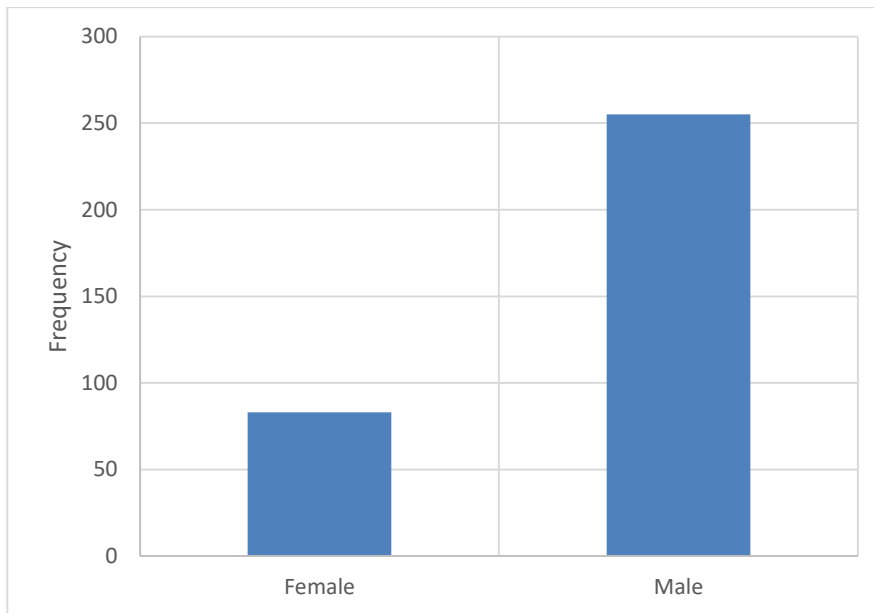
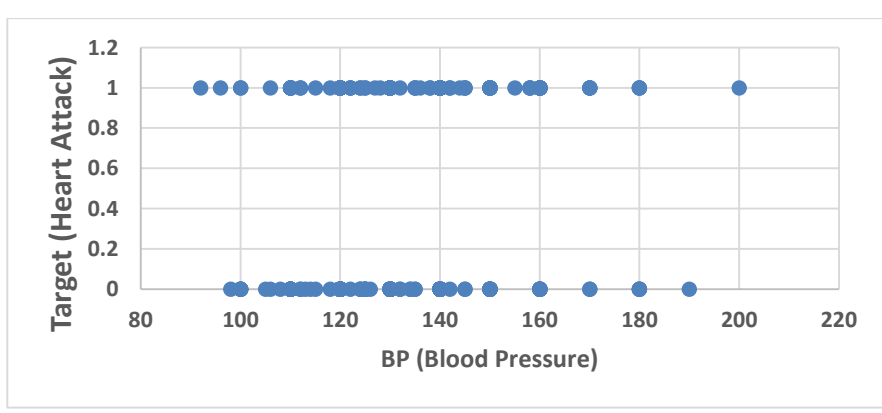
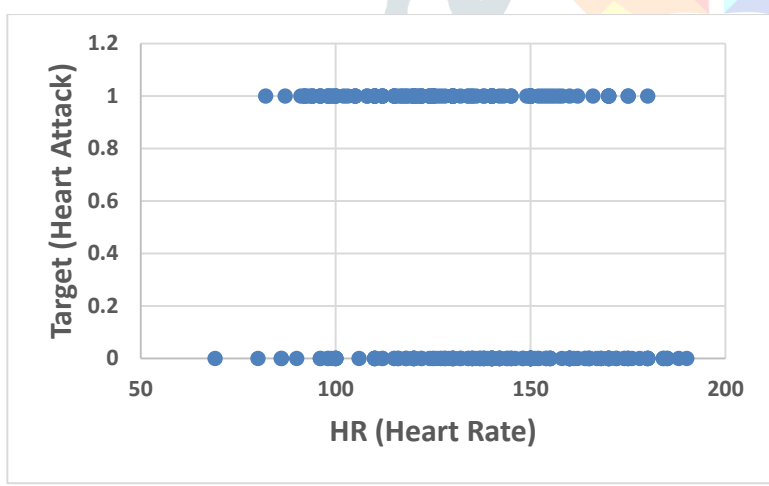


Figure 2

Result Analysis: In this research, author predicts the Myocardial infarction and compare the accuracy with different predicting techniques such as ANN, BLR and MLR. Figure shows the effect of Heat Beat, Blood pressure and cholesterol level on the chances of the heat attack possibilities (o-No, 1-Yes). In the heart diseases data set, the effect of few attributes such as Heat Beat, Blood pressure and cholesterol level.



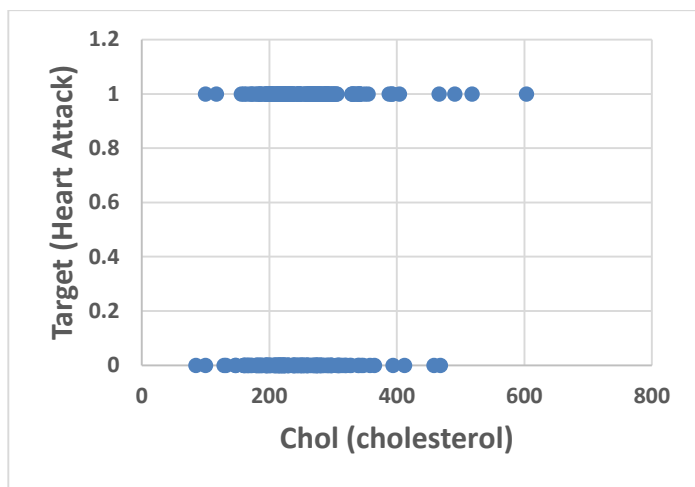


Figure 3: Effect of Heart rate, Blood pressure and cholesterol on the risk of heart attack (1- yes, 0-No)

Development of Multi Linear Regression Model:

Multi linear regression is a statistical model that can be used to describe the data and explain the relationship between one or more dependent variable and two or more independent variable.

1. Analyzing the co relation directionality of the data.
2. Fitting the line.
3. Evaluating the validity and usefulness of the model are the different stages of multi linear regression .The regression line represent the approximated disease chance for a given combination of input factors. The coefficient of determination(R squared) is a statistical matrix which is used to measure how much of the variation in the outcome R square always increases as more predictor are added to the MLR model even though the predictor may not be related to the outcome variable.

OVERALL FIT

Multiple R	0.659019
Standard Error	0.374519
Observations	338

Development of Neural Network model:

Artificial Neural Network (ANN) is used for predicting the heart attack on the above mentioned dataset of myocardial infarction. ANN performs a simple computing task normally a basic yes or no decision and they consist of highly interconnected nodes with three different layers; Input layer, Hidden layer and an Output layer as shown in figure. The neural network require three main parts training, validation and testing.

It is used to prevent over fitting in neural network and it is used the output from the input layer is passed through the weight to the hidden layer and the hidden layer is made up of some function and activation function. The activation function used is shown in figure :

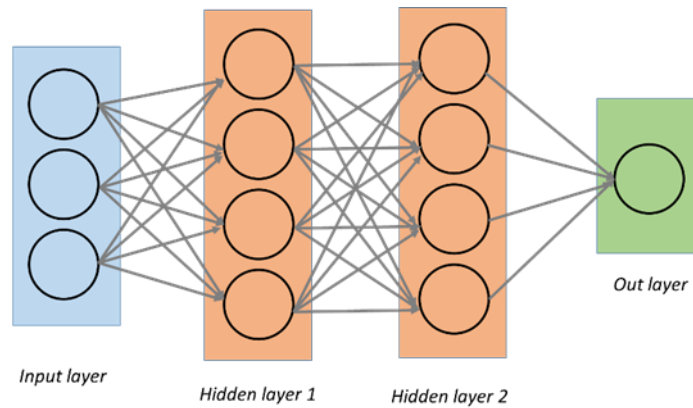


Figure: 4 ANN model with different layers

We have trained the data and then tested on the selected data set of myocardial infarction. For neural network analysis 80% training and 20% testing dataset was used. Following layers have been used for the ANN analysis including one input, two hidden layers and one output layer. The details of Neurons for each layer given as, Layer 1: Input layer-10 Neurons, Layer 2: Hidden layer- 6 Neurons activation (Function = Relu), Layer 3: Hidden layer- 6 Neurons activation (Function = Relu) And Layer 4: Output layer- 1 Neuron activation.

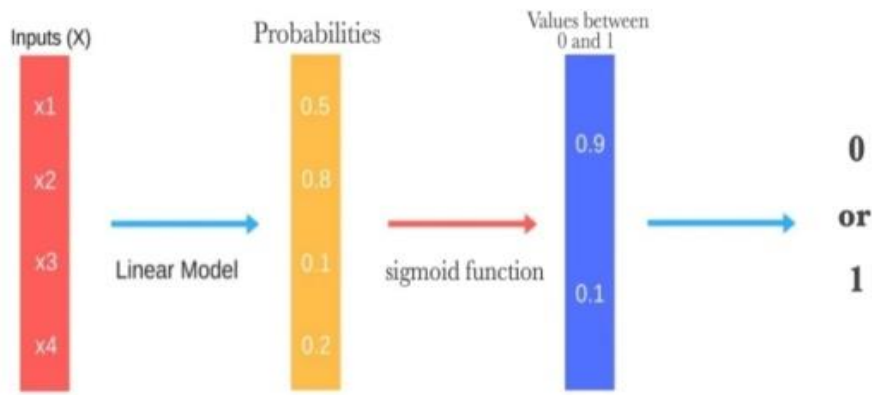
Table2: Confusion matrix for ANN analysis

	Prediction class 1	Prediction class 0
Actual class 1	34	11
Actual class 0	4	19

The confusion matrix for ANN analysis is shown in table 2. The ANN analysis of tested data results in true positive-34, true negative -19, false positive-4, false negative-11. This information helps to calculate some of the required characteristics of the ANN model. Hence, the accuracy, recall, specificity and precision of ANN for predicting myocardial infarction is 77.9 %, 75.5%, 82.6% and 89.4%.

DEVELOPMENT OF LOGISTIC REGRESSION:

Logistic regression is also applied on the selected dataset of myocardial infarction for predicting the heart attack. The logistic regression is a predictive modeling technique and also analyze the data. This technique find the Probability of event (Success), event(failure). A probability out point 0.5 was used to classify observation as events or non-event.



Flow chart of work progress

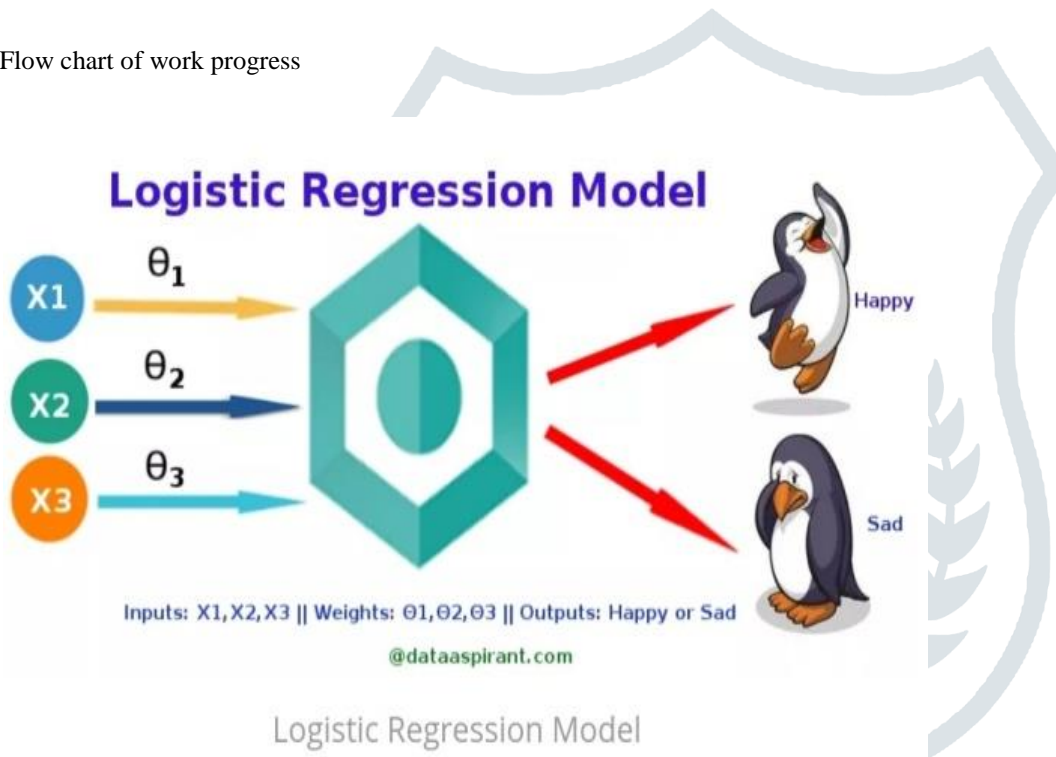


Figure: 5

The overall accuracy of the logistic regression model is defined by comparing the predicted values with actual events. The main motive of the logistic regression is to determine the result of each variable correctly. The logistic regression also refer to as the logistic model that provide the categorical variable with 2 categories health / unhealthy, with risk / without risk. If X_1, X_2, \dots, X_M denotes X is a predicted variable Y denotes the presence $Y = 1$, absence $Y = 0$ disease and P denotes the probability of disease presence that is $Y = 1$

The ROC curve (true positive rate Vs false positive rate) for logistic regression analysis for selected dataset of myocardial infraction for predicting the heart attack.

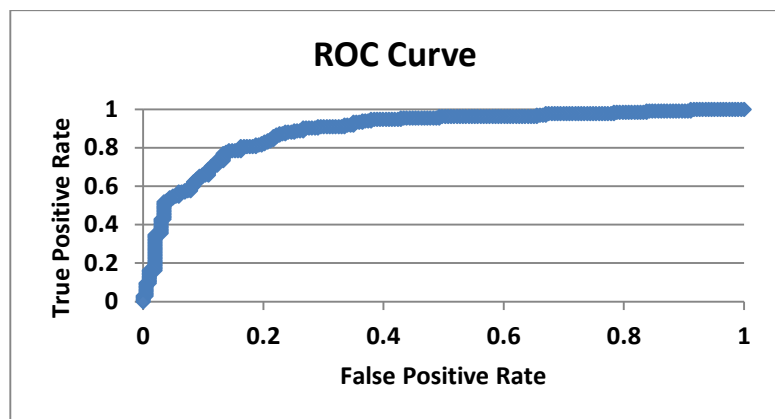


Figure: 6 ROC curve for logistic regression analysis

The classification table for the logistic regression analysis is given below. The accuracy achieved with logistic regression analysis on the selected data set is 81.9%. This accuracy is calculated based on the success objects and fail objects for selected dataset.

Table 3: Classification Table for logistic regression model

	Suc-Obs	Fail-Obs	
Success - prediction	99	25	124
Fail-prediction	36	178	214
	135	203	338
Accuracy	0.733333	0.876847	0.819527

RESULT ANALYSIS:

TECHNIQUES	ACCURACY
Multi linear regression	65.9%
Neural network	77.9%
Logistic regression	81.9%

Here, the result analysis between these techniques shows the accuracy and also shows which one is better technique for the myocardial infarction detection. Logistic regression can predict only discrete dependent variable of binary form. It should not be correlated models the data in binary values. Here, the logistic regression techniques accuracy is much better from other techniques that is neural network and multi linear regression for the prediction of myocardial infarction. This paper proposed a comparative study between the accuracies of applying these techniques that is multi linear regression, neural network and logistic regression with the individual result and also by applying these techniques on the heart attack dataset in order to predicting the heart disease.

CONCLUSION:

The study compared model for the detection of myocardial infarction using heart attack dataset from kaggle website repository. This dataset was applied on a multi linear regression, neural network and logistic regression techniques. This detection ability of the neural network model was organized to be comparable to that of the logistic regression technique. The comparative study of the multi linear regression, neural network and logistic regression. Here, the multi linear regression accuracy is 65.9% and neural network accuracy to detect the myocardial infarction that is 77.9% while logistic regression accuracy to detect the myocardial is

81.9% .Logistic regression model shows better accuracy as compared to the ANN and MLR for selected dataset of myocardial infarction. Better accuracy shows that logistic regression more accurate for the myocardial infarction and from this disease we can also determine that male or female which one is more survival from myocardial infarction. Here, the logistic regression technique is much better accurate for the detection of myocardial the performance was applied to get an accurate measure. In future, the reliability of the model can be verified by applying both the predicting model in larger dataset of myocardial infarction.

REFERENCES:

- 1.K.Subhadra , Vikas B “.Neural networks based Intelligent system for prediction heart disease”. ISSN-2278-3075.(2019)
- 2.AbdulGhaffar Khan “Prediction of heart disease using artificial neural network. (2018)
3. Sofia Tomov “For deep neural network for detecting heart disease”.2018
- 4..Kartik Chopra , C. Srimathi “Logistic regression and CNN performance analysis based on the size of the dataset. ISSN -2321-9939” . 2018
5. AnimeshHazra , Subrata Kumar Mandal , Amit Gupta , Arkomita Mukherjee and Asmita Mukherjee to summarize some of the current research on predicting heart disease using data mining technique. ISSN-0973-6107. (2017)
- 6.K.Polaraju, D.Durga Prasad “Prediction of heart disease using multi linear regression model”. ISSN-2321-9939.(2017)
- 7.Parisa Naraei , AbdolrezaAbhari and AlirezaSadeghian “Application multi layer perceptron neural network and support vector machine in classification of health care data IEEE. (2016).
- 8.<http://www.statisticssolutions.com/multiple-linear-regression/>
- 10.EbenezerOlaniyi , Adnan Khashman , OyebadeK.Oyedotun “Heart disease diagnosis using neural network arbitration
11. A.T.Sayad ,P.P.Halkarnikar , presented diagnosis system for heart disease using neural network approach data mining and Artificial neural network techniques. ISSN-2321-9009(2014)
- 12.S.Florence , N.G BhuvaneshwarAmma . G. Anna poorani ,K.malathi “ Predicting the risk of heart attack usingneural network and decision tree.
- 13)<http://pubs.rsna.org/doi/full/10.1148/rg.301095057>
- 14)<http://mindmajix.com/artificial-neural-network-and--how-it-works>
- 15.M.AkhilJabbar , B.L.Deekshatulu and Preeti Chandra “Classification of heart disease using artificial neural network and feature subset selection” ISSN-0975-4172(2013).
- 16)<https://archive.ics.uci.edu/ml/datasets/heart=disease>
- 17.ChaitraliDangare (A data mining approach for prediction of heart disease using neural network) 2012