EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

Risk Assessment of Myocardial Infarction for Diabetics through Multi-Aspects Computing

Shiva Shankar Reddy[1,*], Nilambar Sethi[2], R. Rajender[3]

[1]Research Scholar, Department of Computer Science and Engineering, Biju Patnaik University of Technology, Rourkela, Odisha, India
[2]Department of Computer Science and Engineering, GIET, Gunupur, Odisha, India
[3]Department of Computer Science and Engineering, LENDI Engineering College, Vizianagaram, India

Abstract

INTRODUCTION: Myocardial infarction (MI) is a type of cardiovascular disease. Cardiovascular disease is the major side effect of diabetes. It causes damage to heart muscle due to interruption in the blood flow. The chance of getting this disease is high in diabetes patients.
OBJECTIVES: To choose a dataset with features related to diabetes, parameters of ECG and risk factors of MI for effective prediction. Predict myocardial infarction in both type-1 and type-2 diabetic patients using regression techniques. Recognise the best algorithm.
METHODS: Multiple linear regression, ridge regression and lasso regression are existing techniques in addition to which proposed technique lasso regression is used to develop a model for prediction. The trained models are compared to know better performing algorithm. Estimation statistics namely confidence and prediction intervals are used to show the amount of uncertainty in predicted values. The statistical measures in regression analysis namely root mean squared error and r_squared value are used to evaluate and compare algorithms.
RESULTS: The proposed algorithm 'lasso regression' has achieved better values of RMSE and r_squared as 0.418 and 0.2278 respectively compared to remaining techniques.
CONCLUSION: Best performance of proposed algorithm was noticed and hence using lasso regression for prediction of myocardial infarction in diabetes patients gives better results.

Keywords: Myocardial infarction, Diabetes, Multiple linear regression (MLR), Ridge regression (RR), Lasso regression, Confidence and prediction intervals, Root mean squared error (RMSE) and r_squared.

Received on 07 July 2020, accepted on 10 December 2020, published on 16 December 2020

doi: 10.4108/eai.17-12-2020.167655

*Corresponding author. Email:shiva.shankar591@gmail.com

1. Introduction

Cardiovascular disease is one of the major complications of diabetes. Myocardial infarction is one type of cardiovascular disease. It is also known as heart attack which is caused because of interruption in the blood flow that damages the heart muscle. There is more chance for diabetic patients to have a heart attack compared to non-diabetic patients. In a study related to diabetes at oxford centre stated that patients who are affected with fatal myocardial infarction are having higher HbA1C compared to normal myocardial infarction patients. [1] HbA1C is the observed average blood sugar levels from past two to three months.
A heart attack or myocardial infarction can be considered as a form of acute coronary syndrome. Acute coronary syndrome is a condition that occurs when the arteries got blocked. These arteries play a major role to carry oxygen, blood and nutrients to heart. [2] There are three types of acute coronary syndrome. They are STEMI, NSTEMI and CAS or unstable angina.

- ST segment elevation myocardial infarction (STEMI) is a condition which occurs when a coronary artery got blocked completely and doesn't allow any blood flow through the heart and damages the muscle.
- Non-ST segment elevation myocardial infarction (NSTEMI) is a condition which occurs when a coronary artery is blocked partially.
- Coronary artery spasm (CAS) or unstable angina is a condition which occurs when the arteries of heart are tightened and reduces or stops the blood flow. [3]

There are several signs one can observe before affecting to heart attack. The chest pain is an early sign when treated can reduce complication. The common symptoms for myocardial infarction include shortness of breath, tiredness, indigestion, heart burn and nausea. [4] More than 68% of diabetic people are affected to heart diseases with age greater than or equal to 65. [5]

Several risk factors which increase the chances of heart attack are high blood pressure, high blood cholesterol, obesity, diabetes, age, family history of heart attack, lack of exercise and use of tobacco. [4] In order to avoid myocardial infarction one should follow a better lifestyle and take medication if having any one of the risk factors. In a survey conducted across the world it was noticed that 32.2% of people affected to cardiovascular disease are having type-2 diabetes. The type-2 diabetes patients have 53% chance of affecting to myocardial infarction. [6]

A Doctor will diagnose heart attack or myocardial infarction by performing tests like electrocardiogram (ECG), blood test, echocardiogram and angiogram. ECG is performed to monitor the electrical activity of heart which shows the measures in the form of graph. A blood test is done to identify the leak of proteins present in heart into the blood. An angiogram test will identify the areas where arteries are blocked. An echocardiogram provides the images of heart which is used to monitor the functioning of valves and for identifying blood clots. [7]

In section 2 the literature work is provided. In section 3 the methodology of this work is provided that explains about the objectives of the work, dataset used and system architecture in detail. In section 4 the proposed work is provided in which MLR and RR algorithms are explained briefly and Lasso regression algorithm is explained in detail. In section 5 the analysis of obtained results is provided. The detailed explanation of estimation statistics and statistical measures used in the work is also given. In section 6 the conclusion of this work is provided.

2. Literature Survey

Xingjin Zhang et al. [8] used long short term memory (LSTM) neural network algorithm to diagnose myocardial infarction. The dataset used by them is an ECG database named as physikalisch technische bundesanstalt (PTB). They performed pre-processing on the ECG signals and divided that into a sequence containing heartbeat. The performance measures used are accuracy, sensitivity, specificity and positive predictivity. Only overall accuracy of the model is provided in this paper and the values of remaining three measures are obtained for each target class. An overall accuracy of 99.91% is obtained for LSTM which is better than other works.

Wenhan Liu et al. [9] focused on diagnosis of myocardial infarction. The proposed algorithm is multiple feature branch convolutional neural networks (MFB-CNN). They have implemented the algorithm on both ECG classification data and patient physiological data for automated MI detection and localization. The PTB dataset is considered for ECG classification. Accuracy, sensitivity and specificity are the metrics chosen for MI detection and accuracy for MI localization. In case of ECG classification data, the values of accuracy, sensitivity and specificity obtained for MI detection were 99.95%, 99.97%, 99.9% respectively and 99.81% overall accuracy was obtained for MI localization. In case of patient data, the values of accuracy, sensitivity and specificity obtained for MI detection were 98.79%, 98.73%, 99.35% respectively and 94.82% overall accuracy was obtained for MI localization. Thus, the proposed algorithm by them obtained better values compared with other existing methods.

Rajendra et al. [10] highlighted their work on detecting myocardial infarction using ECG signals. The used dataset is containing ECG signals taken from PTB database. They proposed deep convolution neural networks algorithm to implement on data with noise and without noise form ECG dataset. The performance measures used are accuracy, sensitivity, positive predictive value (PPV) and specificity. They concluded that the proposed algorithm has obtained good measures for data with noise and without noise. In case of data with noise the values of accuracy, PPV, sensitivity and specificity obtained were 93.53%, 98.03%, 93.71% and 92.83% respectively. In case of data without noise the values of accuracy, PPV, sensitivity and specificity obtained were 95.22%, 98.43%, 95.49% and 94.19% respectively.

Umamaheswari and Isakki [11] performed k-medoid clustering technique to predict myocardial infarction. They used dataset gathered from UCI machine learning (ML) repository. The dataset contains 14 variables out of which 13 are predictor variables and one is target variable. They performed data pre-processing and feature selection on the dataset. Then the proposed clustering algorithm is implemented on the features selected by applying oneR and relief feature selection techniques. They provided results showing that the data is clustered into groups and can be predicted easily.

Runchuan Li et al. [12] presented their work on prediction of cardiovascular disease using random forest (RF) algorithm. They considered data set from UCI ML repository. They compared RF with five existing ML algorithms namely SVM, logistic regression, decision tree, naive bayes and radial basis function (RBF). The metrics used for comparison are sensitivity, specificity, precision and AUC. From result analysis they summarized that RF has performed better. The values of sensitivity, specificity, precision and AUC obtained for RF were 0.880, 0.876, 0.880 and 0.947 respectively.

Procheta Nag et al. [13] developed a system for predicting myocardial infarction using data mining techniques. They collected the dataset from different hospitals containing 25 attributes related to acute myocardial infarction. The data mining techniques used are C4.5 decision tree (DT) and random forest. The percentage split of 70%, 60% and 55% were performed on dataset using seed values 1 to 4. They considered evaluation metrics like accuracy, precision, recall and ROC curve. C4.5 has obtained best values of metrics for all three percentage splits using seed 3. Random forest algorithm is implemented using seed 3 and compared with C4.5. The values of accuracy, precision, recall and ROC curve obtained for random forest in case of 70% percentage split were 96, 0.94, 1 and 0.99, in case of 60% percentage split are 96, 0.97, 0.97 and 0.99, and in case of 55% percentage split the values were 95, 0.95, 0.97 and 0.99 respectively. From the results obtained random forest algorithm has performed well so they used that algorithm to develop an app for predicting myocardial infarction using data provided.

Polaraju and Durga Prasad [14] used MLRtechnique to predict heart disease. They used the data collected from patients. They implemented this algorithm in C# language using .Net framework. They divided training and test datasets using 70% and 30% respectively. The training dataset consist of 13 attributes and 3000 instances. They concluded that the result of MLRmodel is more appropriate for prediction of heart disease.

Neel Adwani [15] presented his work on predicting the probability of affecting to heart attack. He used three attributes from heart disease dataset in kaggle namely age, cholesterol level and target class for prediction purpose. The machine learning regression algorithm called MLRis used by him to predict heart attack. The implementation is done using GNU octave which is an open source software. When age and cholesterol level is given as input the chance of heart attack is predicted by the model. He concluded that adding more predictor attributes can increase the accuracy of the model.

Madhubala et al. [16] provided their work for prediction of diabetes using multiple linear regression. The dataset used by them is a diabetes dataset taken from kaggle. The dataset attributes considered are glucose, BP, insulin, age, BMI and outcome (target variable). They calculated correlation between each predictor variable and target variable. Among these correlation values best two predictor attributes glucose and BMI are used to train MLRmodel. The visualization of the output is provided. From analysis of results they noticed that glucose level > 100 and BMI value > 20 indicates the presence of diabetes.

Muthukrishnan and Rohini [17] presented their research work on using regression techniques for developing predictive models in machine learning. They implemented OLS regression, RR and LASSO regression on real time diabetes dataset. They used these algorithms for feature selection which provided coefficients of predictor attributes. They concluded that the LASSO regression technique has performed better with minimum attributes than the other two techniques.

Jeena and SukeshKumar [18] used RR for predicting risk of stroke. They used clinical dataset collected from a hospital in Trivandrum. The dataset contains 14 attributes and 531 instances. They used bootstrap validation to validate the model developed using RR. They calculated risk score based on which they predicted the chance of effecting to stroke.

Huan Lio and Yahui Liu [19] focused on using RR for developing prediction model. They used forest fire prediction dataset for implementing RR. First they performed RR for feature selection to get efficient features then those attributes are used to develop the support vector machine model using radial basis function as kernel type. They concluded that the prediction is accurate by combining regression technique with SVM.

Avinash Golande and Pavan Kumar [20] highlighted their work on prediction of heart disease. They used some effective ML techniques like DT, k-mean clustering, Ada-boost and k-nearest neighbour on the heart disease dataset. They have compared the output of these algorithms with classifiers used in already existing research papers.

Alexander Schlemmer et al. [21] predicted cardiac diseases using several ML algorithms. They used the data collected from a cardiological study which contains information of 261 patients. The algorithms they have used are support vector machine with linear and RBF kernels, k-nearest neighbours with k values as 1, 3, 5, 8 and random forest. The leave one out (LOO) test and Matthews correlation coefficient was performed on each classifier. They concluded that linear SVM has performed better in terms of Matthews correlation coefficient with a value of 0.28.

Santhana Krishnan and Geetha [22] presented their work on prediction of heart disease. The dataset used by them is in terms of medical data taken from UCI ML repository. They used algorithms like decision tree and naive bayes for prediction of heart disease. By comparing results obtained for both algorithms they concluded that decision tree has obtained better accuracy of 91%.

Arash Farbahari et al. [23] focused on using linear, ridge and lasso regression for determining influential variables that affect fasting sugar levels in type-2 diabetic patients. After determining the influential variables they implemented logistic regression to predict type-2 diabetes. The dataset used by them is collected from 380 healthy persons and 270 type-2 diabetic patients. The three regression algorithms are compared using mean squared error (MSE). They concluded that among all the attributes HbA1C attribute is more influential predictor attribute for fasting sugar level.

3. Methodology

This section comprises of objectives of this work, detailed explanation of dataset used and system architecture.

3.1. Objectives of the work

In recent days ML algorithms are most prominently used in various areas including medical research (disease predictions). The chance of affecting to myocardial infarction is high in diabetic patients of both type-1 and type-2. Early and effective diagnosis of MI in diabetic patients will help to avoid further risks. The objectives of this work to accomplish the proposal are

- To consider dataset attributes related to diabetes, risk factors of MI and parameters from ECG tests.
- Choose ML algorithms for developing a myocardial infarction predictive model.
- Evaluate each algorithm using performance metrics.
- Recognize the best performing one among all the algorithms.

The dataset considered in this work comprises of variables or attributes related to diabetes, risk factors of MI and the parameters from ECG test. The dataset attributes that effect type-1 and type-2 diabetic patients are considered. These attributes are necessary to effectively build a predictive model. Three regression algorithms were chosen for developing a predictive model. They are MLR, RR and lasso regression. These algorithms are implemented in R programming. The statistical metrics should be considered to evaluate the performance of regression algorithms. The estimation statistics namely confidence interval and prediction interval are used to show uncertainty in predicted values. The statistical metrics namely RMSE and R squared were chosen for evaluating and comparing algorithms to recognize the best performing algorithm.

3.2. Dataset

The dataset considered to implement the algorithms consists of 22 attributes and 133 instances. Among these attributes 21 are predictor attributes and one is target attribute. As the main aim of this work is to predict the myocardial infarction in diabetes patients, most of the predictor attributes are related to diabetes and electrocardiogram parameters. The predictor attributes related to diabetes are body mass index (BMI), type of diabetes, duration of diabetes, fasting blood sugar level, HbA1C and type of treatment for diabetes.

Table 1. Dataset description

Att. no	Att. Name	Description
1.	Age	This attributes provides the age of the patient.
2.	Gender	The gender of the patient. 1-Female, 2-Male.
3.	BMI	Body Mass Index is calculated using height and weight of the person. The formula is given by weight in kg / (height in m)2.
4.	DM_type	Type of diabetes mellitus the person is having. 1 indicates type-1 diabetes, 2 indicates type-2 diabetes.
5.	DM_duration	Duration of diabetes mellitus of the person. The number of years the patient is suffering from diabetes.
6.	FBS	Fasting Blood Sugar level is the sugar level obtained by performing blood test after fasting for at least 8 hours. The normal range is 70-110 mg/dL.
7.	HbA1C	HbA1C is known as Haemoglobin A1C test which gives the average blood sugar level for previous two to three months.
8.	LDL	Low density lipoprotein cholesterol is called as bad cholesterol. The normal range for both men and women is less than 100 mg/dL.
9.	HDL	High density lipoprotein cholesterol is called as good cholesterol. The normal range for men is greater than or equal to 40 mg/dL and for women is greater than or equal to 50 mg/dL.
10.	TG	Triglyceride means the fats from the food carried in the blood. The normal range is less than 150 mg/dL.
11.	DM_treat	Type of diabetes treatment the patient is taking. 1- Oral(medicine), 2-Insuin(injecting insulin), 3- Both oral and insulin.
12.	Statin	Drug suggested by the doctor to use for reducing the cholesterol levels in the patients.
13.	Dose	Dosage of statin the patient should take daily. The values 20, 40, 80 indicates milligram of dosage and 0 means no statin usage.
14.	Sys_bp	Systolic blood pressure is the pressure in the flow of blood during contraction of heart muscle. The normal range of SBP is less than or equal to 120 mmHg.
15.	Dia_bp	Diastolic blood pressure is the pressure in the flow of blood between the heart beats. The normal range of DBP is less than or equal to 80 mmHg.
16.	Smoking	Whether the person is having habit of smoking or not. 0- No, 1-Yes.
17.	Restecg	Resting electrocardiographic results obtained by observing ECG graph. The result is expressed using classification method where 0- normal, 1-ST-T wave abnormality and 2- definite or probable left ventricular hypertrophy (LVH).
18.	Thalach	Maximum heart rate achieved for the patient.
19.	Old peak	Exercise induced ST depression related to rest in ECG.
20.	Slope	Slope of peak exercise ST segment in ECG. 1- up sloping, 2- flat, 3- down sloping.
21.	Thal	Thallium heart scan output 3 means normal, 6 means fixed defect and 7 means reversible defect. Here 1 indicates normal, 2 indicates fixed defect and 3 indicates reversible defect.
22.	Class	This is the target attribute which indicates whether the patient is tested 1- positive or 0- negative for chance of affecting to myocardial infarction.

The attributes related to electrocardiogram are results of resting electrocardiograph, maximum heart rate achieved, slope of the ST segment, exercise induced ST depression and thallium heart scan. The attributes results of resting electrocardiograph, slope of the ST segment and thallium heart scan are not in the original graphical form of ECG but they are represented in the form of different classes which are described in below table 1.

Some attributes are related to risk factors of myocardial infarction like cholesterol, age, use of tobacco (smoking) and high blood pressure. The attributes related to cholesterol are parameters of lipid profile test namely LDL, HDL and triglyceride, statin and dosage of statin. The attributes related to blood pressure are systolic blood pressure and diastolic blood pressure. In addition to these the gender attribute was also included.

Figure 1 describes the density plot for all the 22 attributes in the dataset. The name of the attribute is mentioned on the top of each plot. The N value represents number of observations in the dataset which is 133. Kernel smoothing is used in the density plot. In kernel smoothing each data point is represented as Gaussian shaped kernel and all these Gaussian kernels are combined to obtain density plot. The bandwidth under each attribute density plot represents the standard deviation of the smoothing kernel for that respective attribute.

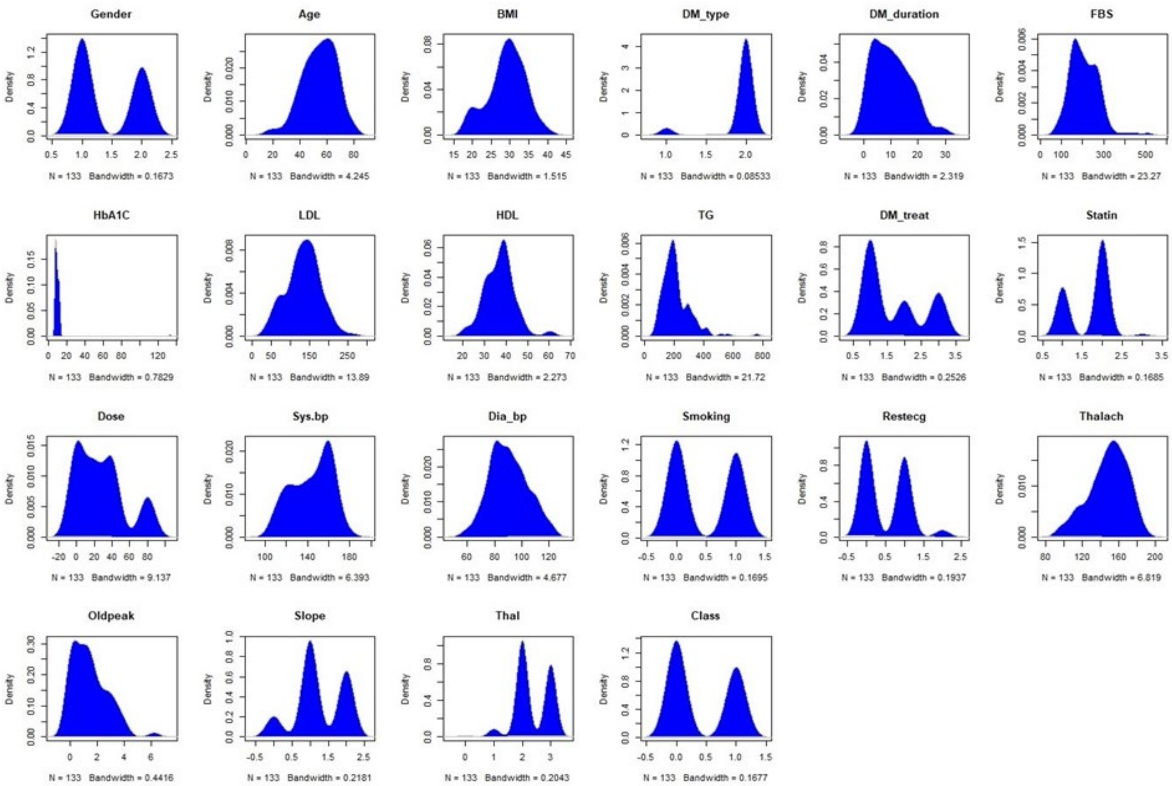

Figure 1. Density plot of attributes in the dataset

3.3. System architecture

A good architecture that describes the entire process is necessary to better understand the work. In few works, only the process is mentioned without presenting any flowchart. Rather than stating only the process, a flowchart is included to describe the process clearly. The system architecture of this work is presented in figure 2. It describes the step by step methodology to obtain best performing algorithm. First the myocardial infarction dataset described in the above section is loaded. The size of the dataset is 133 instances and 22 attributes. Then data pre-processing is performed. The percentage split of 80% is performed on the pre-processed data. The training and test datasets contains 107 instances and 26 instances respectively. Using the training dataset the three algorithms multiple linear regression, ridge regression and lasso regression are implemented which gives the trained model for each algorithm. The each trained model is evaluated using test data and provides the result. These results are compared to obtain the best performing model finally.

This entire process is implemented using R programming. The confidence intervals, prediction intervals, r_squared value and RMSE are used for evaluating each algorithm. The confidence and prediction intervals are used to show the uncertainty in the predicted values. r_squared and RMSE are the statistical measures based on which the best performing algorithm was recognized. The small value of RMSE and large value of r_squared is the criteria for the best algorithm. After comparison lasso regression was recognized as the best one and it was explained in a detail way in result analysis section

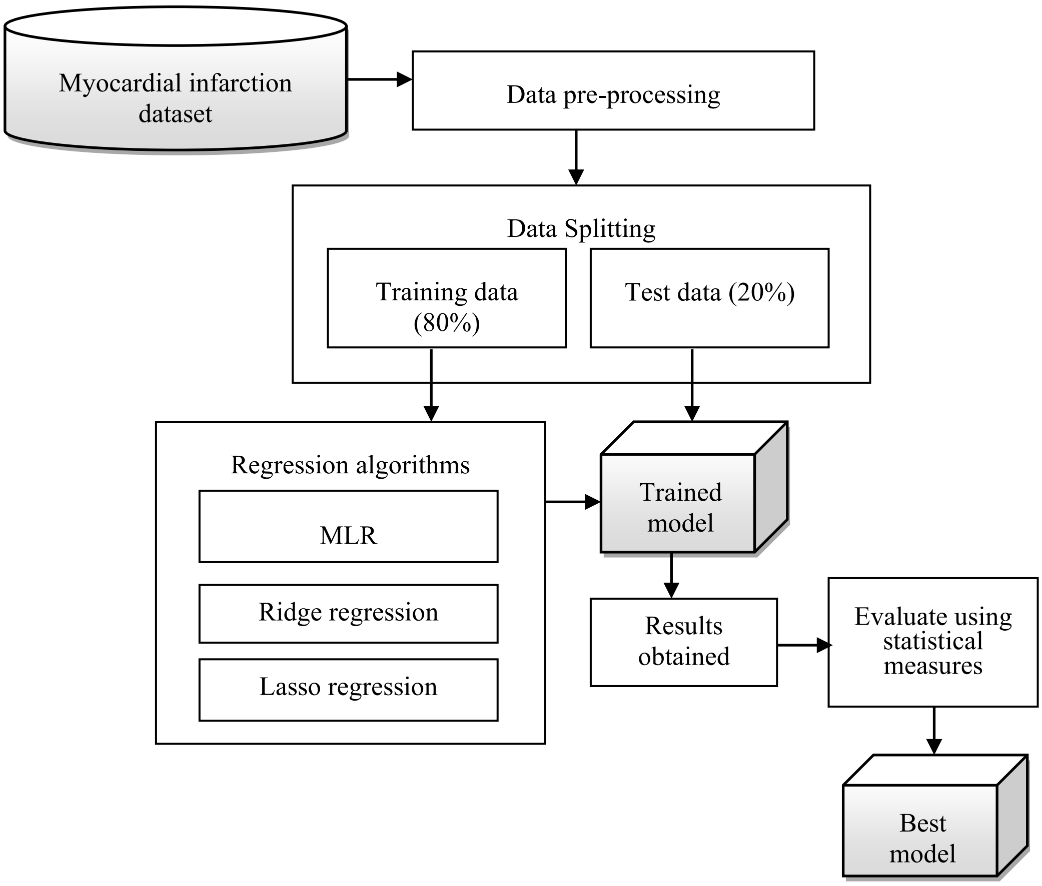

Figure 2. System architecture

4. Proposed work

In this section the algorithms multiple linear regression and ridge regression are explained briefly. The proposed algorithm lasso regression is explained in detail. The value of the target variable is calculated using the below formula[24] in all the three algorithms.

$$\hat{y}_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_p x_{tp}$$

Where p=1,2,3,...,P, P is the number of predictor variables, x_{tp} is the value of predictor variable p in observation t, β_0 is the y intercept value and $\beta_1,\beta_2,\ldots,\beta_P$ are regression coefficients for each predictor variable . Here y_t is actual value of target variable for observation t, $\hat{y}_t$ is predicted value of target variable for observation t. In the following formulae n is number of observations.

4.1. Multiple linear regression

MLR is an extension to the linear regression technique. This algorithm is used to model the linear relationship between several predictor (independent) variables and one target (dependent) variable. In linear regression the value of target variable is estimated by using only one predictor variable. In multiple linear regression two or more predictor variables are used to estimate the value of target variable. The value of target variable is calculated using formula mentioned above. After predicting all the values the error is calculated using cost function which is defined below.

$$CF = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2$$

4.2. Ridge regression

Ridge regression is a regularization of linear regression. This method reduces the complexity of the model by shrinking the regression coefficients using a tuning parameter called lambda. It also reduces the multi collinearity which is said as correlation between predictor variables. The value of the target variable is predicted same as in MLR but the cost function to calculate the error is modified into following formula. The extra term is called the penalty term which is the square of regression coefficients. Here β_k represents regression coefficients for values of k=0,1,2,...P. λ is a tuning parameter. [25]

$$CF = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2 + \lambda \sum_{k=0}^{P} \beta_k{}^2$$

4.3. Lasso regression

Algorithm: Lasso regression

Input: Each observation in the training dataset.
Output: Predicted value of the target variable for given observation.
Assumptions: t is a specific observation, n is number of observations, x_p is a specific predictor variable value where p=1,2,3,...,P, P is number of predictor variables, y is target variable, y_t is value of target variable in observation t, x_{tp} is value of predictor variable p in observation t, $\hat{y}_t$ is predicted value of target variable for observation t, β_k represents regression coefficients for k=0,1,2,...P, λ is a tuning parameter.

Step 1: Start
Step 2: Calculate mean of each predictor variable xp and target variable y.

$$\bar{x}_p = \frac{\sum_{t=1}^{n} x_{tp}}{n}; \quad \bar{y} = \frac{\sum_{t=1}^{n} y_t}{n}$$

Step 3: Calculate regression coefficient of each predictor variable.
Step 3.a: For each predictor variable (p=1,2,3,..,P) compute.

$$\beta_p = \frac{\sum_{t=1}^{n} (x_{tp} - \bar{x}_p)(y_t - \bar{y})}{\sum_{t=1}^{n} (x_{tp} - \bar{x}_p)^2}$$

Step 3.b: End for loop in step 3.a
Step 4: For each observation t= 1,2,3,....,n
Step 4.a: Predict value of target variable in observation t. Here β_0 is the intercept value which is calculated placing all predictor variables equal to 0.

$$\hat{y}_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_p x_{tp}$$

Step 4.b: Return value of target variable obtained in 4.a
Step 5: End for loop in step 4
Step 6: Calculate cost function

$$CF = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2 + \lambda \sum_{k=0}^{P} |\beta_k|$$

Step 7: Stop

Lasso stands for Least Absolute Shrinkage and Selection Operator. It is also a regularization of linear regression. It reduces the model complexity and problem of over fitting by using magnitude of regression coefficients while calculating cost function. The cost function will give the error after predicting values. By using this technique the errors obtained can be reduced when compared to RR. The process of predicting the value is same in all the three algorithms but differs when it comes to cost function. [26]
In step 2 the mean is calculated for each predictor variable and target variable. These values are used for calculating regression coefficient of each predictor variable in step 3. The regression coefficients of predictor variables are used to obtain the predicted value of the target variable in step 4 and return the value in 4.b. The overall error after predicting all the values is calculated using cost function in step 6. Thus, the trained model is obtained which is further evaluated on test dataset.
The figure 3 represents the flowchart of regression algorithms. All the three algorithms works in a similar way, there is a change in only the formulae. The formulae used in each algorithm has already mentioned in this section. So, in this figure only the steps involved in working of algorithm is described through a flowchart.

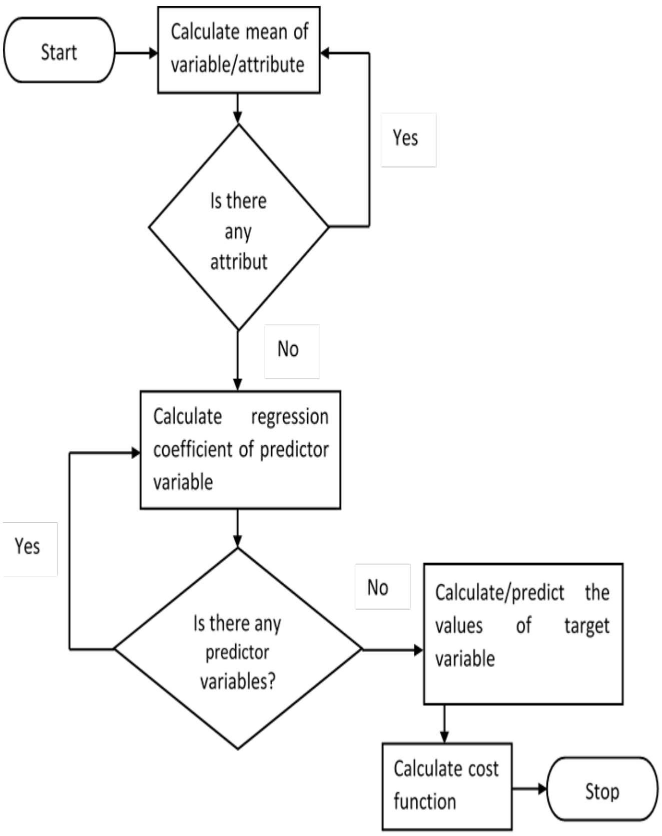

Figure 3. Flowchart of used regression algorithms

5. Result analysis

In this section the results of the three regression algorithms are provided. The algorithms MLR, RR and lasso regression are compared with each other. The uncertainty in predicted values of each algorithm is shown using estimation statistics confidence interval and prediction interval. Comparison of three algorithms is done using statistical measures RMSE and r_squared.

5.1. Statistical measures

R_Squaredvalue

R_Squared is used to know whether the obtained regression line is better than normal horizontal regression line generated through mean of data points. Its value should be between 0 and 1. Its value increases if the errors are less, vice versa. While comparing models the higher value of r_squared value indicates best model. The value of r_squared is explained for lasso regression model below.

$$R^2 = 1 - \frac{\sum_1^n (y_t - \hat{y}_t)^2}{\sum_1^n (y_t - \bar{y}_t)^2}$$

The numerator $\sum_1^n (y_t - \hat{y}_t)^2$ is called as sum of squared errors (SSE). In this equation n is number of instances in test dataset (n=26), y_t is the actual value of test instance i, $\hat{y}_t$ is predicted value for test instance i. The SSE value for lasso regression model is obtained as 4.5436.
The denominator $\sum_1^n (y_t - \bar{y}_t)^2$ is called as sum of squared total (SST). In this equation n is number of instances in test dataset, y_t is the actual value of test instance i, $\bar{y}_t$ is mean of test dataset. The SST value for lasso regression model is obtained as 5.8846.
R_Squared for lasso regression = 1 – (4.5436 / 5.8846) = 0.2278

Root Mean Squared Error (RMSE)

RMSE is defined as root mean square of the errors. In regression RMSE is very important measure to check the performance of the developed model. The lower value of RMSE indicates less errors and better performance. The numerator is called as sum of squared errors (SSE) which is given in r_squared value above. The value of RMSE for lasso regression model is explained below.

$$RMSE = \sqrt{\frac{\sum_1^n (y_t - \hat{y}_t)^2}{n}}$$

RMSE for lasso regression= sqrt (4.5436 / 26) = 0.418

5.2. Estimation statistics

Confidence interval

Confidence interval is an estimation statistic used to give bounds on estimation of population parameter like mean and standard deviation. In regression when 95% confidence interval is used the upper and lower limits of mean are obtained. The estimated value of each observation should lie between these limits.

$$CI = \hat{y}_t \pm T_{(n-2)} \sqrt{MSE\left(\frac{1}{n} + \frac{(y_t - \bar{y})^2}{\sum_{p=1}^n (y_p - \bar{y})^2}\right)}$$

$\hat{y}_t$ is the predicted value of observation t, T_{n-2} is the value in 95% column from t-distribution table and n is the number of observations in test dataset. y_t is actual value of observation t. $\bar{y}$ is mean of the predictor attribute. P is specific observation. The denominator $\sum_{p=1}^n (y_p - \bar{y})^2$ is the sum of squared total (SST). MSE is the mean squared error calculated by formula SSE / (n-q) where q is the number of coefficients in the model. The value of MSE lasso regression was obtained as 0.1747 by using a built-in function.
CI of first instance in test dataset for lasso regression model =

$$0 \pm 2.064 \sqrt{0.1747\left(\frac{1}{26} + \frac{(0-0.3461)^2}{5.884}\right)} = (-0.2299, 0.3687)$$

Prediction interval

Prediction interval is an estimation statistic used to give bounds on estimation of single observation in dataset. As explained in confidence interval 95% prediction interval also gives upper and lower limits. The estimated value of each observation should be in this limit. The description of formula is same as in confidence interval formula.

$$PI = \hat{y}_t \pm T_{(n-2)} \sqrt{MSE\left(1 + \frac{1}{n} + \frac{(y_t - \bar{y})^2}{\sum_{p=1}^n (y_p - \bar{y})^2}\right)}$$

PI of first instance in test dataset for lasso regression model

$$= 0 \pm 2.064 \sqrt{0.1747\left(1 + \frac{1}{26} + \frac{(0 - 0.3461)^2}{5.884}\right)}$$

$$= (-0.8438, 0.9826)$$

5.3. Results obtained

The estimation statistics for predicted values obtained after implementing three algorithms are provided below. The figure 4, figure 5 and figure 6 correspond to MLR, RR and lasso regression respectively. In these figures the left side data is the 95% confidence intervals and right side data is 95% prediction intervals obtained as output for each test instance.
In figure 4 the test instances numbered from 7 to 131 represent the instances selected from original dataset to form the test dataset. In figures 5 and 6 these instances are represented from 1 to 26 because of calculating confidence and prediction intervals by implementing the formulae and storing the results in a data frame. So, the instances were mentioned in a sequence 1 to 26, but the instances are same for all three algorithms only numbering differs. In case of multiple linear regression (figure 4) the estimation statistics are obtained using built-in method.

```
> confidence_int_ml
            fit          lwr          upr
7   -0.52632624 -0.954458661 -0.09819383
15   0.85344781  0.430133964  1.27676165
16   1.03179282  0.541387193  1.52219845
27  -0.21984818 -0.623213235  0.18351687
30  -0.13400754 -0.403315180  0.13530010
33   0.93638540  0.634488296  1.23828251
38   0.56590067  0.180984029  0.95081732
43   0.56789045  0.176073760  0.95970714
52  -0.27034923 -0.664706522  0.12400805
60   0.60371673  0.328940229  0.87849322
63   0.28223655 -0.003937817  0.56841092
77   0.80523753  0.472412234  1.13806282
79   1.15994431  0.676191195  1.64369742
89   0.30749323  0.017040233  0.59794623
91   0.86579470  0.572670451  1.15891894
92  -0.19343425 -0.474530308  0.08766181
93   0.68926285  0.394280226  0.98424547
104 -0.12227811 -0.419488209  0.17493199
110 -0.04953928 -0.483257007  0.38417844
111 -0.09536144 -0.436007657  0.24528477
113 -0.32976633 -0.663866974  0.00433432
120  0.45329654  0.095917688  0.81067540
124  0.24653957 -0.144720907  0.63780005
129  0.22176346 -0.072525402  0.51605232
130  0.45016272  0.164310434  0.73601501
131  1.11410612  0.674627461  1.55358478
>
```

```
> prediction_int_ml
            fit         lwr       upr
7   -0.52632624 -1.33227018 0.2796177
15   0.85344781  0.05005322 1.6568424
16   1.03179282  0.19111099 1.8724747
27  -0.21984818 -1.01291289 0.5732165
30  -0.13400754 -0.86802039 0.6000053
33   0.93638540  0.18979976 1.6829710
38   0.56590067 -0.21794184 1.3497432
43   0.56789045 -0.21936337 1.3551443
52  -0.27034923 -1.05887059 0.5181721
60   0.60371673 -0.13232022 1.3397537
63   0.28223655 -0.45813094 1.0226040
77   0.80523753  0.04561858 1.5648565
79   1.15994431  0.32312573 1.9967629
89   0.30749323 -0.43453858 1.0495250
91   0.86579470  0.12271322 1.6088762
92  -0.19343425 -0.93185368 0.5449852
93   0.68926285 -0.05455367 1.4330794
104 -0.12227811 -0.86698080 0.6224246
110 -0.04953928 -0.85846408 0.7593855
111 -0.09536144 -0.85843949 0.6677166
113 -0.32976633 -1.08994493 0.4304123
120  0.45329654 -0.31739659 1.2239897
124  0.24653957 -0.54043757 1.0335167
129  0.22176346 -0.52177820 0.9653051
130  0.45016272 -0.29008034 1.1904058
131  1.11410612  0.30207795 1.9261343
>
```

Figure 4. Confidence and prediction intervals for multiple linear regression predicted values

```
> confidence_int_ridge
   predictions        lower       upper
1    -0.526028 -0.835817417 -0.21623858
2     0.853318  0.482291739  1.22434426
3     1.031760  0.721970583  1.34154942
4    -0.219756 -0.529545417  0.09003342
5    -0.134100 -0.443889417  0.17568942
6     0.936718  0.565691739  1.30774426
7     0.565896  0.256106583  0.87568542
8     0.568210  0.197183739  0.93923626
9    -0.270974 -0.580763417  0.03881542
10    0.604048  0.233021739  0.97507426
11    0.282310 -0.027479417  0.59209942
12    0.805540  0.495750583  1.11532942
13    1.160240  0.789213739  1.53126626
14    0.307732 -0.002057417  0.61752142
15    0.866210  0.495183739  1.23723626
16   -0.193434 -0.503223417  0.11635542
17    0.689340  0.379550583  0.99912942
18   -0.122470 -0.432259417  0.18731942
19   -0.049608 -0.359397417  0.26018142
20   -0.095532 -0.405321417  0.21425742
21   -0.329500 -0.639289417 -0.01971058
22    0.453306  0.082279739  0.82433226
23    0.247330 -0.062459417  0.55711942
24    0.221656 -0.149370261  0.59268226
25    0.450184  0.140394583  0.75997342
26    1.114927  0.743900739  1.48595326
>
```

```
> prediction_int_ridge
   predictions       lower     upper
1    -0.526028 -1.47117874 0.4191227
2     0.853318 -0.11363646 1.8202725
3     1.031760  0.08660926 1.9769107
4    -0.219756 -1.16490674 0.7253947
5    -0.134100 -1.07925074 0.8110507
6     0.936718 -0.03023646 1.9036725
7     0.565896 -0.37925474 1.5110467
8     0.568210 -0.39874446 1.5351645
9    -0.270974 -1.21612474 0.6741767
10    0.604048 -0.36290646 1.5710025
11    0.282310 -0.66284074 1.2274607
12    0.805540 -0.13961074 1.7506907
13    1.160240  0.19328554 2.1271945
14    0.307732 -0.63741874 1.2528827
15    0.866210 -0.10074446 1.8331645
16   -0.193434 -1.13858474 0.7517167
17    0.689340 -0.25581074 1.6344907
18   -0.122470 -1.06762074 0.8226807
19   -0.049608 -0.99475874 0.8955427
20   -0.095532 -1.04068274 0.8496187
21   -0.329500 -1.27465074 0.6156507
22    0.453306 -0.51364846 1.4202605
23    0.247330 -0.69782074 1.1924807
24    0.221656 -0.74529846 1.1886105
25    0.450184 -0.49496674 1.3953347
26    1.114927  0.14797254 2.0818815
>
```

Figure 5. Confidence and prediction intervals for ridge regression predicted values

```
> confidence_int_lasso
   predictions        lower     upper
1   0.06940969 -0.229934532 0.3687539
2   0.47965313  0.121136797 0.8381695
3   0.50832265  0.208978434 0.8076669
4   0.14524082 -0.154103396 0.4445850
5   0.31164126  0.012297043 0.6109855
6   0.57307728  0.214560945 0.9315936
7   0.48489080  0.185546583 0.7842350
8   0.46303489  0.104518558 0.8215512
9   0.21452286 -0.084821357 0.5138671
10  0.41488114  0.056364804 0.7733975
11  0.39491649  0.095572267 0.6942607
12  0.47590888  0.176564666 0.7752531
13  0.64406797  0.285551639 1.0025843
14  0.44624309  0.146898868 0.7455873
15  0.57062303  0.212106697 0.9291394
16  0.24895210 -0.050392122 0.5482963
17  0.54385930  0.244515084 0.8432035
18  0.36429950  0.064955278 0.6636437
19  0.35306612  0.053721905 0.6524103
20  0.27776604 -0.021578177 0.5771103
21  0.27300817 -0.026336048 0.5723524
22  0.49937618  0.140859843 0.8578925
23  0.29822293 -0.001121285 0.5975672
24  0.37809827  0.019581933 0.7366146
25  0.48063414  0.181289926 0.7799784
26  0.57503029  0.216513960 0.9335466
> |
```

```
> prediction_int_lasso
   predictions      lower     upper
1   0.06940969 -0.8438733 0.9826927
2   0.47965313 -0.4546984 1.4140047
3   0.50832265 -0.4049604 1.4216057
4   0.14524082 -0.7680422 1.0585238
5   0.31164126 -0.6016417 1.2249243
6   0.57307728 -0.3612743 1.5074288
7   0.48489080 -0.4283922 1.3981738
8   0.46303489 -0.4713167 1.3973865
9   0.21452286 -0.6987601 1.1278059
10  0.41488114 -0.5194704 1.3492327
11  0.39491649 -0.5183665 1.3081995
12  0.47590888 -0.4373741 1.3891919
13  0.64406797 -0.2902836 1.5784195
14  0.44624309 -0.4670399 1.3595261
15  0.57062303 -0.3637285 1.5049746
16  0.24895210 -0.6643309 1.1622351
17  0.54385930 -0.3694237 1.4571423
18  0.36429950 -0.5489835 1.2775825
19  0.35306612 -0.5602169 1.2663491
20  0.27776604 -0.6355170 1.1910490
21  0.27300817 -0.6402748 1.1862912
22  0.49937618 -0.4349754 1.4337277
23  0.29822293 -0.6150601 1.2115059
24  0.37809827 -0.5562533 1.3124498
25  0.48063414 -0.4326489 1.3939171
26  0.57503029 -0.3593213 1.5093819
> |
```

Figure 6. Confidence and prediction intervals for lasso regression predicted values

Table 2. Coefficients estimated by algorithms

Predictor attribute	Coefficients		
	Multiple linear regression	**Ridge regression**	**Lasso regression**
(Intercept)	-0.7556	-0.7557	0.02935
Gender	0.1270	0.1270	.
Age	0.0124	0.0124	0.0078
BMI	-0.0042	-0.0043	.
DM_type	-0.0113	-0.0113	.
DM_duration	0.0047	0.0047	.
FBS	0.0011	0.0011	.
HbA1C	-0.0014	-0.0015	.
LDL	-0.0016	-0.0016	.
HDL	-0.0041	-0.0041	.
TG	0.0002	0.0003	.
DM_treat	-0.0121	-0.0121	.
Statin	0.0402	0.0403	.
Dose	0.0005	0.0005	.
Sys_bp	-0.0007	-0.0008	.
Dia_bp	0.0094	0.0095	.
Smoking	0.1129	0.1129	.
Restecg	-0.0315	-0.0316	.
Thalach	0.0025	0.0025	0.0011
Oldpeak	-0.0579	-0.0579	-0.0020
Slope	0.0077	0.0078	.
Thal	-0.2671	-0.2672	-0.0838

The statistical measures RMSE and r_squared that have considered are obtained for each algorithm. The values of RMSE and r_squared obtained for multiple linear regression are 0.4325 and 0.1732, for ridge regression are 0.4326 and 0.1730 and for lasso regression are 0.4180 and 0.2278.

The value of tuning parameter for multiple linear regression and ridge regression are considered by the functions used for implementation lm() and lmridge() respectively. Those values are not visible instead the best value was selected by those functions. For the lasso regression algorithm the function cv.glmnet() was used. The best value of tuning parameter (lambda) was obtained and used as a parameter in cv.glmnet() which is 0.1584893 in this work. The table 2 comprises of estimated coefficients obtained for each predictor attribute by implementing each algorithm.

In this work the attributes thal, age, thalach and old peak from table 1 are identified as important attributes by lasso regression and attributes thal, age, gender, FBS, dia_bp and old peak are identified as important attributes by ridge and multiple linear regression algorithms. The remaining attributes are least important attributes. The attribute thal - Thallium heart scan is identified as the most important attribute.

5.4. Comparing results of three algorithms

In table3 all the results obtained for three algorithms are provided which are statistical measures namely root mean squared error (RMSE) and r_squared. By comparing the results of MLR and RR there is no significant change in the values obtained. The lasso regression has obtain good values of RMSE and r-squared as 0.4180 and 0.2278 respectively.

Table 3. Comparing results of algorithms

Algorithm	RMSE	R_squared
Multiple linear regression	0.4325	0.1732
Ridge regression	0.4326	0.1730
Lasso regression	0.4180	0.2278

Figure 7 shows the comparison graph of root mean squared error (RMSE) value for the three regression algorithms. By observing the graph the value of RMSE obtained for lasso regression (0.4180) is less compared to remaining algorithms. Thus in terms of RMSE lasso regression can be said as the best model.

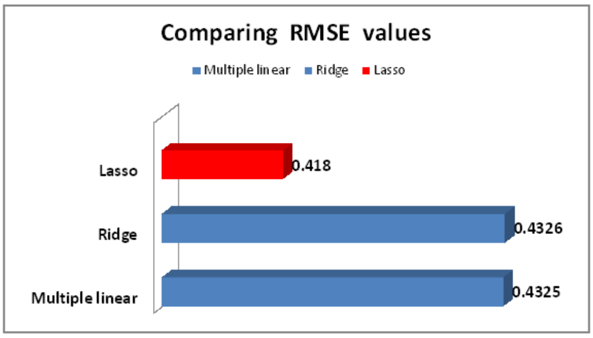

Figure 7. Comparing the algorithms using RMSE value

Figure 8 shows the comparison graph of r-squared value for the three regression algorithms implemented. By observing the graph the value of r_squared obtained for lasso regression (0.2278) is more than other algorithms. Thus in terms of r-squared lasso regression can be said as the best model.

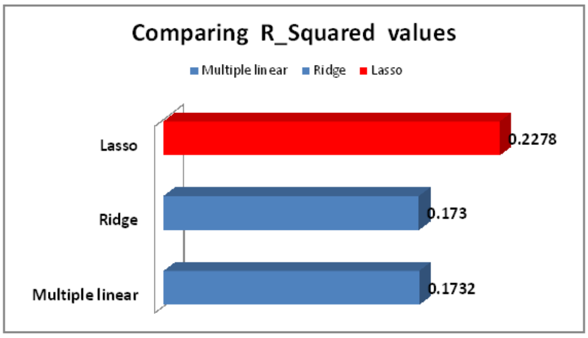

Figure 8. Comparing the algorithms using R_squared value

5.5. Comparing proposed work with literature work

The literature works related to myocardial infarction and heart diseases are provided in the table 4. In some works ([8], [9] and [10]) only ECG data was considered for implementation. To increase the quality of research, attributes related to diabetes, ECG data and risk factors of myocardial infarction are considered in this work. Considering the attributes on the above criteria will increase the scope of prediction. In [11] clustering technique is used along with feature selection technique but the evaluation of the model was not provided. In this work statistical evaluation metrics were considered. In some works ([12], [13], [21] and [22]) few ML classification algorithms are used. These algorithms are the most commonly used algorithms. In this work regression algorithms are considered instead of regularly used algorithms.

Table 4. Comparing literature work with proposed work

Author	Algorithms considered in the work	Findings	Algorithm recognized as better	Values of metrics obtained for the proposed algorithm
Authors of this paper in this work.	Multiple linear regression, ridge regression and lasso regression	Prediction of myocardial infarction is done using regression algorithms. Confidence and prediction intervals are obtained for each algorithm. RMSE and r_squared are the metrics chosen for evaluating and comparing algorithms.	Lasso regression	RMSE - 0.4180 and r_squared-0.2278

Xingjin Zhang et al. [8]	Long Short Term Memory (LSTM)	LSTM algorithm is implemented on an ECG dataset to predict myocardial infarction. The comparison between the proposed work and other works is done by considering overall accuracy. Metrics like sensitivity, specificity and positive predictivity are obtained for each target class but overall values are not provided.	LSTM has obtained highest accuracy than models in other works namely SVM and CNN.	Overall accuracy of LSTM-99.91%
Wenhan Liu et al. [9]	Multiple feature branch convolutional neural networks (MFB-CNN).	Diagnosis of myocardial infarction (MI) is done using MFB-CNN. They considered ECG classification data and patient physiological data for automated MI detection and localization. In both these cases metrics like accuracy, sensitivity and specificity are chosen for MI detection and only accuracy for MI localization.	MFB-CNN performed better than other existing methods in both the cases (ECG classification data and patient physiological data).	ECG data: MI detection-values of accuracy, sensitivity and specificity are 99.95%, 99.97%, 99.9% respectively and MI localization-99.81% overall accuracy is obtained. Patient data: MI detection- Values of accuracy, sensitivity and specificity are 98.79%, 98.73%, 99.35% respectively and MI localization-94.82% accuracy is obtained.
Rajendra et al. [10]	Convolution neural networks (CNN)	Detection of myocardial infarction (MI) using ECG signals is performed. CNN is implemented on data with noise and without noise. The accuracy, positive predicted value (PPV), sensitivity and specificity metrics are chosen for evaluation.	CNN	Data with noise: accuracy-93.53%, PPV-98.03%, sensitivity-93.71% and specificity-92.83%. Data without noise: accuracy-95.22%, PPV-98.43%, sensitivity-95.49% and specificity-94.19%
Umamaheswari and Isakki [11]	OneR & relief feature selection techniques and k-medoid clustering	Prediction of myocardial infarction is performed. 8 among 14 attributes were selected by performing oneR and relief feature selection techniques. Then k-medoid clustering is performed. The clustering plot obtained clearly differentiates the two target classes.	OneR & relief feature selection techniques and k-medoid clustering.	Performance metrics were not considered instead a plot obtained after clustering technique is provided.
Runchuan Li et al. [12]	SVM, logistic regression, decision tree, naive bayes, radial basis function (RBF) and random forest	Machine learning algorithms are performed to predict cardiovascular disease. Sensitivity, specificity, precision and AUC are the metrics considered for evaluating and comparing all algorithms.	Random forest	Sensitivity-0.880, specificity-0.876, precision-0.880 and AUC-0.947
Procheta Nag et al. [13]	C4.5 decision tree and random forest	Developed a system to predict myocardial infarction. 70%, 60% and 55% percentage splits were considered using seed values 1-4. Accuracy, precision, recall and ROC curve metrics are considered for evaluation and comparison.	C4.5 has performed better in case of seed 3. Random forest using seed 3 has performed even better than C4.5 decision tree.	Percentage splits for seed 3: 70% - accuracy, precision, recall and ROC curve values are 96%, 0.94, 1 and 0.99 60% - accuracy, precision, recall and ROC curve values are 96%, 0.97, 0.97 and 0.99 and 55% - accuracy,

				precision, recall and ROC curve values are 95%, 0.95, 0.97 and 0.99.
Alexander Schlemmer et al. [21]	SVM with linear and RBF kernels, k-nearest neighbours with k values as 1, 3, 5, 8 and random forest.	Cardiac disease is predicted using ML algorithms. The leave one out (LOO) test was performed for each algorithm. Matthews correlation coefficient (MCC) is the metric considered for evaluating the model.	Linear SVM	MCC – 0.28
Santhana Krishnan and Geetha [22]	Decision tree and naive bayes	Prediction of heart disease is done using two ML algorithms. Accuracy is the metric considered for evaluating and comparing algorithms.	Decision tree	Accuracy – 91%

The literature works related to MLR are [14], [15] and [16]. In [14] MLR algorithm is used for predicting heart disease but its evaluation was not performed. In [15] and [16] only two attributes were chosen for their work. In [15] they concluded that adding few more predictor attributes will give better results. The works [18] and [19] are comprises of RR. In [18] only 14 attributes were used in the dataset and evaluation of the model was not done. In [19] RR was used for feature selection.

In the works [17] and [23] lasso regression was used. The works [24], [25] and [26] are referred to study about theoretical concept of the three algorithms. In [17] lasso regression algorithm has outperformed OLS and RR. Ordinary Least Squares (OLS) regression will minimize the sum of squares obtained by calculating the difference between actual and predicted values. They have stated that OLS regression has a disadvantage of over fitting. In [23] HbA1C attribute was the significant one to predict type-2 diabetes.

In this present work a 22 attributes dataset was considered and evaluation of the model was also done. This makes the work different from other related literature work and has performed well.

MLR is type of linear regression that deals with two or more predictor attributes and a target attribute. Linear regression has a drawback of over fitting. So, regularization techniques were used to overcome it by penalizing the coefficients. The Ridge and lasso regression are the regularization techniques of linear regression. The lasso regression will perform better in the case where there are few significant attributes but whereas RR will perform better when there are more significant attributes. As there are few significant attributes lasso regression has obtained better results than remaining two.

There is a limitation in this work. The dataset used does not contain more instances, it has only 133 instances. Choosing correct predictor attributes is more important to develop an efficient model. Though instances were less the 22 attributes considered has covered all the required data for accurate prediction of myocardial infarction. Finally it is recommend that a large dataset with more instances could be taken to obtain even more efficient model in further works. Also to consider lasso regression as it has performed better than remaining two algorithms.

6. Conclusion

Myocardial infarction is also known as heart attack which is a type of cardiovascular disease. It is one of the complications of diabetes and its prediction is of utmost importance. The regression algorithms namely multiple linear regression, ridge regression and lasso regression are used in this work to predict myocardial infarction for type-1 and type-2 diabetic patients. The algorithms are implemented in R programming. After rigorous analysis of the considered algorithms it is found that the proposed 'lasso regression algorithm' has performed better in predicting Myocardial infarction in terms of statistical measures RMSE and r_squared. The lasso regression obtained values of RMSE and r_squared as 0.418 and 0.2278 respectively. The attribute "Thallium heart scan" is identified as the most important attribute. By considering the statistical measures lasso regression was suggested to predict myocardial infarction in diabetic patients.

References

1. Stevens RJ, Coleman RL, Adler AI, Stratton IM, Matthews DR, Holman RR. Risk factors for myocardial infarction case fatality and stroke case fatality in type 2 diabetes. Diabetes Care. 2004; 27(1):201-207.
2. Types of Heart attacks [online]. Healthline; [cited 2020 Aug 31]. Available from: https://www.healthline.com/health/heart-disease.
3. Acute coronary syndrome [online]. Medicalnews today; [cited 2020 Aug 31]. Available from: https://www.medicalnewstoday.com/articles/315332.
4. Heart attack [online]. Mayo Clinic; [cited 2020 Aug 31]. Available from: https://www.mayoclinic.org/diseases-conditions/heart-attack/symptoms-causes/syc-20373106.
5. Cardiovascular Disease and Diabetes [online]. American Heart Association; [cited 2020 Aug 31]. Available from: https://www.heart.org/en/health-topics/diabetes/why-diabetes-matters/cardiovascular-disease--diabetes.

6. Einarson TR, Acs A, Ludwig C, Panton UH. Prevalence of cardiovascular disease in type 2 diabetes: a systematic literature review of scientific evidence from across the world in 2007-2017. Cardiovasc Diabetol. 2018;17(1):83.
7. Acute Myocardial Infarction [online]. Healthline ; [cited 2020 Aug 31]. Available from:https://www.healthline.com/health/acute-myocardial-infarction.
8. Zhang X, Li R, Hu Q, Zhou B, Wang Z. A New Automatic Approach to Distinguish Myocardial Infarction Based on LSTM. In: Proceedings of 8th International Symposium on Next Generation Electronics (ISNE); 9Oct 2019; Zhengzhou, China. IEEE;2019,p. 1-3.
9. Liu W, Huang Q, Chang S, Wang H, He J. Multiple-feature-branch convolutional neural network for myocardial infarction diagnosis using electrocardiogram. Biomedical Signal Processing and Control. 2018; 45:22-32.
10. Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Information Sciences. 2017; 415:190-198.
11. Umamaheswari M, Devi PI. Prediction of myocardial infarction using K-Medio clustering algorithm. In: Proceedings of IEEE international conference on intelligent techniques in control, optimization and signal processing (INCOS); 2017 Mar 23;Srivilliputhur, India. IEEE; 2017, p. 1-6.
12. Li R, Shen S, Zhang X, Li R, Wang S, Zhou B, Wang Z. Cardiovascular Disease Risk Prediction Based on Random Forest. In: Proceedings of The International Conference on Healthcare Science and Engineering; 2018 Sep 10; Singapore. Springer; 2019, p. 31-43.
13. Nag P, Mondal S, Ahmed F, More A, Raihan M. A simple acute myocardial infarction (Heart Attack) prediction system using clinical data and data mining techniques. In: Proceedings of 20th International Conference of Computer and Information Technology (ICCIT); 2017 Dec 22; Dhaka, Bangladesh.IEEE; 2018,p. 1-6.
14. Polaraju K, Prasad DD. Prediction of heart disease using multiple linear regression model. International Journal of Engineering Development and Research Development. 2017; 5(4):1419-25.
15. Adwani N. Using Multivariate Linear Regression to estimate the probability of having a Heart Attack.2019: 8(11): 112-114.
16. Madhubala V, Porkodi P, Selvapriya R, Tamilzhchelvi P. Diabetics Prediction Based on Multi-Linear Regression Using R Language.Asian Journal of Computer Science and Technology.2019; 8: 17-19.
17. Muthukrishnan R, Rohini R. LASSO: A feature selection technique in predictive modeling for machine learning. In: Proceedings of IEEE international conference on advances in computer applications (ICACA); 2016 Oct 24; Coimbatore, India. IEEE;2017, p. 18-20.
18. Jeena RS, SukeshKumar A. Stroke Risk Assessment Using Ridge Regression Model. In: Proceedings of 2nd International Conference on Trends in Electronics and Informatics (ICOEI); 2018 May 11; Tirunelveli, India. IEEE; 2018,p. 1-3.
19. Luo H, Liu Y. A prediction method based on improved ridge regression. In: Proceedings of 8th IEEE International Conference on Software Engineering and Service Science (ICSESS); 2017 Nov 24; Beijing, China. IEEE; 2018, p. 596-599.
20. Golande A, Kumar TP. Heart disease prediction using effective machine learning techniques. International Journal of Recent Technology and Engineering (IJRTE). 2019 Jun; 8(1):944-950.
21. Schlemmer A, Zwirnmann H, Zabel M, Parlitz U, Luther S. Evaluation of machine learning methods for the long-term prediction of cardiac diseases. In: Proceedings of 8th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO); 2014 May 25; Trento, Italy. IEEE; 2014, p. 157-158.
22. Santhana KJ, and GeethaS. Prediction of Heart Disease Using Machine Learning Algorithms. In: Proceedings of 1st International Conference on Innovations in Information and Communication Technology (ICIICT); 25-26 April 2019; CHENNAI, India. IEEE; 2019, p. 1-5
23. Farbahari A, Dehesh T, Gozashti MH. The Usage of Lasso, Ridge, and Linear Regression to Explore the Most Influential Metabolic Variables that Affect Fasting Blood Sugar in Type 2 Diabetes Patients. Romanian Journal of Diabetes Nutrition and Metabolic Diseases. 2019 Dec 1;26(4):371-9.
24. Zhang J, Li Y, Cao J. Sensor situation based on the multiple linear regression forecast. In: Proceedings of IEEE International Conference on Computer Science and Automation Engineering; 2011 Jun 10; Shanghai, China. IEEE;2011, p. 47-50.
25. Boonyakunakorn P, Nunti C, Yamaka W. Forecasting of Thailand's Rice Exports Price: Based on Ridge and Lasso Regression. In: Proceedings of the 2nd International Conference on Big Data Technologies; 2019 Aug 28; Jinan, China. ACM; 2019, p. 354-357.
26. Musoro JZ, Zwinderman AH, Puhan MA, ter Riet G, Geskus RB. Validation of prediction models based on lasso regression with multiply imputed data. BMC medical research methodology. 2014 Dec 1;14(1):116.