

CARDIOVASCULAR DISEASE PREDICTION USING CLASSIFIER ALGORITHM

Ankur Sharma¹, Neha Arora²

¹Student, ²Assistant Professor

Department of Computer Science and Engineering

IILM College of Engineering and Technology, Greater Noida, U.P, India

Abstract: Machine learning has introduced new and easiest ways of solving problems in data science. One common application of machine learning is the prediction of an outcome based upon existing data. The machine learns the data in the form of patterns and then applies them in order to predict an outcome. This paper searches a method termed ensemble classification, which is used for improving the accuracy of weak algorithm. R programming is used in order to visualise and analyse the data and to make prediction on the outcome. In this paper, research is done to apply the properties of machine learning in healthcare sector.

Cardiovascular Disease or Heart Disease is one of the most common and deadliest diseases now-a-days. There are various parameters that results in such kind of disease. In today's world, data is the major entity which defines one's characteristics/identity whether it's a name, age etc. Similarly, in healthcare doctor diagnose the patients and keep a track of their record in a form of datasets. This vast data can be sort into the necessary data which is required to a doctor.

Here comes the role of Big data analytics, it helps the user to visualize the data which is relevant to use and sort it down. Then the data can be analysed using classifiers. And as we all know Technology is booming day-by-day, which is resulting not only in making the work easier but also effective at the same time.

This paper is focused onto the situation of predicting the rate of heart attack in a patient by visualizing and analysing their dataset. The data plays an important role in analytics, if the data is not refined and has flaws in it the accuracy rate will be much lower. This idea is based on providing a structured base to the doctors to predict the chances of CVD in a patient.

By examining the social determinants of health for instance, smoking, education etc. we (doctors) would be able to predict more accurate results. The visualization technique will help the doctors to distinguish between the factors more easily and effectively, it will also help in determining the prevail and the least measure/factor for the variation in the data. As we know Big data is known for its valuable use of data controlling, handling, contrasting, managing the large dataset which gives more accurate results through analysing the whole data. The prediction rate depends upon the structure of the data i.e. if the data is unstructured then there is possibility of least accurate results.

This paper illustrates the way how the Big data can be used in a significant way to predict the rate of heart attack in people through analysing their daily routine or habits and their way of understanding. This study describes the use of Framingham dataset study in which the parameters such as CurrentSmoker, Blood pressure, Diabetes, Cholesterol etc. are measured to fetch out the most possible accurate results. Framingham heart study allows us to understand the factors responsible for such heart disease.

Researches are still going on for this topic to create a possible accurate model to predict CVD in people.

Keywords: CVD, packages, missmap, age, TenYearCHD etc.

1. Introduction:

One of the most eminent diseases that affect many people during middle or old age is Cardiovascular disease or Heart disease, and in many cases, it eventually leads to fatal complications and results if no actions have been taken in response to it. According to the data of WHO, an estimated 17.9 million people died from CVD in 2016, representing 31% of all global deaths. And there were about 116 million cases of Hypertension which escalates the CVD risk. Around 17 million people die due to cardiovascular disease (CVD) every year worldwide, and the disease is highly prevalent in Asia. The Framingham heart study helps in understanding the risk factors which lead to CVDs. The Cleveland data is also a de facto database for heart disease research.

Age, sex/gender, smoking, Current Smoker, Cholesterol, Blood pressure, Sugar/Diabetes etc. are the risk factors for heart disease. Some risk factors are controllable. Apart from the above factors, lifestyle habits such as eating habits, physical inactivity, and obesity are also considered to be the major risk factors. There are different types of Heart diseases for instance Coronary heart disease, congestive heart failure etc. It is a difficult task to manually examine each factor of risk and to determine the probability of getting the kind of heart disease in a person. However, machine learning is advancing day-by-day which help in managing the big data and try to use it in a meaningful way. Machine learning techniques are useful to predict the output from existing data. Hence, this paper illustrates the use of Classification algorithm to predict the rate of Heart attack in patients. Classification algorithm is a

machine learning algorithm which is used to analyse the data and database. It helps in analysing and making predictions in Big data. Big or vast data sources are difficult to handle manually.

2. Literature Review:

Machine learning is useful in handling Big data and datasets. As the healthcare sector consists large sets of data which are difficult to handle manually, it is as the application area of data mining. One of the reasons for casualty due to heart disease is since the risks or symptoms are either not identified, or they are identified at only the later stages. However, machine learning techniques and algorithms can be useful to beat this problem and to predict the risk of heart attack at early stage so that the patients can be cured accordingly in advance. Some of the techniques for these kinds of predictions are the Support Vector Machines (SVM), Naïve Bayes classifiers, Decision trees, Regression.

Age group, Smoking, Diabetes, blood pressure these are considered as the risk factors.

Analytical studies have shown that techniques such as Neural Networks, Decision tree and Naïve Bayes classifiers can be used for prediction of heart attack in patients. These are called Associative Classifier. We prefer associative classifier over Traditional classifiers as because Associative classifiers provide high flexibility and accuracy. Traditional classifiers lack this ability. Associative classifiers are useful even in handling unstructured data.

Decision tree algorithm is the simple and accurate classifier. Naïve Bayes is considered as the best algorithm for such kinds of prediction followed by Decision tree and Neural Network. In case of Artificial Neural Networks, supervised networks have been used for diagnosis which are trained through Back Propagation Algorithm. The experimental results have shown satisfactory accuracy.

The literature research is done using e-books, journals and articles. It includes the stats of CVD patient i.e. the rate of CVD deaths, the causes of CVD's deaths, the symptoms of CVD, the precautions that can be taken for this disease, how can we use these stats to predict the rate of death. The model reads the data and visualise the data for a clear pattern which helps to understand the data more easily. The literature research also include the study of R language in Big data to handle the vast data and to manage it to have a clear perception about the dataset and to analyse the data using some classifier to gain an expected output. R programming language helps in analysing the data and make predictions on the given dataset. R language helps in applying various operations right through the data and to visualise the change per step which enable the user to maintain a pace in handling the data.

Though various methods were used for predicting heart disease risks with good accuracy in state-of-the-art research, some classification algorithms identify heart disease risk with poor accuracy. Most of the state-of-art research that produces high accuracy employs a hybrid method which include classification algorithms. This study tries on improving the weakness of the classification algorithm. The results are a measure to indicate how these classifiers can effectively be used in the medical field.

3. Dimensionality Reduction:

Dimensionality reduction is an important step in representing the data. It involves selecting non-mathematical representation such that one can relate the majority of, but not all, the variance within the given data, thereby including only most significant information. The data considered for the task may include number of attributes, but not all of these attributes may equally influence the output. Many attributes, or features, may affect the computational complexity and may even lead to overfitting which leads to poor results.

Dimensionality Reduction is generally achieved by two methods – **Feature Extraction** and **Feature Selection**.

3.1 Feature Extraction:

In this, a subset of features is derived from the original feature set. Feature extraction involves a transformation of the features. This transformation is often not reversible as few, maybe many, useful information is lost in the process. Principal Component Analysis (PCA) is used for feature extraction. Principal Component Analysis is a popularly used linear transformation algorithm. In the feature space, it finds the directions that maximize variance and finds directions that are mutually orthogonal. It is a global algorithm that gives the best reconstruction.

3.2 Feature Selection:

In this, a subset of original feature set is selected. The key features are selected by CFS (Correlation based Feature Selection) subset evaluation combined with Best First Search method to reduce dimensionality.

Data Sorting:

The first and foremost step in the data handling is to make it easier to read and to make it work. That's where we talk about the property of dimensionality reduction. The dataset contains vast data among which not all of it is necessary and relevant. To reduce the complexity of the dataset and to make it easier to work upon we apply the reduction property.

The available dataset is composed of several attributes and entries among which most of the data is unnecessary and is of no need. Therefore, the data has to be sorted and filtered. Through using R programming language and packages, the data is being sorted.

Here, use of `missmap()` function is used. This function is used to track for the missing data and irrelevant entries in the data so in this case the irrelevant entries are observed, and the data is checked for the total available data or values. In Fig. 1, the `missmap()` function is used and the missing data is represented in Red colour while the observed or available data is represented in Blue colour. From the graph, there is total of 19% data is missing and 81% data is observed. But here we need only the observed data which can be used to plot the visualisations and graphs.

Here comes the use of `complete ()` function, this function is used to filter the data and to remove the missing values or entries from the data. This function eliminates all the missing data and provide the observed data. This helps in deleting the useless data which gives the data on which operations can be performed. This data is then computed, and all the entries are checked if they are valid or not.

To visualise the data, it is important that the data should be filtered and sorted. The graphical representation will be much clear and accurate if the data is clean. The visualisation include each attribute and it shows the dependency of one attribute onto another.

Clean data is data that is free of errors or anomalies which may make it hard to use or analyse the data. Starting from a clean dataset allows you to focus on creating an effective visualization rather than trying to diagnose and fix issues while creating visualizations.

Data cleaning include following steps –

- Develop a Data quality plan. Set the expectations for the available data i.e. what is required and what is needed.
- Standardized the contact data at the point of entry.
- Validate the accuracy of the data. Validate the data accuracy in the real-time.
- Identify duplicates. Duplicate data creates the situation of redundancy. Redundant data makes the data more complex which makes the data difficult to understand.
- Append the data. Finally append the data and apply the operations that are required.

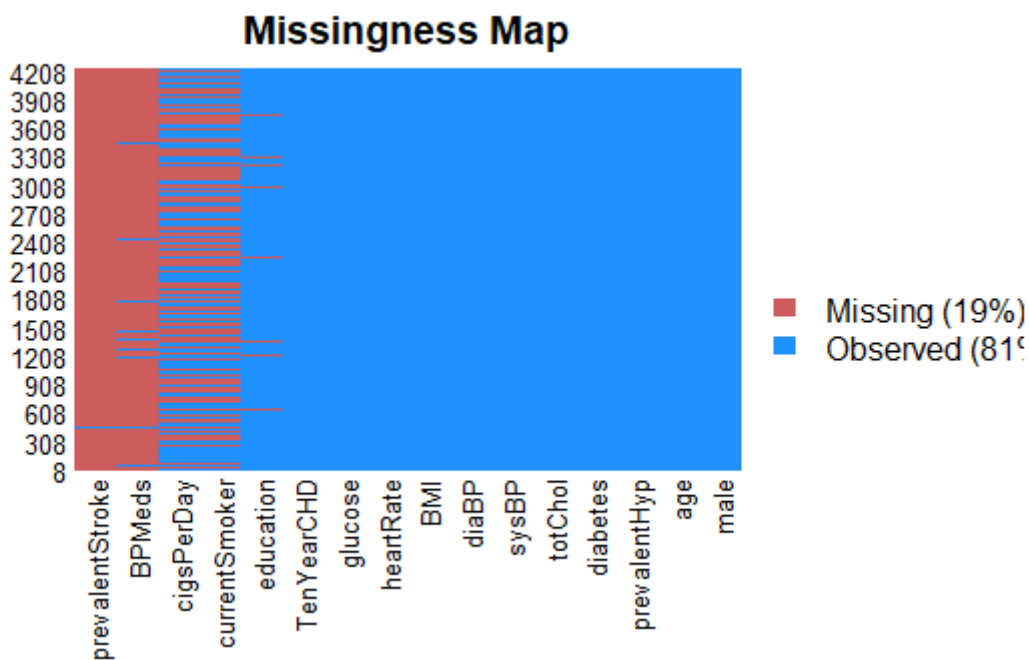


Fig. 1 (Reading the missing data and available data)

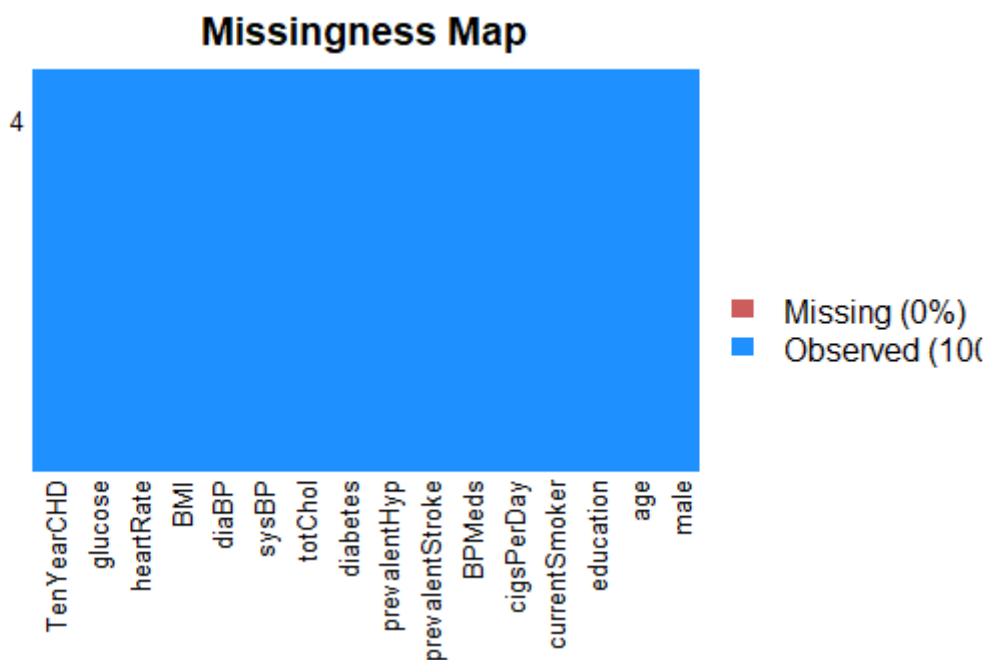


Fig. 2 (Removal of missing data)

4. **Proposed Model:**

It could be utilized to give better access to human services to those living in provincial territories, or to empower elderly individuals to live freely at home for more. Basically, it can enhance access to social insurance assets while decreasing strain on human services frameworks. This is especially valid for cardiovascular sickness or disease; its decent variety of powerful intercessions are conceivably befuddling. Medicinal data got from patients must be put away safely for proceeded with utilize. Specialist’s advantage from knowing a patient’s medicinal history, and machine learning isn’t successful unless substantial databases of data are accessible to it.

The dataset used in this study is based upon the Framingham Heart study. This database contains 16 attributes which are filtered attributes which are only to be used in the experiment. In the first stage of the cleaning process we developed customized procedures to remove noises from the data and prepare it for mining. Additionally, a synthetic dataset was developed for preliminary evaluation for the model. Each attribute plays an important role in establishing the model. Each attribute is correlated to each other which

shows the dependency of data. This dependency helps in visualising the data more accurately and to analyse the data which can give results more accurately.

The value of attributes is given in the table:

Table (1) attributes of dataset

male	age	education	currentSmoker	cigsPerDay	BP Meds	prevalentStroke	prevalentHyp	diabetes	totalChol	sysBP	diaBP	BMI	heartRate	glucose	TenYearCHD
1	39	4	0	0	0	0	0	0	195	106	70	26.97	80	77	0
0	46	2	0	0	0	0	0	0	250	112	81	28.73	95	76	0
1	48	1	1	20	0	0	0	0	245	7.5	80	25.34	75	70	0
0	61	3	1	30	0	0	1	0	225	15	95	28.58	65	103	1
0	46	3	1	23	0	0	0	0	285	13	84	23.1	85	85	0
0	43	2	0	0	0	0	1	0	228	18	0	30.3	77	99	0

4.1 Description of the Dataset:

The dataset is simply consisted of the attributes responsible and useful in making prediction. Patients from age 20 to 70 have been selected in this dataset. Male patients are denoted by a gender value 1 and female patients are denoted by a gender value 0. Education plays an important role in CVDs, as the person is aware about the disease and the symptoms it would be much easier and fast to take best possible actions.

The education level in the dataset is indicated as 1 for secondary standard, 2 for higher standard, 3 for graduates and 4 for PG or adults.

It takes the case as if the person is a current smoker or not i.e. if a person smokes the entry will be 1 and if not, the entry will be 0. Then how many cigarettes a person smoke daily. Blood pressure medication if exists. Prevalent Stroke i.e. if a person has had a stroke in the past.

If the person is diabetic or not. If a person has diabetes, then the dataset will show the value 1 if a person does not have diabetes then the dataset will show the value 0.

The total cholesterol in a person's body will be accounted for pressure on heart. As we know that cholesterol directly affect the heart of a person if the cholesterol is high then the chances of person getting a heart attack are high.

Blood Pressure flow should be subsequent in a person's body. The reading of Systolic Blood Pressure and Diastolic Blood pressure gives proper result and situation of a person's heart.

Body mass Index is also accounted in the dataset. Glucose reading is also present.

Ten Year Coronary Heart Disease record is counted if any. If a person has had any CHD then it will be entered as 1 otherwise 0.

Table (2) Attributes description

S.No.	Attribute Name	Description
1	Male/Gender	1 – Male, 0 - Female
2	Age	20 to 70 years
3	Education	1 – Secondary level, 2 – Higher Secondary, 3 – Graduates, 4 – Adults
4	currentSmoker	Person who is a current smoker
5	cigsPerDay	Number of cigarettes per day
6	BPmeds	Blood pressure medicines
7	prevalentStroke	Person having the condition of prevalent stroke
8	prevalentHyp	Person having condition of prevalent hyperness
9	Diabetes	Diabetes
10	totChol	Total Cholesterol
11	sysBP	Systolic Blood Pressure
12	diaBP	Diastolic Blood Pressure
13	BMI	Body Mass Index
14	heartrate	Current heart beating rate

15	Glucose	Glucose level
16	TenYearCHD	Ten years Coronary Heart Disease record

4.2 Classification and ensemble algorithms

Classification is a supervised learning procedure that is used for predicting the outcome from existing data. This paper proposes an approach for the diagnosis of heart disease using classification algorithms, and to improve the classification accuracy using an ensemble of classifiers. The dataset has been divided into a training set and a test set, and individual classifiers are trained using the training dataset. The efficiency of the classifiers is tested with the test dataset. The working of the individual classifiers is explained in the next section.

4.2.1. Bayesian Network

The Bayesian network is a graphical prediction model based on probability theory. Bayesian networks are built from probabilistic distributions, and they utilize the laws of probability for prediction and diagnosis. Bayesian networks support both discrete and continuous variables. The network is represented as a set of variables whose conditional dependencies are described using acyclic directed graphs. In a Bayesian network, edges between the nodes represent dependent features, whereas nodes that are not connected are conditionally independent. Let X be an evidence that is dependent on n attributes $X = \{A_1, A_2, \dots, A_n\}$. Let H be a hypothesis that the evidence belongs to class C . The probability of hypothesis H , given the evidence X is represented as $P(H|X)$. $P(X|H)$ is the posterior probability of X conditioned on H . The posterior can be calculated using the Bayes theorem as shown in the following equation –

$$P(H|X) = P(X|H) P(H) / P(X)$$

Where $P(H)$ is the probability of the hypothesis being True. $P(X)$ is the probability of the evidence. $P(X|H)$ is the probability of the evidence given that hypothesis is True and $P(H|X)$ is the probability of the hypothesis given that the evidence is present.

4.2.2. Naïve Bayes

The Naïve Bayes classifier or simply, the Bayesian classifier, is based on the Bayes theorem, it is a special case of the Bayesian network, and it is a probability-based classifier. In the Naïve Bayes network, all features are conditionally independent. The changes in one feature therefore does not affect another feature. The Naïve Bayes algorithm is suitable for classifying high dimensional datasets. The classifier algorithm uses conditional independence. Conditional independence assumes that an attribute value is independent of the values of the other attributes in a class.

Let D be a set of training data and associated class labels. Each tuple in the dataset is defined with n attributes that are represented by $X = \{A_1, A_2, \dots, A_n\}$. Let there be m classes represented by C_1, C_2, \dots, C_m . For a given tuple X , the classifier predicts that X belongs to the class having the highest posterior probability, conditioned on X . The Naïve Bayes classifier predicts that the tuple X belongs to the class C_i if and only if

$$P(C_i|X) > P(C_j|X) \quad \text{for } 1 \leq j \leq m, j \neq i$$

Thus, $P(C_i|X)$ is maximized. The class C_i for which $P(C_i|X)$ is maximized is called the maximum posteriori hypothesis. According to Bayes' theorem,

$$P(C_i|X) = P(X|C_i) P(C_i) / P(X)$$

If the attribute values are conditionally independent of one another,

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i)$$

Where x_k refers to the value of attribute A_k for tuple X .

If A_k is categorical, then $P(x_k|C_i)$ is the number of tuples of class C_i in D having the value x_k for A_k , divided by $|C_i, D|$, the number of tuples of class C_i in D . The classifier predicts the class label of X is the class C_i if and only if,

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \text{ for } 1 \leq j \leq m, j \neq i$$

Bayesian classifiers are effective in the sense that they have the minimum error rate for classification.

4.2.3 Uses

Using medical profiles such as age, sex, blood pressure and blood sugar, chest pain, ECG graph etc.

1. It can predict the likelihood of patients getting a heart disease.

- 2. It will be implemented in PYTHON as an application which takes medical test’s parameter as an input.
- 3. It can be used as a training tool to train nurses and medical students to diagnose patients with heart disease.

Visualisation for Dimensionality Reduction:

As discussed, earlier Dimensionality Reduction plays an important role in making the data clear and classified. Visualisation is a technique which is used to look over the data more clearly and in a sorted form. We need data visualisation because a visual summary of information makes it easier to identify patterns and trends than looking through thousands of rows on a spreadsheet. It’s the way the human brain works... Charts and graphs make communicating data findings easier even if you can identify the patterns without them. Since the purpose of data analysis is to gain insights, data is much more valuable when it is visualised. Even if a data analyst can pull insights from data without visualisation, it will be more difficult to communicate the meaning without visualisation.

Without visual representation of the insights, it can be hard for the audience to grasp the true meaning of the findings. Visualisation is carried out in different ways -

- Determining Relationships (Correlations)
- Line Chart
- Bar Chart
- Histogram

Correlation is used to describe the linear relationship between two or more than two continuous variables (for instance - height and weight). In general, correlation tends to be used when there is no identified response variable. It measures the strength and direction of the linear relationship between two or more variables.

A correlation can range in value between -1 and 1.

It can also be shown in different aspects as per the requirement of the situation. For instance, in our case the dataset includes different types of attribute which are correlated to each other in different ways.

The correlation graph for the available dataset can be shown as –

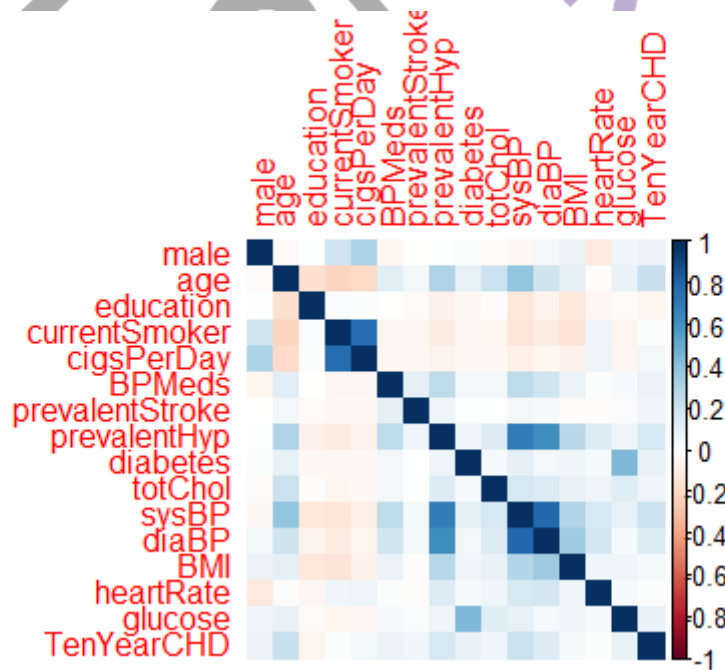


Fig. 3 (Correlation graph of given attributes in block form)

This same graph can be represented in different form such as in pie form –

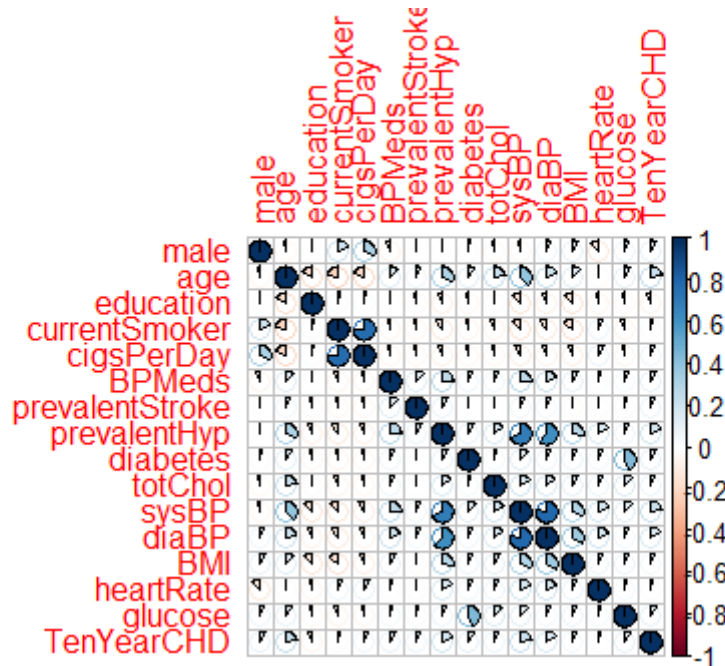


Fig. 4 (Correlation graph in pie form)

The correlation graphs represented above are the graphs representing the correlation inbetween the present attributes. The data can be read and presented in both the forms. This correlation graph shows the dependency of elements on one another. The prediction is based on the relativeness of these attributes which provide best possible accurate results. The representation shows the contribution of each element in the prediction i.e. how the attributes affect the heart disease. The relevancy of attributes must be accurate.

Visualisations:

1. The age distribution in the database can be numbered theoretically but can be understand more effectively through representation or visualisation. The number of patients of different ages in the dataset can be represented as –

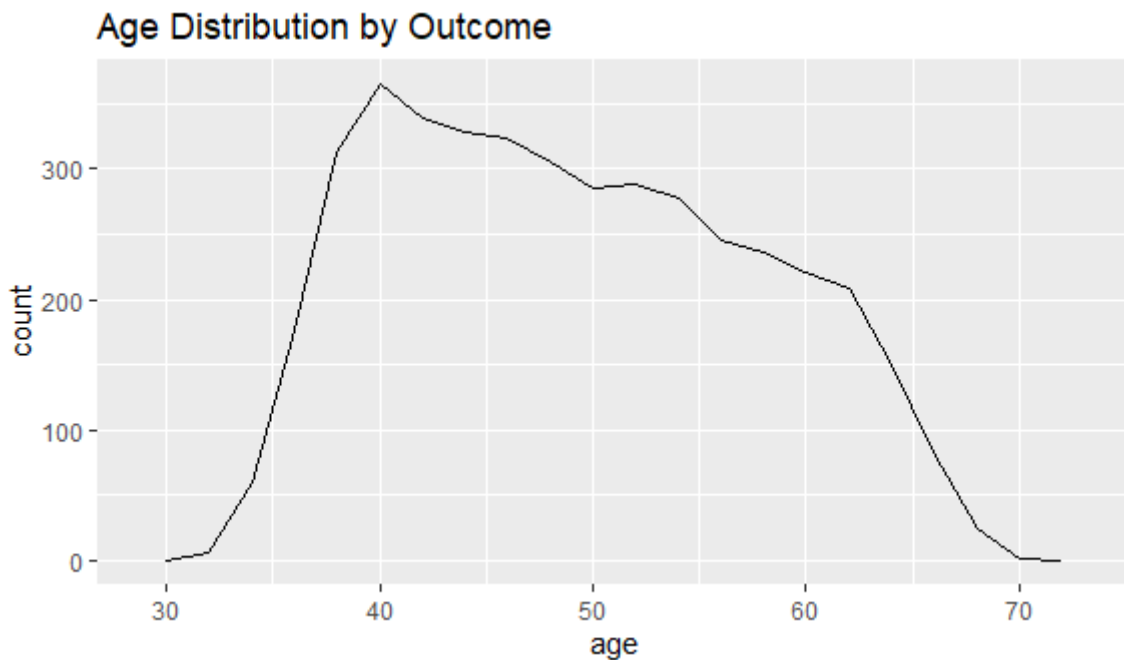


Fig. 5 (Age distribution representation in the database)

Representation gives a proper and overall idea about the data taken in the database. The graph is representing the age distribution of the patients in the database. We can notice the age distribution on the horizontal axis and the number of counts on the vertical axis.

The graph includes the age group of about 20 to 70 years. From which most of the people are in between 40-50.

2. A visualisation can be created to represent the number of males and females in the dataset. It can be represented as –

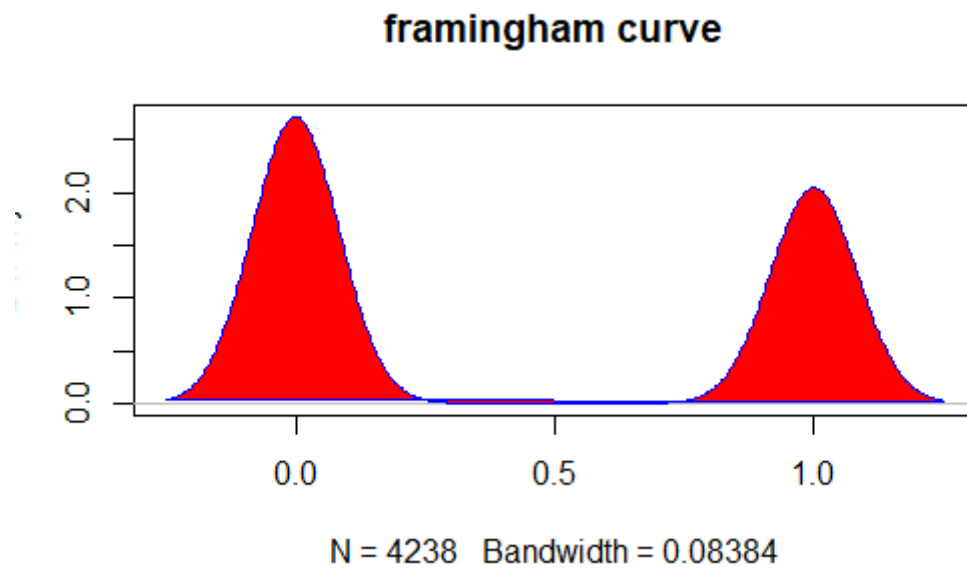


Fig. 6 (Framingham curve to represent the ratio of males and females)

The above graph is showing the ratio of males and females in the data. Framingham curve is used to represent the number of patient's type. In the graph, the **0.0** is showing the Female ratio and **1.0** is showing the male ratio and here we can see that the number of female patients is more than the number of male patients.

Reason: The main reasons of more female patients are –

- Women have risk factors men don't have: There are certain diseases that only affect women, such as endometriosis, ovary disease, diabetes and high blood pressure that develop during pregnancy, increase the risk of coronary heart disease.
- CHD (Coronary heart disease) in women is sometimes hard to diagnose. CAD in women often affects small arteries, which cannot be clearly seen.
- Women are at greater risk of developing a blood clot that can cause another heart attack.

3. The affect of Body Mass Index on the risk of CVD is severe, sometimes. Higher BMI results in high BP, it increases the risk of obesity in a person which directly affect person's heart. As the age increases, the risk of obesity and BMI also increase which leads to CVD.

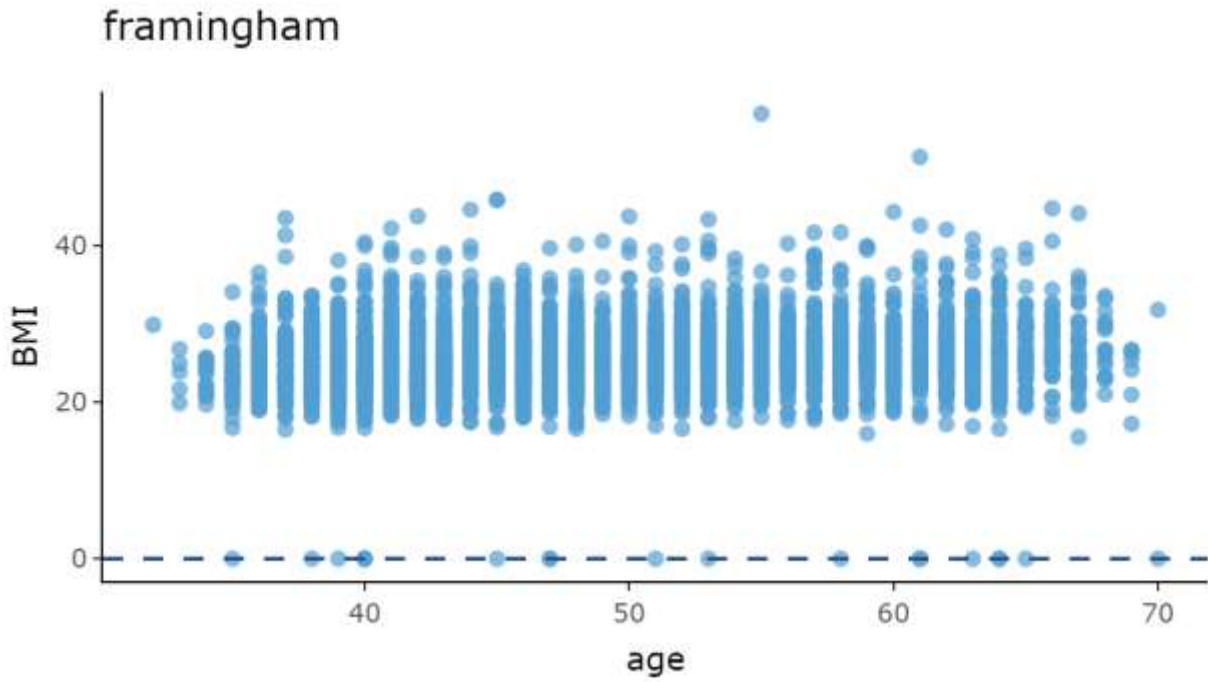


Fig. 7 (Framingham curve to represent rate of BMI vs age)

The above graph is representing the rate of BMI vs age i.e. when the age is increasing BMI is also ascending with it. The maximum BMI can be observed in between 40-50 years of age. After 60 years of age BMI generally remains constant. Overweight individuals (BMI 25-29.9 kg/m²) also have an approximately 60% higher risk of CVD.

These are the local observations from the database in which we can see the effect of attributes on the heart and each other.

4. The effect of diabetes on hypertension is direct and specific i.e. if a person has diabetes there are certain chances of hypertension. Diabetic patients get hypertense easily than that of normal people which affects heart directly. The relative graph of diabetes vs hypertension can be shown as –

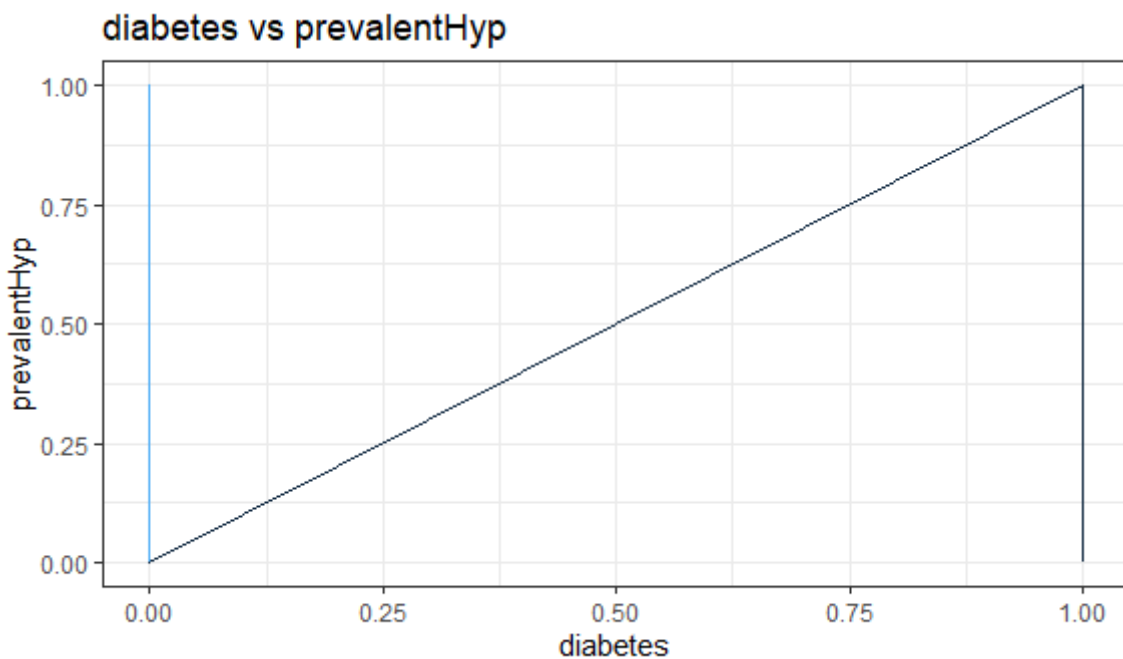


Fig. 8 (Graphical representation for diabetes and prevalent hypertension)

The above graph illustrates the direct effect of diabetes onto prevalent hypertension. The fluid in the body increases, especially if diabetes is already affecting kidneys. Insulin resistance may involve processes that increase the risk of hypertension. From the graph

we can depict that as the diabetes is increased to 1 unit the prevalent hypertension also responds to that by ascending to 1 unit, that is why, the graph is showing a straight line.

5. The Body Mass Index effects the heart rate of a person. As discussed earlier, the BMI ranging from (25 – 29.9 kg/m²) is unhealthy for heart. This can be seen from the following representation –

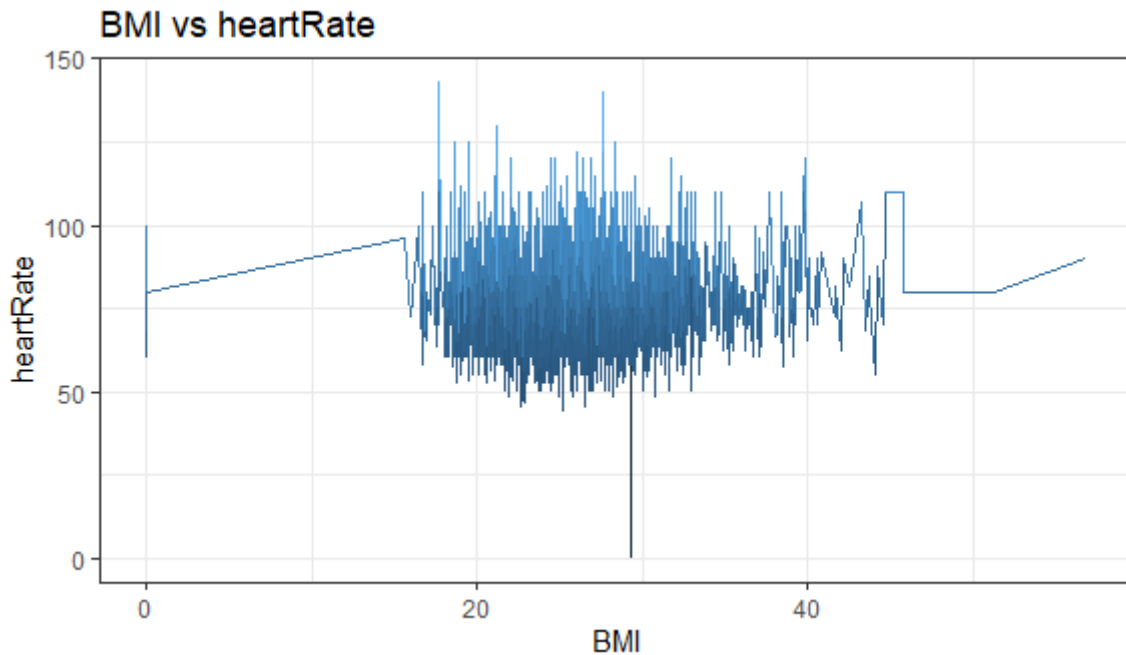


Fig. 9 (BMI vs Heart rate)

As we can see the graph represents that when the BMI is in the range of 20-40, the heart rate fluctuates repeatedly and shows signs of instability.

Methodology:

The methodology followed in this research includes –

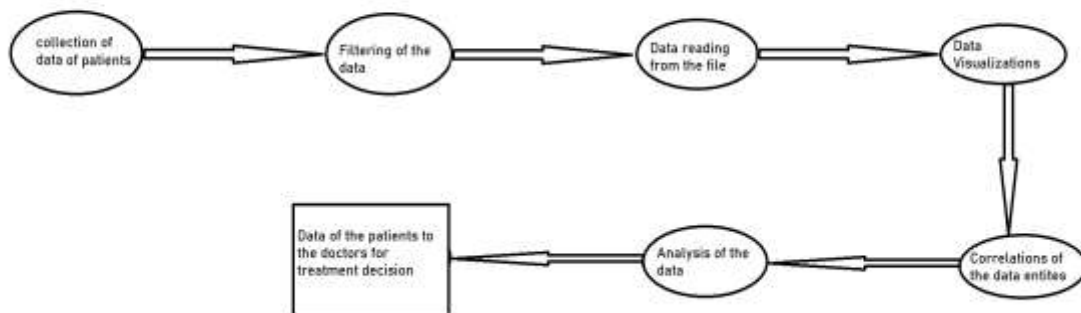


Diagram – Methodology for the proposed work

- **Search Strategy**

The articles for the Big Data in Healthcare were used from GutenbergR. The search strategy included all English language peer reviewed articles and studies. The selected title “DISEASES AND DISORDERS OF THE HEART AND ARTERIES IN MIDDLE AND ADVANCED LIFE” consists of all the articles for heart disease, symptoms, causes, syphilis, tobacco etc.

- **Search terms**

Some of the search terms applied to the selected database include: Big Data Analytics, Big Data in Healthcare, Heart Attack Prediction, Visualization and Big Data.

- **Review Procedure**

The database study is conducted for India using Framingham’s study Rule which consists of mainly statistical or quantitative data and uses different definitions.

- **Article selection Criteria**

Inclusion/exclusion criteria

The article encompasses various types of studies, for example, case-control studies, descriptive and qualitative studies. The article focuses on the scenario where we can associate our society/societal issues with Big Data (here in case heart disease).

Testing Technology Used:

RStudio supports the direct execution of code from within the source editor (the executed commands are inserted into the console where their output also appears).

A script is used to write the text file containing a set of commands and comments. Then the program is put to analysis and visualisation run onto the R console.

Future Scope:

This system can be used as a training tool for Nurses and medical students who are freshly introduced in the field related to Heart Diseases. Moreover, if implemented on a larger scale it can be used in medical facilities like Hospitals, clinics where a patient would not have to wait in long queues if he is feeling symptoms related to heart disease.

Artificial intelligence models can be prepared on this objective so that machine can predict the chance of heart attack in a person or patient. Machine learning can be used to develop effective prediction models for more accurate prediction.

Use of MLNN (Multilayer Neural Network) can be done to provide accurate results.

The proposed model requires an efficient processor with good memory configuration to implement it in real time. The proposed model has wide area of application like grid computing, cloud computing, robotic modelling, etc.

Conclusion:

Big data proposes vast promises for detecting interactions and nonlinearities in relationships among variables. Mobile devices, such as smartphones and tablets, and sensors, will continue to be the most indispensable tools available to deliver heart attack prediction and telecardiology services over wireless networks to reduce cardiovascular disease morbidity and mortality. The most important factor, however, in the development and application of big data, telecardiology, sensor use, mobile phone or tablet use and landline use is patient privacy and to safeguard the patient’s ability to direct and discover the use of his or her health care information.

Machine learning is getting advanced which is helpful in making correct choices and taking best decision. Machine learning has proved to be a best tool in making prediction in the healthcare sector. The study has shown that machine learning algorithms are better at predicting the absolute number of cardiovascular disease cases correctly. Through the study, we have found that awareness towards such kind of disease is important. So, correct prediction towards these will help the people and the doctors to take precautions as required. CVD is a type of disease that must be controlled to decrease the death ratio worldwide.

In addition, given the magnitude of increase in the prevalence of heart disease cases worldwide, it is important to look at different ways of prevention. The data clearly shows that individual based risk factors evaluation is insufficient; And that social determinants of health must be looked at.

References:

1. Journal Paper Reference

- Al Mamoon I, Sani AS, Islam AM, Yee OC, Kobayashi F, Komaki S (2013) A proposal of body implementable early heart attack detection system, 1-4. Citation: Alexander CA, Wang L (2017) Big Data Analytics in Heart Attack Prediction. J Nurs Care 6: 393. doi:10.4172/2167-1168.1000393 Page 8 of 9 J Nurs Care, an open access journal ISSN:2167-1168 Volume 6 • Issue 2 • 1000393.
- Ghadge P, Girme V, Kokane K, Deshmukh P (2015) Intelligent heart attack prediction system using big data. International Journal of Recent Research in Mathematics Computer Science and Information Technology 2: 73-77.
- Wanaskar UH, Ghadge P, Girmev V, Deshmukh P, Kokane K (2016) Intelligent Heart attack prediction system using big data. The International Journal of Advanced Research in Computer and Communication Engineering 5: 723-725.
- Kumar KG, Arvind R, Keerthan PB, Kumar SA, Dass PA (2014) Wireless methodology of heart attack detection. International Journal for Scientific Research and Development 2: 673-676.
- C. Beulah ChristalinLatha, S. Carolin Jeeva (2019)Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques.

2. Conference Paper Reference

- Demchenko Y, Grosso P, De Laat C, Membrey P (2013) Addressing big data issues in scientific data infrastructure. In Collaboration Technologies and Systems (CTS) 2013 International Conference on IEEE, pp: 48-55.
- Lee VYA, Tan SC, Chee JKK (2016) Idea identification and analysis (I2A): A search for sustainable promising ideas within knowledge-building discourse. In: CK Looi, J Polman, U Cress, P Reimann (eds.), Transforming learning, empowering learners: He International Conference of the Learning Sciences (ICLS), International Society of the Learning Sciences, Singapore.
- Sun J, Reddy CK (2013) Big data analytics for healthcare. Tutorial presentation at the SIAM International Conference on Data Mining, Austin, TX.

3. Book Reference

- Megahed FM, Jones-Farmer LA (2013) A Statistical Process Monitoring Perspective on “Big Data”, Frontiers in Statistical Quality Control, Springer, New York, pp: 1-21.