Predictive model for CAD in Diabetic Patients using Machine Learning Models

Project report submitted in fulfillment of the requirement for the degree of Bachelor of Technology

In

Computer Science and Engineering/Information Technology

By

Devyani Kaulas (151269)

Siddhesh Sharma (151298)

Under the supervision of

Mr. Nitin Kumar

to

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology Waknaghat, Solan- 173234, Himachal Pradesh

Candidate's Declaration

I hereby declare that the work presented in this report entitled "Diabetes Prediction System Using Machine Learning" in fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering/ Information Technology submitted in the department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology Waknaghat is an authentic record of the work carried out over a period from January 2019 to May 2019 under the supervision of Mr. Nitin Kumar (Assistant Professor, Computer science and Engineering).The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Devyani Kaulas (151269) .....
Siddhesh Sharma (151298).....

This is to certify that the above statement made by the candidates is true to the best of my knowledge.

Mr. Nitin Kumar
Assistant Professor
Computer Science & Engineering
Dated:

ACKNOWLEDGEMENT

We would like to express our special thanks of gratitude to our project guide Mr. Nitin Kumar who helped us in conceptualizing the project and actual building of procedures used to complete the project. We would also like to thank our Head of department for providing us this golden opportunity to work on a project like this, which helped us in doing a lot of research and we came to know about so many things.

Secondly we would like to thank our family and friends who guided us throughout the project so as to complete our project on time.

Thanking you,

Devyani Kaulas(151269)
Siddhesh Sharma (151298)

Table of Contents

List of Abbreviations

ACEI	Angiotensin-converting-enzyme inhibitor
ANN	Artificial Neural Networks
BUN	Blood urea nitrogen
CAD	Coronary Artery Disease
CKD	Coronary Kidney Disease
COPD	Chronic obstructive pulmonary disease
DL	Deep Learning
DM	Diabetes Mellitus
FN	False negatives
FP	False positives
Hb	Hemoglobin
HTN	Hypertension
IHD	Ischemic heart disease (IHD)
LR	Logistic Regression
ML	Machine Learning
NaN	Not-A-Number
PCO	Polycystic Ovary Syndrome
RFC	Random Forest Classifier
SES	Social Economic Status
SVM	Support Vector Machine
T3	Triiodothyronine
TN	True negatives
TP	True positives
WHO	World Health Organization

List of Figures

List of Tables

Abstract

Machine Learning and Deep Learning methods are applied to diagnose diseases and give a better insight to understand them, whether it is through predictive modeling or reducing the dimensions of feature space. There was utilization of various ML and DL algorithms on Diabetes patients to produce binary classification on Coronary Artery Disease (CAD). Here a deduced relationship between various features with Coronary Artery Disease (CAD) was established. There was utilization of various data preprocessing techniques and created classifiers using Logistic Regression, Random Forest Classifier, Support Vector Machine, Naive Bayes and Artificial Neural Networks (ANN). The best classifier is chosen which gave the best results based on the calculation of sensitivity, specificity and accuracy which has been computed using Confusion Matrix. These results can help Diabetic Patients in early detection of Coronary Artery Disease (CAD) and ways to avoid developing it to acute levels.

Chapter-1

Introduction

1.1 Introduction

Nowadays, we are blessed with technology of machine-learning and deep-learning which is applied on different fields such as crime detection, fraud detection, predictive models, various finances etc. One of the most evolved fields and common use case of this boon is in Medical applications to understand any disease better [1].

Diabetes (DM) is one of the most common diseases where patient's body ability to produce and respond to insulin is impaired which may results in increased glucose level in blood. Long-term complications of diabetes include vision problems, more thirst, frequent urination while on critical stages it can lead up to death. There are various other diseases that take emerge along with Diabetes such as Coronary Artery Disease (CAD), Coronary Kidney Disease (CKD), Chronic obstructive pulmonary disease (COPD), Hypertension (HTN), Hypothyroidism. These diseases do not give much of a clue until they become uncertain. Hence, early detection of these diseases alongside with proper treatment can help the patient to revive to a better condition [2].

Coronary Arteries are the major blood vessels that supply our heart with oxygen, blood and nutrients. When these arteries become damaged or impaired, the patient develops Coronary artery disease (CAD). Oily plaque builds up in these arteries which results in narrowing them, Hence, dropping the blood flow to patient's heart. These cardiovascular diseases develop over decades. They silently kill patient without giving any major symptoms to them until acute circumstances are not created such as Heart Attack which many time results in patient's death. Due to these reasons, research on this topic is escalating. Hence, dependency of diseases such as Coronary Kidney Disease (CKD), chronic obstructive pulmonary disease (COPD), Hypertension (HTN), Hypothyroidism can help in early detection and help in management of this disease [3]. In this study, we created a predictive model for Coronary Artery Disease (CAD) with the use of Diabetic patients' dataset. The rest of the paper is organized as follows: The information about dataset is in Section 2, Section 3 shows Data Preprocessing used on the dataset, Section 4 is regarding various Machine Learning and Deep Learning methods which have been

applied on preprocessed dataset, Section 5 is on results and evaluation by these various models and finally Section 6 concludes the paper.

1.2 Problem Statement

Diabetes is a major problem killing many people throughout the world. With the advancement in technologies, human life is prospering. Therefore why not use the technologies for the betterment of healthy lifestyle. The deep learning technologies and various machine learning algorithms are used for much type of prediction facilities. Often used by business giants for profits and sales. Here we are presented with the question of how can we use these technologies for the betterment of mankind. The various algorithms we have used and learned with time are to be challenged for prediction of something whose specialization only resides in the hands of experts. The machine has to be trained with the mind of doctors in order to learn the complexity of various features of bio mechanics of human beings and predict the complicated problems of living beings. These algorithms have to be implemented for the prediction of complicated diseases using various features and external factors provided from an authentic dataset.

1.3 Objectives

The main objective of this prediction system for CAD in diabetic patients is to find out a useful model for the benefit of mankind and can be understood be the following points:

a) Prediction for CAD (Coronary artery disease) in Diabetic Patients
b) Feature Relations with CAD using Machine Learning
c) Implementing the fundamentals of machine learning
d) To find relation between Diabetic Patient and his various factors that affects the disease

1.4 Methodology

It is a predictive model for Coronary Artery Disease (CAD) with the use of Diabetic patient’s dataset. The dataset is an authentic data collected from an expert supervisor. This set of data has been use on various important algorithms. Coronary Artery Disease is a brutal condition of heart in which there is a blockage of arteries which results in lack of

blood supply and hence untimely death of the patients. Diabetes Metticuls is a disease of metabolism with imbalance of glucose in their body. Scientists and various researchers have observed a connection between these two catastrophic diseases. This connection between these diseases can be established by various parameters taken under consideration. These parameters often include genetics, life style, drug abuse, body weight, socio economic factor. Hence by making a data set on these factors and having an authentic record of these observations the wonders of advanced sciences can be used for greater cause of humanities.

Machine Learning is a methodology for machines in which we can train the computer to learn from various data set. By applying various algorithms and going through a bunch of cost function the concept of machine learning algorithms can be used in predictive model often applying a relationship between these factors. With advancement of deep learning in this project we have implemented certain intricate points for better accuracy and performance of our model. This model will use the modern science for an integral function.

1.5 Conclusion

We have successfully created a predictive model for CAD. We'll be exploring various algorithms and their implementations in our model. By using the research of past scientist and taking many things under consideration, we'll be successful in making a predictive model for this chronic disease.

Chapter-2

Literature Survey

2.1 Introduction

Diabetes is a disease which results in various problems. The World Health Organization (WHO) [7] have surveyed and learned that 1.2 million people died due to this chronic disease. Moreover 2.2 million people suffered the same fate due to cardiovascular diseases. American Heart Association have come to the conclusion that diabetes is a relatively controllable risk for CAD [8]. The huge level of glucose affects the stroke levels for heart. Moreover people from Type 2 diabetes are prone to biological factors of hypertension, chest pain, obesity etc. making them more vulnerable to this disease.

Blood pressure, cholesterol, triglycerides, obesity, sedentary lifestyle, abnormal sugar levels and smoking etc are factors that are mutually exclusive for both of these chronic diseases. There have been abundant amount of research in this field in past 10 years that there research has created an opportunity to develop an important tool for future references.

2.2 Research about CAD and Diabetes.

There was a basic research in the basics of these diseases. Careful observation for the patients there health, important factors affecting them was taken under consideration. The understanding accumulated from these disease has been mentioned in the following paragraphs in depth.

2.2.1Coronary artery disease (CAD):

This disease is also termed as ischemic heart disease (IHD). The reduce blood flow builds tension in the arteries leading to the blockage. When the there is a blockage in the arteries the blood is unable to reach the heart and provide basic functioning as required by the heart to carry on the process of functioning of our body, which leads to the immediate death of the patient. The major symptom of this disease remains heart burning pain in the

chest, tensions in the movement of shoulder, neck, and jaw. Patient often experience heartburn.

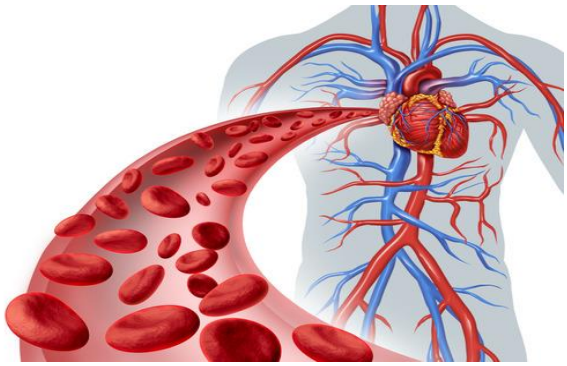

Figure 1- A Picture of arteries connection to heart

The affects of this disease is often used in classifying into four major categories. These categories are devised from the result or after implication of this fatal disease and major classification are as following:

i. Myocardial infarction
ii. Stable angina
iii. Sudden cardiac death
iv. Unstable angina

2.2.2 Diabetes:

Diabetes is a chronic disease in which levels of blood sugar and glucose is quit unstable. The result of this instability in the levels of glucose results in severe health issues. Sometimes these health issues can cause sudden death also. This disease is a disorder for metabolism hence it can be classified in three types.

There are many people suffering from this ailment. The number of people having this illness is growing gradually. According to a recent survey it has been found out that one in 11 adults are suffering from this disease. It is quite a dangerous statistic for a disease to spread like that.

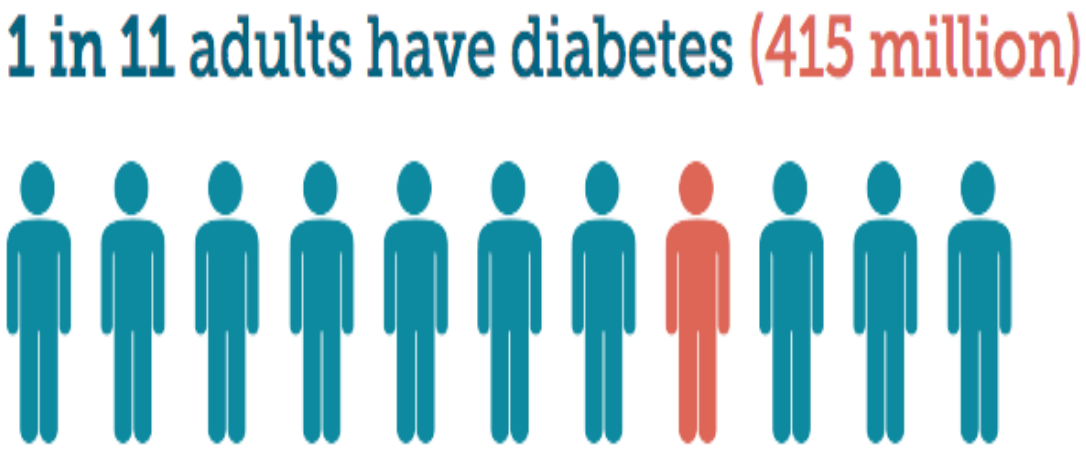

Figure 2- Diabetic Rate among adults

Type 1:

The body of the patient cannot produce insulin or produce too little insulin for the body which can cause certain serious damage. There is a great risk on pancreas of the person suffering from type 1 of the disease. According to recent statistics it has shown that people with type 1 diabetes are usually under 20.

Type 2:

The body is repulsive or cannot take the insulin produced in the body hence there is instability of insulin. People with type 2 are quit prone too many heart related ailments and according to report from WHO, maximum patients of Diabetes are suffering from Type 2.

Type 3:

Cases of Type 3 are quite rare, as in Type 3 there is a serious damage on the brain of the person. It is also known as Gestational Diabetes.

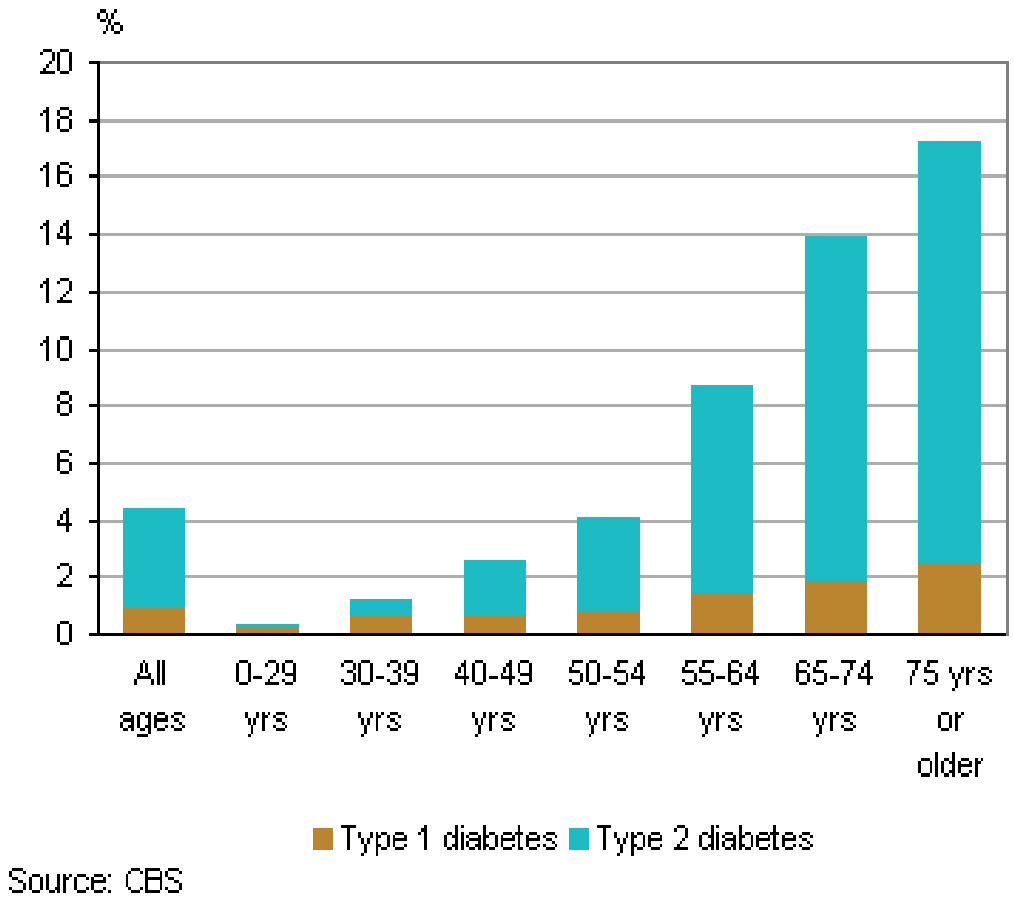

Figure 3- Statistic of Type 1, type 2 and type 3 diabetes among people

2.2.3 Relation between CAD and Diabetes:

In majority of cases of patients of various records it has been observed that people suffering from diabetes are quite prone to many heart related ailments. Scientifically also it has been established by many known scientists that diabetes and CAD are interrelated, as unstable blood sugar levels ultimately results in the blockage of arteries hence hampering the blood flow from the heart, which can sometimes result in sudden death of the person or patient.

Our machine learning algorithms with major tools of deep learning would establish related features in the two diseases. Upon running many test observation we came to a conclusion that there were 11 features that were better for the prediction of these results.

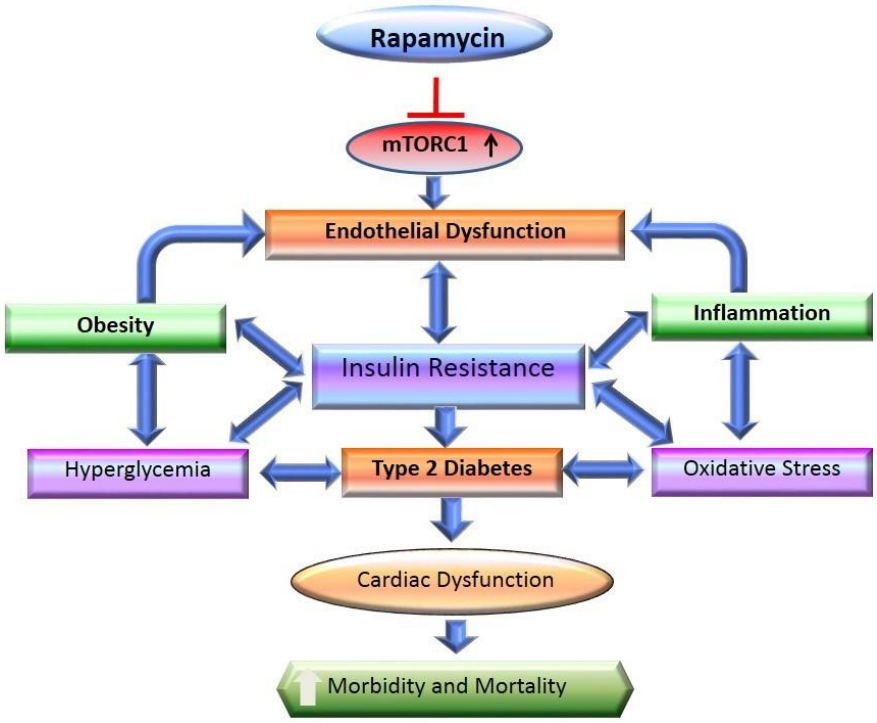

Figure 4-Relation between Heart disease and diabetes

Diagram above shows a clear relation between disease related to heart and there relation with diabetes. The above chart very well explains that imbalance of insulin in the body of the patient results in secretion of various enzymes that causes indirect and direct problems to the heart. Since CAD (coronary artery disease) is the most common heart disease there are many chances of having this disease with those patients who are suffering from the diabetes. With the help of data like that we can prepare a model that can establish a number of features which will connect or find a relation between theses disease. Upon finding this relation we will carry out our research for development of a model which help in prediction of this disease with the help of the data set and many past records.

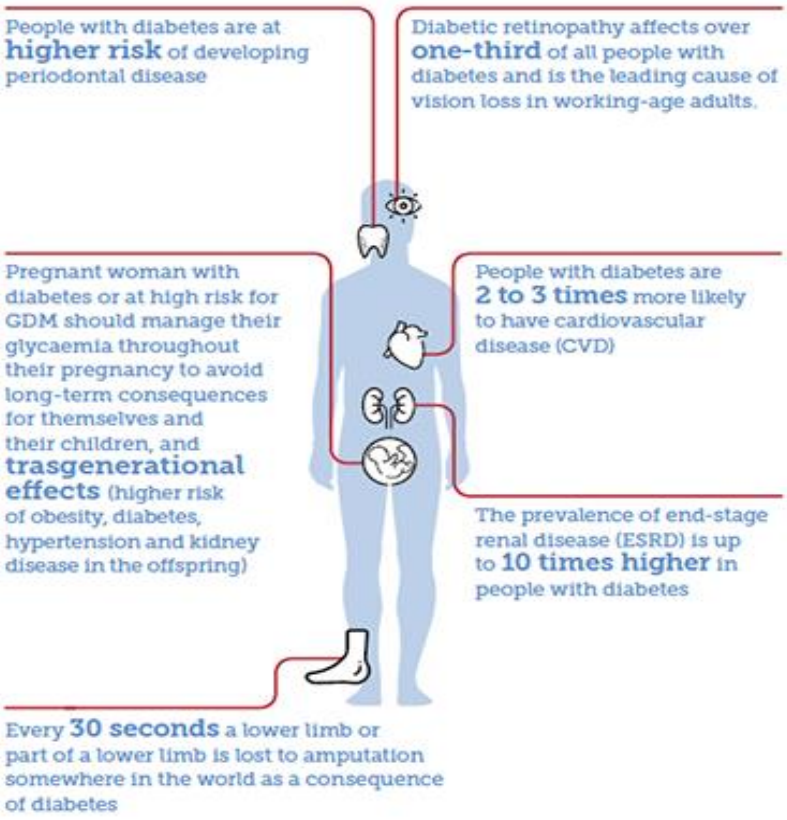

Figure 5- Diabetes effect on various body parts

Diabetes is a type of disease that affects many body parts in many different ways. It affects kidney, brain , heart etc. It causes many painful and chronic affects on the human body.

Hence there remains a terrific relation between the CAD and diabetes. Researchers have already established a relation between various enzymes and chemicals that link both of this disease. They have come to a conclusion that this disease is caused by certain common features. People sharing the common features in both of the disease would create a helpful database for the prediction analysis. People having a past history of this ailment or the ailment running in their families, drug abuse at a very early stage, alcohol addiction etc are certain major factors common to the disease.

2.3 Conclusion

CAD and Diabetes are both very harmful disease and have interrelated side effects on the human body. There can be many features that are common in both the disease. Through those features here it can be established that there can be a necessary prediction from the relevant data.

Chapter -3

Test Plan

The test plan for this predictive model for CAD in diabetic Patients has been taken from actual data of real patients from an authentic survey with permission of the owner. We have developed various methods or approach towards developing a systemized and synchronized way of using the data for the purpose of our model. The data specifically targets the crucial points required for our research. Moreover, the test plan is in accordance with our model and can be helpful for further improvements or developments in our model or models related to this type of research.

3.1 Dataset

The Diabetes patients consist of 8699 records of patients who are suffering from diabetes with 25 features (including dependent feature CAD). There has been collected by various reputed doctors. Most of the features are binary-categorical features either 1 (Presence of Particular Feature) and 0 (Absence of Particular Feature). The can be divided in following groups: Demographic, Patient Diagnosis, Lifestyle Factor, Lab Investigation and hemodynamic Variables.

Demographic features consist of patient's Sex, Age, Social Economic Status and Diabetes Duration. Diagnosis features informs whether patient is suffering from diseases such as Coronary Artery Disease (CAD), Coronary Kidney Disease (CKD), Chronic obstructive pulmonary disease (COPD), Hypertension (HTN), Hypothyroidism, Obesity, Polycystic Ovary Syndrome (PCO) and Poor Growth. Lifestyle Factors shows whether patient do physical activity, smoking, alcohol consumption. Lab Investigation tells the details about various hormonal tests such as Hb, T3, BUN and Hba1C. Hemodynamic Factors consists of drugs patient is currently using to cure his condition such as Anti-platelets, ACEI, Ca Channel blocker, Beta blocker, Statins.

Table-1: Features of Diabetes patients to predict CAD

Feature Type	Feature Name	Range
Demographic	Sex	1 / 2 (1 for male and 2 for female)
	Age	20 - 95
	Education Years	0 - 30
	SES(Social Economic Status)	1 / 2 / 3 (1 for low,2 for moderate and 3 for high)
	Diabetes Duration(Years)	0 - 40
Diagnosis	CAD(Resultant Feature)	0 / 1 (0 for no and 1 for yes)
	CKD	0 / 1 (0 for no and 1 for yes)
	COPD	0 / 1 (0 for no and 1 for yes)
	Hypothyroidism	0 / 1 (0 for no and 1 for yes)
	Obesity	0 / 1 (0 for no and 1 for yes)
	PCO	0 / 1 (0 for no and 1 for yes)
	Poor Growth	0 / 1 (0 for no and 1 for yes)
	HTN	0 / 1 (0 for no and 1 for yes)
Lifestyle Factor	Physical Activity	0 / 1 (0 for no and 1 for yes)
	Smoke	0 / 1 (0 for no and 1 for yes)
	Alcohol	0 / 1 (0 for no and 1 for yes)
Lab Investigation	Hb	6.0 - 17.4

	T3	0.6 - 190.2
	BUN	7.6 - 177.0
	HbA1C	4.2 - 21.9
Hemodynamic Calculation	Antiplatelet	0 / 1 (0 for no and 1 for yes)
	ACEI	0 / 1 (0 for no and 1 for yes)
	Ca channel blocker	0 / 1 (0 for no and 1 for yes)
	Beta Blocker	0 / 1 (0 for no and 1 for yes)
	Statins	0 / 1 (0 for no and 1 for yes)

The above table is the observation made from the data set. This table provides a brief summary of the integral features of the data set. The account of certain important features that are credible for accurate result are computed here for better understanding of the concept and future references in an easy manner.

3.2 Data preprocessing

Data Preprocessing is one of the major important features required for the training of the model. Since all the columns or rows might not be useful for the model or the data set that is available is not in the form in which it can be used for the training of the machine in all these cases data preprocessing is an important factor that determines the healthy start of the models.

Data preprocessing is a technique which is used to turn raw data to useful format. It can be through filling empty values, converting categorical variables to binary format through encoding. Initially we have 8699 patients and 25 features; we removed the rows where our dependent variable (CAD) which we are going to predict is NaN or Empty. A seaborn heat map can be seen in Figure-1. Heat map clearly indicates about lots of missing values in columns like PCO, Poor growth, t3 etc. Finally we removed such columns and all the

rows with empty values. The main reason behind handling missing values by removing them is to avoid biased estimates, which can further result in invalid conclusions. Finally, we are left information about 1199 patients and 11 features (including dependent feature CAD) [4].

Out of these 249 patients are suffering from Coronary Artery Disease (CAD) while remaining 950 does not experience Coronary Artery Disease (CAD).This clearly represents imbalanced dataset. The category Patients not suffering from CAD have more samples in dataset comparing to Patient suffering from CAD. This can result in biased learning, incorrect conclusions and decrease in accuracy rate for minority classes [5].

Hence, we have selected 249 patients suffering from Coronary Artery Disease (CAD) and 351 patients not experiencing Coronary Artery Disease (CAD). The final list of features to predict Coronary Artery Disease (CAD) can be seen in Table-2.

Table-2: Final Features of Diabetes patients to predict CAD

Feature Type	Feature Name	Range
Demographic	Sex	1 / 2 (1 for male and 2 for female)
	Age	20 - 95
	Diabetes Duration(Years)	0 - 40
Diagnosis	CAD(Resultant Feature)	0 / 1 (0 for no and 1 for yes)
	CKD	0 / 1 (0 for no and 1 for yes)
	COPD	0 / 1 (0 for no and 1 for yes)
	Hypothyroidism	0 / 1 (0 for no and 1 for yes)
	HTN	0 / 1 (0 for no and 1 for yes)
Lifestyle Factor	Physical Activity	0 / 1 (0 for no and 1 for yes)
Hemodynamic Calculation	Anti-platelet	0 / 1 (0 for no and 1 for yes)
	Statins	0 / 1 (0 for no and 1 for yes)

These are the final features after deleting many features on certain criteria. There were many factors involve that made certain features not useful for the model upon which the machine is being trained hence those features were deleted.

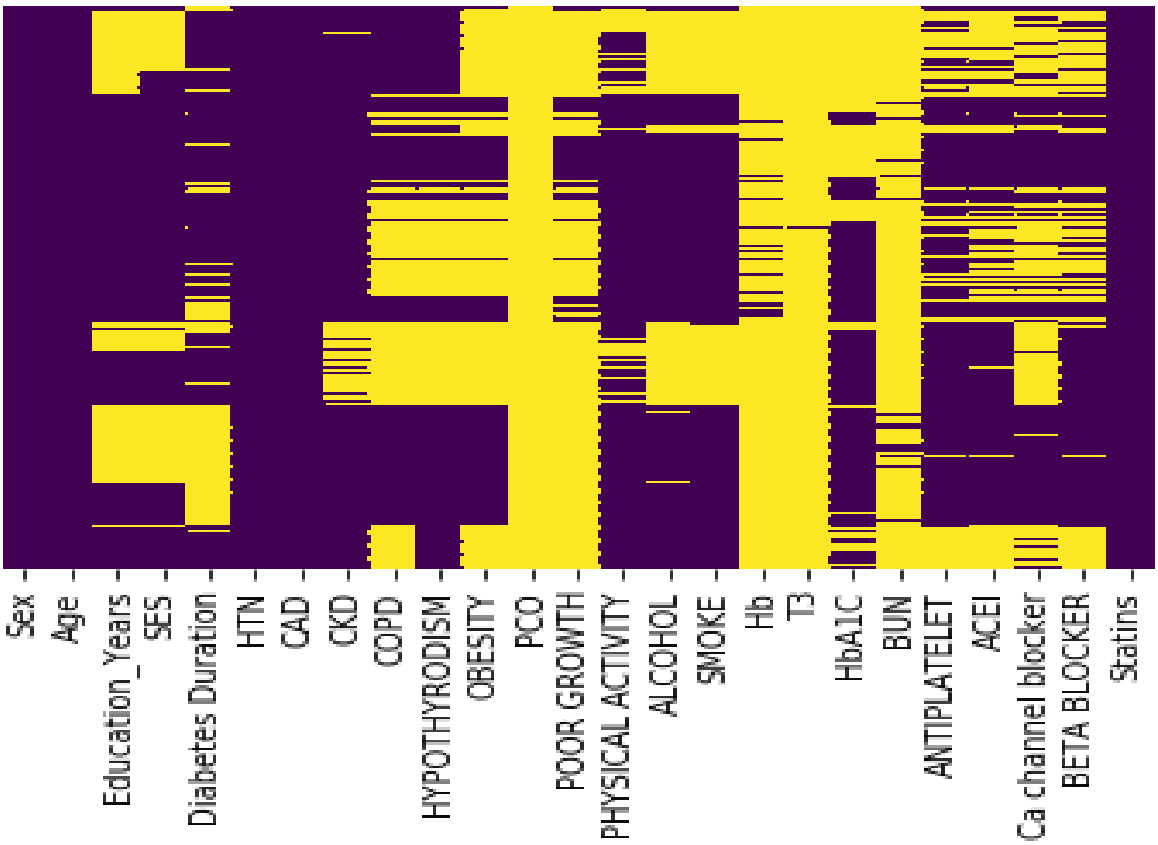

Figure 6- Revealing of empty and non empty values of patient's features. Yellow color shows the empty values (NaN) and violet color shows presence of data.

Through this seaborn chart we can observe that many features were having empty value and hence on the lack of availability of the data those features couldn't be used for the training of the data therefore they were not considered.

3.3 Conclusion

There was abundant amount of data for this model. The data was pre processed in order to make it useful for future implementations. Many features were deleted upon observation of its impact on the result of the model.

Chapter -4

Algorithms and Methodology

In this Chapter, we used various machine learning and deep learning methods on our diabetic patient dataset to predict CAD. We utilized Logistic Regression, Random Forest Classifier, Support Vector Machine, Naive Bayes and Artificial Neural Networks (ANN). These methods are explained in sections 4.1 to 4.5 and finally section 4.6 tells the features importance in predicting CAD through random forest feature importance.

4.1 Support Vector Machine Algorithm (SVM)

SVM is supervised machine algorithm which is mostly used for classification problems. We plot data items as a point in n-dimensional space. Then classes are separated by a hyper plane. A particular side of hyper plane belongs to category 1 while other side belongs to category 2.

It is extremely helpful for data analysis, classification problems, regression analysis etc. The graph for SVM usually looks like the following.

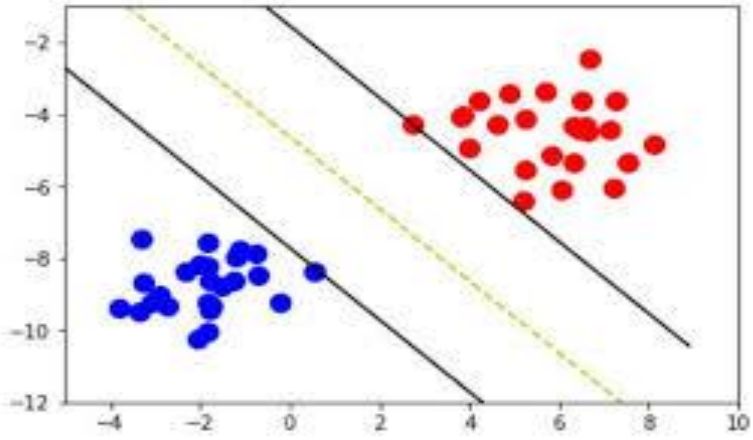

Figure 7- Showing a type of SVM in 2-D

An SVM is a type of classifier in which a hyper plane is separated. In 2-D hyperplane looks like a line as shown in the diagram above which separates two plane classifying both of them in different categories for careful observation and work.

4.2 Logistic Regression

Logistic Regression is mainly used in case when target variable is categorical, It works by the use of logistic function, also called the sigmoid function which takes value between 0 and 1. Based on which we can predict category of the target variable. This is sigmoid function

$$Cost(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

4.3 Random Forest Classifier

Random Forest Classifier is supervised machine learning algorithms, it created multiple decision trees and finally merges them together to give the final prediction. Multiple decision trees are based on various conditional statements which helps in great accuracy. Due to such great number of random decision trees, we can avoid over fitting .

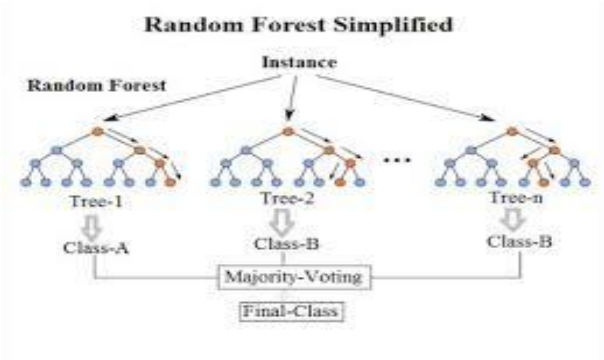

Figure 8-Classification in Random Forest

4.4 Naive Bayes

It is probabilistic algorithm based on Bayes's Theorem. It works great for classification problems based on assumption that every feature is equal and independent. This is Bayes formula

$$P\left(\frac{\mathrm{A}}{\mathrm{B}}\right) = \frac{\mathrm{P}\left(\frac{\mathrm{B}}{\mathrm{A}}\right)\mathrm{P}(\mathrm{A})}{\mathrm{P}(\mathrm{B})}$$

We are finding Probability of event A, given event B is already true.

4.5 Artificial Neural Network (ANN)

The idea of ANN is simply based on working on Human Brain, how humans take various input and finalize output constructed on various decisions. This decisions making is constructed to various layers in ANN.

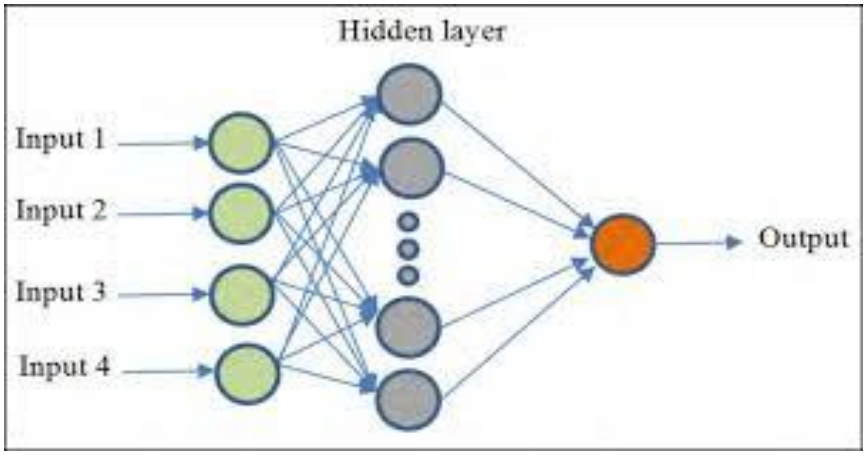

Figure 9 -Structure of a Basic ANN

Each input parameter has some particular weight age for prediction of output parameter, these weights are passed to hidden layers from input layer. In these hidden layer summation and activation function works to predict the final output. It mainly works on feedback method , where for good prediction network is rewarded else punished.

4.6 Feature Importance

Out of all the 10 features Sex, Age, Diabetes Duration, HTN, CKD, COPD, Hypothyroidism, Physical Activity, Antiplatelets, Statins which are going to predict CAD. These all features play an important role in prediction of Coronary Artery Disease (CAD).Their importance can be seen in Figure-2. It clearly tells Antiplatelets, Age, Diabetes Duration, Physical Activity are the major contributors in prediction of CAD.

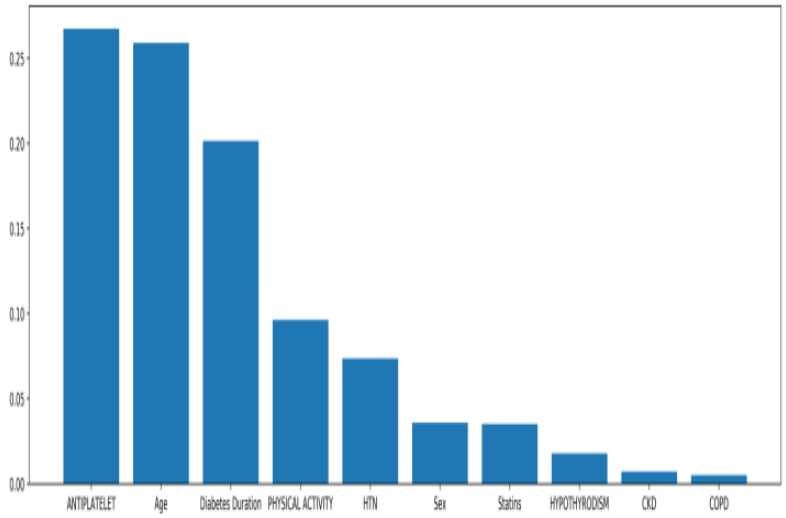

Figure 10-: Tells about the feature importance, here taller bar signifies more importance

4.7 Conclusion

We implemented various algorithms and observed which algorithms was the most suitable one and upon that we discovered that ANN was showing the best result out of all the data

Chapter -5

Result and Evaluation

After a great learning experience and going through careful observations of various models, considering much different approach for a particular result, we can finally evaluate the major results and explore the right outcomes. The various Machine Learning algorithms and deep learning models taken into account proved to be extremely useful for the collection of right data for cross validation.

For every deep learning and machine learning model we have created. We will calculate accuracy, sensitivity and specificity. These are one of the most important performance measures in medical field [6]. These can be calculated with the use of Confusion matrix.

5.1 Confusion Matrix

Confusion matrix can be considered as a table used to describe performance of a classifier. For a binary classifier there are two rows and two columns which consists of True positives (TP), True negatives (TN), False positives (FP), False negatives (FN). Confusion matrix is shown in

TP: Actually it's positive and predicted as positive.

FP: Actually it's negative but predicted as positive.

FN: Actually it's positive and predicted as negative.

TN: Actually it's negative and predicted as negative.

Table-3: A table about confusion matrix

		Predictions	
		Class Positive	Class Negative
Actual	Class Positive	TP	FN
	Class Negative	FP	TN

5.2 Sensitivity

Sensitivity refers to actual positives correctly predicted as positive by the classifier out of all the positives. It can be calculated by Confusion Matrix through this mathematical formula.

$$Sensitivity = \frac{TP}{TP + FN}$$

5.3 Specificity

Specificity refers to actual negatives correctly predicted as negative by the classifier out of all the negatives. It can be calculated by Confusion Matrix through this mathematical formula.

$$Specificity = \frac{TN}{TN + FP}$$

5.4 Accuracy

Accuracy tells about correctly predicted results by the classifier. It tells about how accurate is the classifier .It can also be calculated through Confusion Matrix by using this mathematical formula

Table 4-Comparison between Confusion Matrix

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$$

Classifier	**Confusion Matrix**
Support Vector Machine	105 | 10 13 | 52
Logistic Regression	98 | 17 16 | 49

Random Forest Classifier	105 10 17 48
Naive Bayes	103 12 20 45
Artificial Neural Network	101 14 16 49

The table completely explains about the performance of each ML algorithm. These values can be further used to deduce sensitivity, specificity and accuracy for each model. In medical problems it is important that we choose the model with great sensitivity and specificity and not just accuracy. We can clearly see False positive of SVM and ANN is least. Hence, it's must me the most optimum choice to choose a model among these.

Table 5-Comparison between Sensitivity, Specificity and Accuracy

Classifier	**Sensitivity**	**Specificity**	**Accuracy**
Support Vector Machine	0.86	0.84	0.85
Logistic Regression	0.86	0.75	0.82
Random Forest Classifier	0.90	0.78	0.86
Naive Bayes	0.88	0.69	0.81
Artificial Neural Network	0.93	0.69	0.85

Clearly, highest Sensitivity is given by ANN while highest Specificity is shown by SVM. The accuracy achieved by ANN, SVM is equal. Algorithms tend to show poor specificity.

5.5 Conclusion

In the above table we have clearly mentioned about the matrix formed upon implementing those algorithms. There result and performance can be compared easily. By comparing these algorithms we have concluded that ANN and SVM tends to be better for the operation or use of the function in model.

Chapter-6

Conclusion and future Work

6.1 Conclusion

Here we applied various Machine Learning and Deep Learning methods to build a Coronary Artery Disease (CAD) classifier. We achieved optimal sensitivity, specificity and accuracy through Support Vector Machine (SVM). ANN reached the greatest sensitivity mark up to 0.93 and 0.85 accuracy. Further through Table-3, We can conclude with increase in Age and Diabetic Duration there is an increased chance for patient to develop CAD. To avoid CAD in such diabetic patients, proper need of Physical Activity and drugs such as Anti-platelets are required under proper guidance of doctor. This would help in timely cure of the disease. The harmful effects of this disease can neutralized with early diagnosis and proper medication. The precision and sensitivity observed would be helpful for future models and can create a useful tool for many other ailments as well.

CAD also develops other diseases such as CKD,HTN and COPD. These all can be only minimized by detection of these diseases at proper time and physical activity.

6..2 Future Work

In future, we aim to consider more features such as Patient's Family history and checking their dependency on Coronary Artery Disease (CAD). We also aim to increase patient samples to get more accurate results for giving proper guidance to patients. The more data we get better are the samples for cross validation for the cases run on the model.

This will be an extremely important tool in the futures to come and will be considered as the biggest inventions in order to cure root problems. The rate of people with disease will gradually fall and hence this disease won’t be a threat to the mankind.

Doctors can use it for their own validation, for consultancy etc. People having a doubt or just wanted to have a normal routine checkup can take its consideration into the account

Hence by modern technologies and advancement in future articles this disease would have a better cure and timely medication use to this system.

Therefore with help of machine learning algorithm this model is capable of finding a timely diagnosis of a fatal disease CAD(Coronary Artery Disease) in the mean course of time. Moreover for future work in this field there are many observations that can be taken into account for useful consideration.

Apart from that this model can be deployed as an web application or android application which can be available for free and can be tested by real users. These results can help us improving the model if we shift towards reinforcement learning.

In today's time where ML is growing at higher rate, With proper guidance and using of Principal Component Analysis, This model accuracy can be reached to a mark of 90%.

References

1. A. Cüvitoğlu and Z. Işik, "Classification of CAD dataset by using principal component analysis and machine learning approaches," *2018 5th International Conference on Electrical and Electronic Engineering (ICEEE)*, Istanbul, 2018

2. Kang, Hyun. (2013). The prevention and handling of the missing data. Korean journal of anesthesiology.

3. Bektas, Jale & Ibrikci, Turgay & Özcan, I.T.. (2017). Classification of Real Imbalanced Cardiovascular Data Using Feature Selection and Sampling Methods: A Case Study with Neural Networks and Logistic Regression. International Journal on Artificial Intelligence Tools.

4. Aamodt, A. and Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches, *AI Communications*

5. R. Alexander Pyron, John J. Wiens, A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians, Molecular Phylogenetics and Evolution, Volume 61, Issue 2, November 2011, pp. 543-583

6. R. Alizadehsani, M.H. Zangooei, M.J. Hosseini, J. Habibi, A. Khosravi, M. Roshanzamir, F. Khozeimeh, N. Sarrafzadegan, S. Nahavandi, Coronary artery disease detection using computational intelligence methods, Knowledge-Based Systems, 109 (2016) 187-197.
7. J. Bektas, T. Ibrikci, and I. Ozcan, 'Classification of Real Imbalanced Cardiovascular Data Using Feature Selection and Sampling Methods: A Case Study with Neural Networks and Logistic Regression,' International Journal on Artificial Intelligence Tools, 2017.
8. Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm," Computer Methods and Programs in Biomedicine, vol. 141, pp. 19-26, 2017/04/01/ 2017.
9. D. Giri , U. Rajendra Acharya , R.J. Martis , S. Vinitha Sree , T.-C. Lim , T. Ahamed Vi , J.S. Suri , Automated diagnosis of coronary artery disease affected patients us- ing LDA, PCA, ICA and discrete wavelet transform, Knowl.-Based Syst. 37 (2013) 274–282.
10. S. Patidar , R.B. Pachori , U. Rajendra Acharya , Automated diagnosis of coronary applied on heart rate signals, Knowl.-Based Syst. 82 (2015)