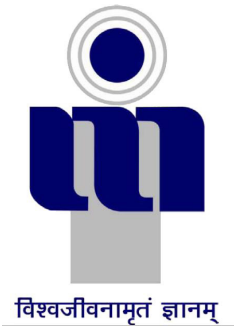
SMART CARDIAC RISK ANALYSIS: A MACHINE LEARNING SOLUTION

A project report submitted in partial fulfillment of the requirements for B.Tech. Project

B.Tech.

by

Aryan Sachan (2016IPG-020)
Mohit Nama (2016IPG-054)
Sanya Verma (2016IPG-092)



ABV INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT
GWALIOR-474 010

2019

CANDIDATES DECLARATION

We hereby certify that the work, which is being presented in the report, entitled **Smart cardiac risk analysis: A machine learning solution**, in partial fulfillment of the requirement for the award of the Degree of **Bachelor of Technology** and submitted to the institution is an authentic record of our own work carried out during the period *May 2019* to *September 2019* under the supervision of **prof. Anurag Srivastava** and **Dr. Pinku Ranjan**. We also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken.

Date: Signatures of the Candidates

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

Date: Signatures of the Research Supervisors

ABSTRACT

India has the second largest population in the world. Due to insufficient medical facilities for such a large population, the deaths caused by cardiovascular disease in India rose by 34 percent from 201.5 to 272 death per lac population between 1990 and 2016. This death rate is much higher than the global average. According to an article by Nicholas Parry based on a survey done in June 2018, in India, we lose approximately 1.7 million lives every year due to heart diseases. There are only 4000 cardiologists while we need approximately 88,000 cardiologists to manage so many patients. It is the need of the hour that the country adopts the population-level strategies to reduce the risk factor. Recent studies have focused on using ML techniques to develop predictive models that can estimate the incidence of heart failure. Most of these research used a binary output class in which either the existence or lack of heart failure would be predicted. A multi-level risk assessment of cardiovascular failure was undertaken in this study using the C4.5 decision tree classifier to estimate different risk types of heart failure. On the other hand, by combining three significant risk factors with data on cardiovascular failure, trying to enhance the early instances of cardiovascular failure. Our predictive model shows 86.5 per cent sensitivity, 95.5 per cent specificity and 86.53 per cent precision improvement in current studies.

Keywords: Heart failure risk, cardiovascular diseases , machine learning, deep learning, decision-tree classifier, cross validation method, accuracy, apache, MySQL, PHP.

ACKNOWLEDGEMENTS

We are deeply indebted to **Prof. Anurag Srivastava** and **Dr. Pinku Ranjan** and are obliged to give us the autonomy to explore and experiment with ideas. We would like to take this opportunity to express our profound gratitude to them not only for the academic guidance they have provided but also for the keen interest they have shown in our project. Their continual assistance proved very fruitful and was instrumental in instilling self-assurance and trust within us. The nurturing and blossoming of the current project is primarily due to their precious advice, suggestions, astute judgment and constructive criticism. Our mentor has always answered myriad of our doubts with smiling grace and prodigious patience, never allowing us to feel that we are novices by always listening to our opinions, appreciating and improving them and by giving us a free hand in our project. It is only because of their overwhelming concern and helpful attitude that the current work has reached its prevailing state.

Finally, we are thankful to our institution and to our colleagues whose continuous encouragement has served to renew our spirit, refocus our attention and energy and helped us to carry out this project.

(Aryan Sachan) (Mohit Nama) (Sanya Verma)

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

ML	Machine learning
DL	Deep learning
DM	Data Mining
ANN	Artificial Nueral networks
CNN	Convolutional nueral network
SVM	Support vector machine
MOM	Middle of Maximum
PHP	Hypertext Preprocessor
MLP	Multi Layer Perceptron
HTML	Hyper Text Mark-up Language
UI	User interface
Avg	Average
Hg	High
Lw	Low
VL	Very Low
VW	Very High

CHAPTER 1

INTRODUCTION AND LITERATURE SURVEY

This chapter has an insight overview for the entire project. It starts with details of heart diseases and the attributes which affect them in the section 1.1. In the next subsequent section has the details about the technologies currently in use. Section 1.3 contains the details about the previous papers following with research gap. The last section has the main objectives of our project.

1.1 Introduction

Heart disease defines a number of circumstances that impact your heart. Diseases under the heart disease category include blood vessel diseases, such as coronary artery disease, heart rhythm problems and heart defects you're born with (congenital heart defects). The word heart disease is usually used interchangeably with the term "cardiovascular disease". Cardiovascular disease generally involves decreased or blocked blood vessels that may lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that influence the muscles, valves or rhythm of your heart, are also regarded forms of heart disease.

Dr. Satyendra Tewari is a leading cardiologist at Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow. He is visited by a large number of patients suffering from cardiovascular disease. He was concerned that sometimes he can not help many severe patients due to inappropriate allocation of appointments. So he asked prof. Anurag Srivastava for solution. Prof. Anurag Srivastava is our project co-ordinator. So we are trying to solve this problem with our project. Our project is currently growing under TIIC.

- **TIIC:** Technology Innovation and Incubation Centre is the Industrial Interface of Atal Bihari Vajpayee-Indian Institute of Information Technology and Man-

agement, Gwalior established in 2008. The main objective of TIIC is to sell the intellectual goods of ABV-IIITM to the sector on competitive terms and at the same moment to give economic significance to the teaching and Research and development operations of ABV-IIITM.

1.1.1 Attributes which affect the heart diseases

In our work, The heart disease data set of the Cleveland Clinic Foundation has been used, which is accessible via internet. We collect these major attributes that have significant role in analysis of heart risk failure for each patient :

(a) **Age:** The risk of heart failure rises with age in both males and females. About eight out of ten individuals who die from coronary heart disease are 62 or older. Heart disease is becoming more prevalent as you age so it is essential to go for frequent check-ups and to be careful.

(b) **Sex:** Several conditions that raise heart disease risk are not unique to women but they are more prevalent in females. For example, autoimmune disorders such as rheumatoid arthritis, lupus, and scleroderma are more common in women and appear to confer a higher-than-normal risk of cardiovascular disease.

(c) **CP (chest pain):** Chest pain or discomfort created by not getting enough oxygen-rich blood from your heart muscle. It is also called angina. It is a natural consequence of a fundamental cardiac issue, generally coronary heart disease (CHD).

(d) **restbps (resting blood pressure):** Uncontrolled increased blood pressure can damage your heart in a variety of ways, such as coronary artery disease. Coronary artery illness influences the arteries that supply your heart muscles with blood. Arteries that reduce coronary artery disease do not allow blood to flow freely through your veins.

(e) **fbsu (fasting blood sugar):** High blood sugar from diabetes can harm the blood vessels and nerves that regulate your heart and blood vessels. The longer you have diabetes, the more probable you are to have cardiovascular disease. Cardiovascular diseases are the most common causes of death in adolescents with diabetes.

(f) **MHRA:** Patients with atrial fibrillation, heart rate is advised to be optimally controlled between 60 bpm and 80 bpm during rest and between 90 bpm and 115 bpm while exercise gently, as a constant heart rate of more than 100 bpm may be harmful to the heart leading to cardiovascular failure.

(g) **restegr (resting electrocardiographic result):** The prognosis of major ECG abnormalities is greatly affected by the presence of symptomatic CHD (coronary heart

disease). In males with symptomatic CHD, residual electrocardiograms may assist to identify a high-risk group that may profit from intervention.

(h) **Chol (serum cholesterol):** It is basically a lipid. Cholesterol and its derivatives are significant components of other steroid compounds cell membranes and precursors however, a large percentage of low-density lipoprotein in the blood (which carries cholesterol to the tissue) is correlated with an enhanced danger of coronary heart disease. High cholesterol is a major cause of coronary heart disease. Cholesterol can be divided into two categories-

LDL(low density lipoprotien) cholesterol (value in mg/dl)	**Heart disease risk**
130-159	Lw risk
160-189	Hg risk
190 and above	Very Hg risk
HDL(high density lipoprotien) cholesterol (value in mg/dl)	**Heart disease risk**
Above 60	Low-level risk
Less than 40	High risk

Table 1.1: Types of Cholesterol

Cholesterol itself does not cause symptoms, so many individuals are unaware that their cholesterol concentrations are too high. It is therefore essential to find out what your cholesterol figures are. Lowering cholesterol concentrations that are too enhanced reduces the risk of developing heart disease.

(i) **peak (ST depression induced by exercise relative to rest):** Stress test for treadmill electrocardiogram (ECG) is usually used to detect disruptive coronary artery disease (CAD). The presence of the ST segment changes, whether it's depression or elevation. The presence of CAD is most often stated on the ECG during the treadmill examination and warrants even more prevention.

(j) **Slope (the slope of the peak exercise ST segment):** A treadmill ECG stress test is regarded abnormal when there is a horizontal or downward ST-segment depression higher than or equal to 1 mm at 60 to 80 ms after point J. In particular, the incidence of horizontal or downward ST-segment depression at a reduced workload (calculated in METs) or heart rate shows a worse prognosis and a greater probability of multi-vessel disease.

1.1.2 Heart risk failure prediction approaches

The health care environment involves an enormous amount of information like genetic data, clinical data and data from electronic health records (EHRs). Machine learning, data mining and deep learning techniques provide methodology and technology for the extraction of relevant information for decision-making. Extensive study on all elements of heart disorder (diagnosis, treatment, ECG, ECHO, etc.) has resulted into enormous information.

1.1.2.1 Machine Learning approach

This approach can be divided into these parts-

Supervised Learning: Supervised learning is the system in which both input and desired output are provided for future data processing. In this there are two kinds of learning tasks: regression and classification. Some of the most common algorithms are Genetic algorithms, Decision Trees (DT) Artificial Neural Networks (ANN) k-Nearest Neighbors (k-NN), SVM.

Unsupervised Learning: Unsupervised learning is used to draw inferences from data set contains input information without a marked response. In this section there are two learning tasks Association and clustering. Some of the most common algorithms are k-means clustering and association rule learning algorithm.

Semi-supervised Learning: It is a learning which is a combination of marked and unmarked information that falls within the area of controlled and unsupervised learning. This learning is mainly used in the classification of web pages, genetic sequences and voice recognition. semi-supervised learning is usually split into two learning tasks, classification and clustering.

Reinforcement Learning: Reinforcement learning is a unique form of ML technique. It concerns how software agents automatically determine the optimum conduct within a special framework to maximize its effectiveness. The Reinforcement Signal gets feedback from the policeman about his behavior. A schematic representation of ML approaches is in Figure 2.1 .

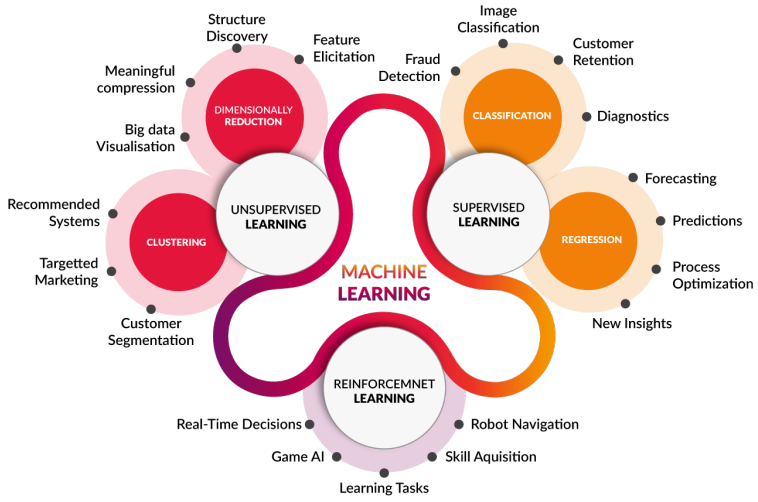


Figure 1.1: Approches of ML [1]

1.1.2.2 Deep learning approach

Deep learning [3] is referred to as profound structured learning or hierarchical learning. Most of the DL architectures and algorithms are modified within the framework of the [4] ANN. The ANN is comprised of neurons (interconnected nodes). Initially, the input layer receives an input and passes the altered version of the input to the next layer. Layers between input and output known as hidden layers and composed of various linear and non-linear transformations. The most prevalent deep learning architectures are multi layer perceptron (MLPs), Convolution neural networks (CNNs) and recurrent neural networks (RNNs).

Multi layer perceptron (MLP): MLP is an artificial neural feed-forward network described as a group of input nodes linked as a directed graph between input and output layers. It utilizes a back-propagation technique for network training and is commonly used to solve supervised learning issues as well as parallel-distributed processing and computational neuroscience studies.

Convolution neural networks : CNN is referred as profound artificial neural networks are widely used to classify pictures, clustering based on resemblance, object recognition and related algorithms used to define faces, road signs, tumours and various elements of visual information.

1.2 Existing architecture for heart risk failure prediction

Significant attempts have been made to detect cardiovascular diseases through ML, data mining and DL methods that enhance the precision of heart disease predictions. Machine learning algorithms are explored for evaluating and predicting the seriousness of cardiac failure by ANN, SVM, classification and regression tree. Some of these major attempts are-

- In the heart disease diagnostic significant attempts are made by authors, best approach is SVM which provides an accuracy of 74.60 percent by [5].
- To improve the performance in cardiovascular diseases classification and prediction Tree based methods are used by [6].
- Artificial neural network back-propagation algorithm is proposed with 13 clinical features to predict heart diseases and proved with the accuracy of 92 percent as in [7].
- Some Characteristics of genetic and phenotype variables used to study the performance of neural network techniques to predict coronary artery disorder. 23.9 percent improvement by using neural network method on the genes data set over a single classifier approach used by [8].

1.3 Literature Review

- An approach of developing a method for prediction of IHD (Ishemic heart disease) using smartphone. In this approach a score is generated using data mining technology and classified the risk into three categories low, moderate and high.[9] mainly focuses on sensitivity and specificity to check risk to have IHD.
- In [10] developed an automated diagnostic system based on chi square statistical model for cardiovascular disease prediction. It focuses primarily on improving the functions and removing the difficulties of the predictive model. It improves the quality of decision making during the diagnosis process of cardiovascular disease.
- This [11] research paper applied artificial neural network (ANN) with machine learning to predict heart disease. They train the system for one year time period to get enough data to make a prediction. Various tools are used like smart mirror, smart mouse, smart chair, smartphone etc. in data gathering.

- Paper [12] introduced modified hybrid algorithm combining particle swarm optimization (PSO) with back propagation neural network (BPNN). In this PSO is based on chaos theory which overcomes premature convergence. Then for training BPNN model they used improvised CAPSO algorithm.

1.4 Research Gap

When several patients are trying to get an appointment there is a possibility that a trivial patient may get an appointment before a more critical patient. This may be a cause of the death of a critical patient due to the unavailability of treatment at the right time. To cure the patients, an appropriate allocation of appointments plays a crucial role. Seeing this scenario, it is very important to make an efficient allotment for a cardiologist in a way that we can prioritize appointment according to the severeness of heart failure risk of patients. This is very beneficial for the patients as well as the hospital delegates for the management of the patients. In this way, we can decrease the death rate caused by cardiovascular disease. Thus we aim to combine biological knowledge and engineering principles to address medical needs that can contribute to the development of revolutionary and life-saving concepts.

- Most of the models have done only the binary classification.
- All models work on a fix number of set of predefined attributes.
- Some models contain kits that are very expensive and is not affordable for a common person.
- Highly technical interfaces and various kinds of inputs with technical terms take away a huge category of population away from it.
- In heart failure cases, accuracy is very important so far applied models could not reach the level of accuracy required to be implemented on practical basis.

1.5 Objectives

The following is the main objectives of our project:-

- Train the clinical data set according to the considered parameters.
- Try to predict how strong is the probability of a patient to have a heart failure using multilevel risk assessment analyzing the attributes such as blood pressure, cholesterol, diabetes, obesity, etc.

- To develop a prototype of a heart risk failure prediction system.
- Prioritize the patients according to their risk of heart failure by providing an optimal classification technique.
- Try to increase the accuracy of our model.
- To develop a user interface to allot an appointment according to this priority.

CHAPTER 2

DESIGN DETAILS AND IMPLEMENTATION

The chapter discusses the intrinsic and comprehensive details of the implementation. In this section we comprehensively see that how cardiovascular diseases can be analyzed with the help of approaches we have discussed earlier-

2.1 Data set

In this research, The heart disease data set of the Cleveland Clinic Foundation has been used, which is accessible online. We collect 10 major attributes that can have significant role in analysis of heart risk failure for each patient. Each instance in output class contains 9 attributes. The attributes of heart failure risk prediction previously discussed in detail. So small description of data set is as follows-

- sex (0=male, 1=female)
- age (in years)
- cp (chest pain)
- restbps (resting blood pressure)
- chol (serum cholesterol in mg/dl)
- fbsu (fasting blood sugar)
- restegr (resting electrocardiograph result)
- mhra (maximum heart rate achieved)
- exing (exercise induced angina)

2.1.1 Pre-processing the data set

Data prepossessing is an essential aspect of ML,as the quality of the data and the valuable information that can be obtained, directly impacts the capacity of our model to learn. Therefore, It is very important that we pre-process our data before we add it to our model. In our model we preprocess our data by handling the null values in some attributes, categorical values, standardization etc.

2.1.2 Train the data set

The training data set is the one used to train an algorithm to understand how concepts such as neural networks can be applied, to learn and to produce results. It involves both input and anticipated output information. Training sets make up the bulk of the complete information, approximately 60 percent or as per our requirement.

2.1.3 Validation of data set

The test information set is used to assess how well your training data set algorithm has been trained. In AI projects, we can not use the training data set at the testing point because the algorithm will already understand the anticipated output, which is not our objective. Testing sets constitute 20 percent of the information. The test set shall ensure that input information are grouped together with verified right outputs, usually by human verification.

We want our model to generalize well without over fitting. We can ensure this by validating the model. Through Cross-validation we take random sub samples and categorize the whole data into testing and training data set. The 10-fold cross-validation technique is used to evaluate the output of the classifiers to analyze the general output of the decision tree classifier. Figure 2.1 represents a clear picture of cross validation method. Usually this technique is used to optimize the use of the set of information.

The set of information is divided randomly into ten equal sub-sets. Each of them has about the same percentage of different class labels Of the ten subsets, one subset is kept as testing data, and the other nine subsets are retained as training data. The cross-validation method is then implemented ten times until each of the ten subsets has been used as a testing set exactly once. Then the outcomes can be averaged to assess the efficiency of the classifier. The benefit of this technique is that all subsets are used for training as well as for testing, and each subset is used to test precisely once.

Figure 2.1: Cross Validation

2.2 Imlementation/Execution of project

The problem has been broken into many sub-parts(as shown in Figure 2.2), the implementation details of which are following:

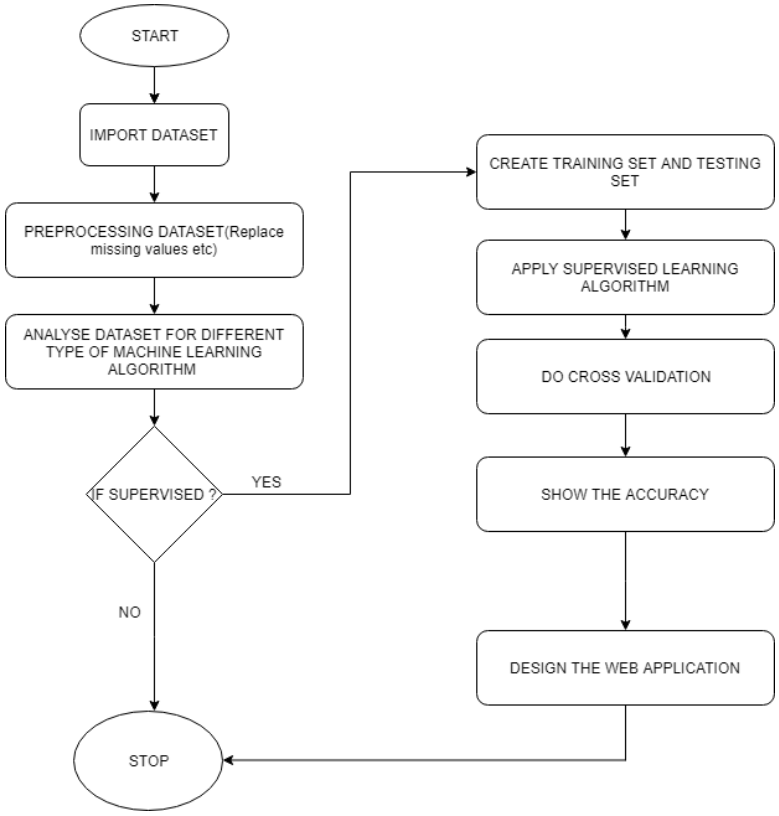


Figure 2.2: Project flowchart

2.2.1 Creation of Model

For solving our problem we use machine learning algorithm which is implemented in python. It's simplicity allows developers to write reliable models. We use an open

source cross-platform SPYDER IDE (integrated development environment) to create our machine learning model. In the collected data set we have fourteen attributes.We consider only the discrete attributes which are important as far as cardiovascular diseases are concerned. We train our data set and then use decision tree classifier along with confusion matrix approach to detect the severeness of the heart risk failure into different levels. We also use Pruning and post-pruning method that decreases the final classifier complexity and thus enhances predictive precision by reducing over fitting.

2.2.1.1 Decision Tree Classifier

Rule-based classifiers, neural networks, SVM, and naive Bayes classifiers are distinct methods to fix a classification problem. Decision Tree Classifier is a straightforward method of classification as shown in figure 2.3. Decision Tree Classifier carries number of carefully designed questions about the test record characteristics. Generally, decision tree having a specific rule for each and every path from root node to the leaf node.So decision tree can be converted into a specific set of rules. The technique of classification used in multiple areas such as medical diagnosis, like classification issues, marketing, detection of fraud, etc. is commonly used. There are different types of decision tree classifier namely ID3, CART and C4.5. We have successfully implemented the C4.5 decision tree classifier to our data set. The C4.5 algorithm is used in ML as a decision tree classifier that is implemented in a data set to generate a decision.

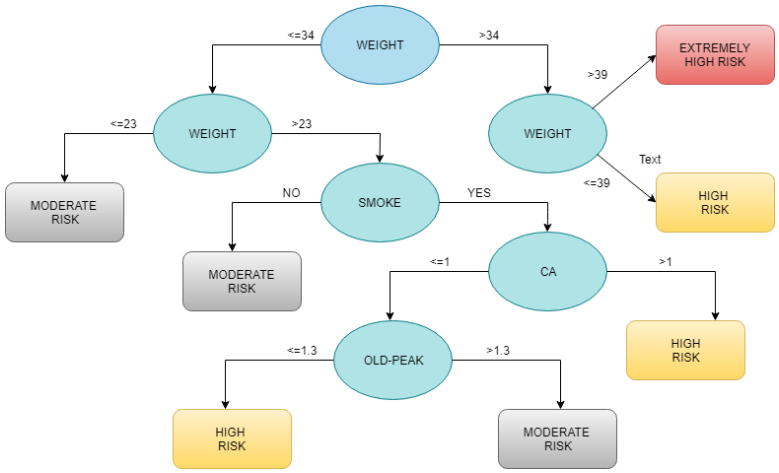


Figure 2.3: Decision tree classifier [2]

It has several advantages over other classifiers-

- It can handle both type of attributes continuous and discrete.
- It can also handle the data with missing attributes values.
- This classifier uses the Single Pass Pruning Process to restrict over fitting.

2.2.2 Analysis of Classifier

The general efficiency of any predictive model can be calculated using a number of statistical functions, that includes true positive rate, true negative rate, and accuracy. These calculations are done on the estimate of the number of risk prediction done correctly and incorrectly, which are depicted through a confusion matrix. A confusion matrix, also known as an error matrix, is a specific table layout that allows you to visualize the results of an algorithm, typically a supervised learning matrix. Each row of the matrix reflects the cases in the predicted class, while each column represents the cases in the actual class (or vice versa) as shown in figure 2.4.The following table represents the confusion matrix for this predictive model. Say, (A, B, C, D and E) are mapped to (No risk, Lw risk, Moderate risk, Hg risk and Extremely Hg risk) levels respectively. Where true positive rate (TPR) also known as the sensitivity relates to the capacity of

		Predicted Classes				
		A	B	C	D	E
Actual Classes	A	146	0	14	0	0
	B	0	32	0	3	0
	C	18	0	36	0	0
	D	0	5	0	30	0
	E	0	0	0	0	13

Figure 2.4: Actual and Predicted classes

the classifier to properly recognize a risk category, whereas the specificity or the true negative rate (TNR) relates to the capacity of the classifier to properly exclude the other risk levels (identifies the negative instances). Accuracy is the general precision of the model, which can be provided as the number of real results (i.e. both true positive and true negative) separated by the total quantity of the test set examined. These statistical methods can be represented mathematically as follows:

- **Sensitivity** =
$$\sum TruePositive / \sum ConditionPositive$$
- **Specificity** =
$$\sum TrueNegative / \sum ConditionNegative$$

- **Sensitivity** =

$$\sum Truepositive / \sum Predictedconditionpositive$$

- **Sensitivity** =

$$\sum (TruePositive + TreuNegative) / \sum ConditionPositive$$

2.2.3 AUROC curve

It stands for AUC (Area Under the curve) and ROC(Receiver operating characteristics). The precision of the test depends on how well the test group is differentiated from those with and without the disease. Accuracy is evaluated by the region of the ROC curve. Area 1 reflects a ideal test, while less than 0.5 represents a useless test.

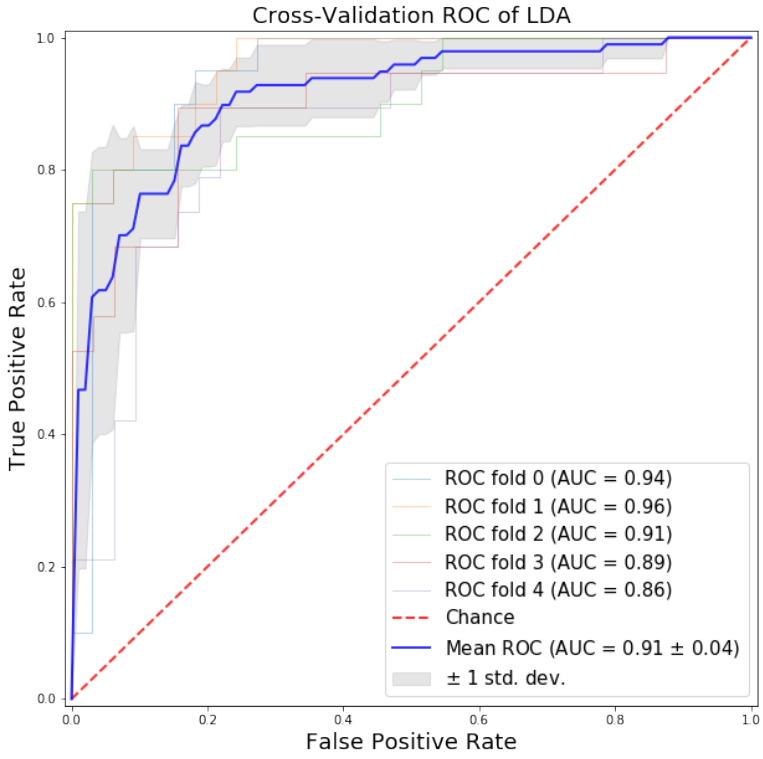


Figure 2.5: Auroc Curve

2.3 Creation of User Interface

With the help of above primary model we will create a UI which is basically a web based application that provide us a better visualization of our model.Based on the different levels of heart failure risk, we prioritize the patients from extreme level of risk to a very low level risk. In this way we can allot appointments to the cardiologist.
These are the following frameworks in which our UI is working-

- **HTML/CSS:** HTML is an acronym of hyper text markup language. HTML is at the core of each web page, regardless of the complexity of the site or the number of technologies involved. It offers a fundamental structure that is enhanced and altered by other methods such as CSS and JavaScript.
  CSS is an acronym of Cascade Style Sheets. This is used to regulate presentation, formatting, and design.

- **PHP:** It is a reverse acronym of Hypertext Preprocessor. PHP is a web scripting language that's integrated in HTML. The output of the PHP characteristics on the website is usually transmitted as an HTML code that can be read by the browser. The PHP code is converted to HTML before the page is loaded, and users can not view the PHP code on the page. This guarantees that PHP pages are secure enough to access databases and other secure information.

- **Apache:** It is the most popular software on a web server. It enables a computer to host one or more websites that can be accessed over the Internet through a web browser. The popularity of Apache in the web hosting sector is primarily owing to its open source and free use. Web hosting companies can therefore offer solutions to Apache-based web hosting at minimal price.

- **MySQL:** MySQL is a relationship open source database management system. It is based on the structure query language (SQL) used to add, delete, and modify information in the database. MySQL may be used for a variety of applications, but is most commonly found on web servers. A website that uses MySQL may contain web pages that access information from a database.

- **Javascript:** JavaScript is a more complex language than HTML or CSS. JavaScript is a programming language that allows web developers to create interactive websites. Most of the vibrant conduct you see on the web page is JavaScript. It is supported by all modern web browsers and is used for stronger and more advanced features on almost every web site.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Results

In this project, dataset acquired from the Cleveland Clinic Foundation cardiac disease information set comprising roughly three hundred instances is used. Out of these cases, 53.87 percent were for healthy people with lowest probability of heart failure, while the other 40.07 percent were for the patients with distinct rates of failure risk. Assessment of our model is shown in figure 3.1. We embrace a predictive model that can predict early the chances of heart risk failure using the C4.5 decision tree classifier and classify them into different levels.

Classes	Sensitivity	Specificity	Precision	F-Measure	ROC Area
A	0.913	0.868	0.89	0.901	0.906
B	0.914	0.98	0.865	0.889	0.966
C	0.667	0.942	0.72	0.692	0.83
D	0.857	0.988	0.909	0.882	0.966
E	1	1	1	1	1
Average	0.865	0.955	0.863	0.864	0.91

Figure 3.1: Assessment result of the model

As can be seen from the above table, the greatest level of risk detection (i.e. sensitivity) was accomplished by class E, i.e. Exceptionally strong risk of developing cardiovascular disease. This is regarded to be an effective result because this level is the most threatening and a right choice can save a patient's life. The next highest level of sensitivity was accomplished by Class B, which demonstrates the first danger of heart failure. Followed by class A which refers to healthy people. While detecting the mild amount of risk (i.e. class C) was the smallest sensitivity.
This is because this phase defines the transition from a lower risk to a greater level of risk and the beginning of a severe threat which might cause a death.

3.2 User interface

Our UI is a web based platform named as TICKERNAL. Some of the glimpse of our UI is as follows-

- **Home page**

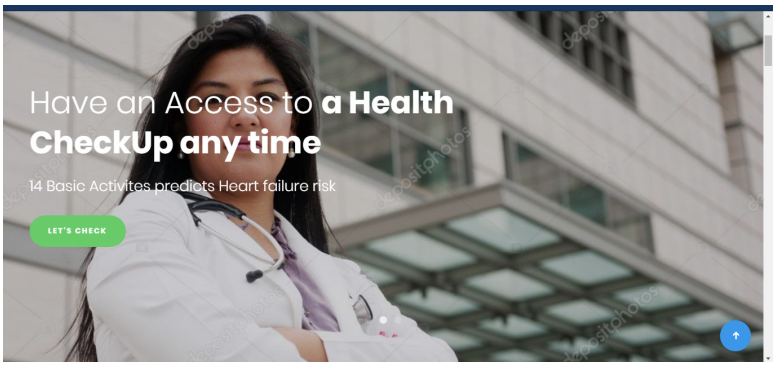


Figure 3.2: Home page of TICKERNAL

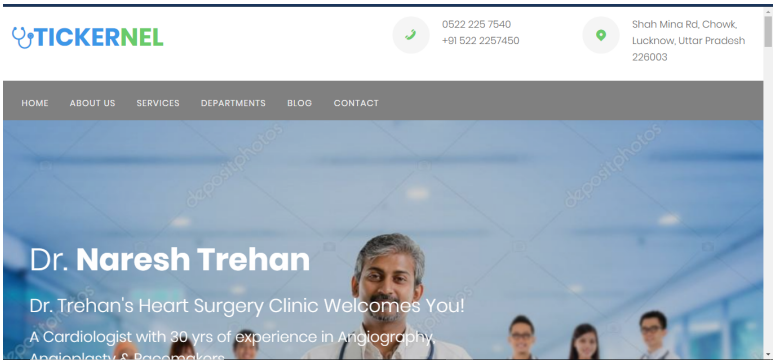


Figure 3.3: Home page (2) of TICKERNAL

- **Functionality** In this UI user can also visit the page for seeking the free advice from various health expertise. It also contains the various information about cardiovascular diseases.

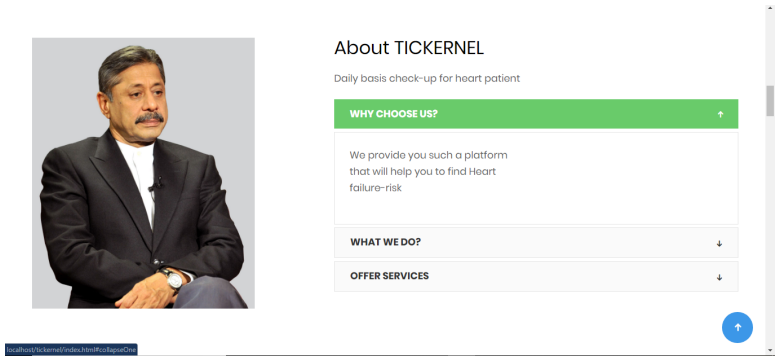


Figure 3.4: Tickernal

In this UI the users register along with their details. In registration process, user give the information which are significantly important i.e. the attributes which affect the cardiovascular diseases, as previously discussed in detail.

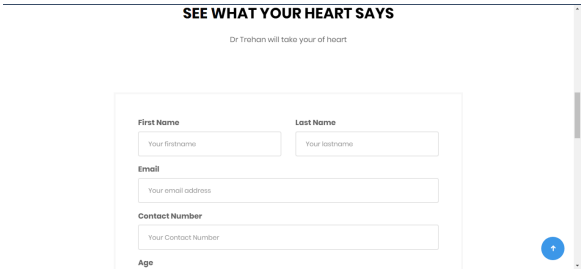


Figure 3.5: login

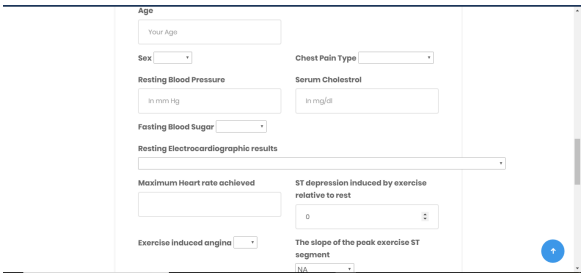


Figure 3.6: User Details

3.3 Discussion

DL and ML methods are a too early-stage detection of heart disease, it also assesses the possible risk factors, which includes-

- Severity reduction, complications delay and quality of life
- Health cost reduction.

The heart disease includes structural problems, diseased vessels, and blood clots. The most common types arecoronary heart disease, high blood pressure, cardiac arrest, arrhythmia, strokes, congenital heart disease, heart failure. The major parameter expected in the prediction is high blood glucose levels since it is the first step involved in the diagnosis of the HD patient. The articles related to drugs and therapy in HD majorly includes medication prescriptions. Therefore, there is a lot of scopes to perform research on drug and therapy with the ML and DM methods.
Machine learning and Deep learning [(13)] articles in the prediction of cardiovascular diseases enhanced accuracy above 80 percent. The most commonly used in cardiovascular disease prediction are Genetic algorithm [(14)], Artificial neural networks [(15)], supervised learning approaches [(16)] in classification [(17)], SVM, Bigdata technologies, Deep belief network [(18)], Convolution Neural network, Data mining algorithms [(19)] [(20)] .

CHAPTER 4

CONCLUSION

Heart disease has been one of the global health challenges of latest years.Currently, a number of studies have been conducted to predict and diagnose heart disease. A systematic attempt has been produced in this study to recognize and review machine learning, data mining and deep learning methods applied to cardiovascular diseases studies. EHRs in health care generating a vast quantity of technologically advanced information also offer rise to in-depth exploration of precise disease prediction, diagnosis, and therapy. The prospective advantages of implementing our model will decrease the rate of illness and death of the worldwide population. Specially, these are the advantages of our model:

(a) We are classifying the heart risk failure into different categories which is much better than binary classification .

(b) We are considering the various attributes which affects the cardiovascular diseases significantly.

(c) Our model is very reliable, user friendly. It's accuracy is much better than the existing solutions.

In this regard, the future state of our model is not only for disease prediction and diagnosis, but also for the other field of bio-informatics.

REFERENCES

[1] T. Sharma, S. Verma, and Kavita, "Intelligent heart disease prediction system using machine learning: A review," *International Journal of Recent Research Aspects*, vol. 4, pp. 94 – 97, 2017.

[2] A. Joshi, E. J. Dangra, and D. M. K. Rawat, "A critical review of various techniques for heart disease prediction," *International Journal of Technology Research and Management, ISSN*, vol. 3, pp. 773 – 778, 2016.

[3] L. BCS and T. PHH, "Deep learning for cardiac computer-aided diagnosis: benefits, issues amp; solutions," *International Journal of Computer Engineering Technology (IJCET)*, vol. 4, pp. 12–19, 2017.

[4] Carneiro, Nascimento, and Freitas, "The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods," *IEEE*, vol. 2, pp. 30–40, 2012.

[5] Akhila and Vidya, "A survey on machine learning approaches for disease predicting system," *Journal of Emerging Technologies and Innovative Research*, vol. 4, pp. 203 – 205, 2005.

[6] P. Austin, J. V., J. ho, D. Levy, and D. S. Lee, "Using methods from the data- mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure sub-types," *Journal of Clinical Epidemiology*, vol. 66, pp. 398–404, 2013.

[7] Abushariah, Alqudah, A. R, and Yousef, "A hybrid classification system of heart disease diagnosis based on the rfrs method," *Journal of Software Engineering and Applications*, vol. 1, pp. 11–18, 2017.

[8] C. Tham, C. Heng, and W. Chin, "Smartphone based ischemic heart disease risk prediction using clinical data and data mining approaches, a prototype design," *IEEE*.

[9] M. Raihan, "Smartphone based ischemic heart disease risk prediction using clinical data and data mining approaches, a prototype design," vol. 3, pp. 299 – 303, 2016.

[10] M. Ali, A. Rahman, A. Khan, A. J. M. Zhou, and J. A. Khan, "An automated diagnostic system for heart disease prediction based on 2 statistical model and optimally configured deep neural network," *IEEE*, vol. 7, 2019.

[11] R. Wijaya, A. Prihatmanto, and Kuspriyanto, "Preliminary design of estimation heart disease by using machine learning," *Joint International Conference on Rural Information Communication Technology and Electric- Vehicle Technology (rICT ICeV-T)*, vol. 1, pp. 1–4, 2013.

[12] Z. Yuan, S. Guo, and L. Han, "Disease diagnostic prediction model based on improved hybrid capso-bp algorithm," *36th Chinese Control council (CCC)*, vol. 3, pp. 11–18, 2017.

[13] Nilakshi, Waghulde, and N. Patil, "Data mining approach for prediction of heart disease using neural network," *Genetic Neural Approach for Heart Disease Prediction*.

[14] Anbarasi, Anupriya, and Iyengar, "Enhanced prediction of heart disease with feature subset selection using genetic algorithm," *International Journal Of Engineering Science And Technology*, pp. 70–76, 2010.

[15] Abushariah, M. Alqudah, Y. Adwan, and R. Yousef, "Automatic heart disease diagnosis system based on artificial neural network (ann) and adaptive neuro-fuzzy inference systems (anfis) approaches," *Journal of Software Engineering and Applications*, vol. 7, pp. 1055–1064, 2014.

[16] M. Fatima and M. Pasha, "Survey of machine learning algorithms for disease diagnostic," *Journal of intelligent learning systems and applications*, vol. 9, pp. 1– 4, 2017.

[17] H. D. Masethe and M. A. Masethe, "Prediction of heart disease using classification algorithms," *World Congress on Engineering and Computer Science*, 2014.

[18] Avendi, Kheradvar, and Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac mri," *IEEE*, 2016.

[19] M. Kumari and S. Godara, "Comparative study of data mining classification methods in cardiovascular disease prediction," *International Journal of Computer Science and Technology*, 2011.

[20] B. Kaur and W. Singh, "Review on heart disease prediction system using data mining techniques," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 2, pp. 30–48, 2014.