

# The Classification Scheme for the Heart Disease Prediction

\*Saloni Kapoor, #Ashwinder Tanwar

\*M.E.Student, Assistant Professor, University Institute of Engineering(CSE), Chandigarh University, Gharuan, Mohali, Punjab, India.

\*salonikapoor31@gmail.com, #ashwinder.cse@cumail.in

**Abstract—** at present time, the diseases in the field of medical is day by day increasing and its analysis must be done at an extremely early stage so that it can be handled properly. Moreover, the accuracy of analysis is deduced whenever there is incomplete data of healthcare. On the other hand, different areas show extremely unique features on regional healthcare diseases through which the prediction of diseases erupts. The paper describes about the machine learning algorithms' so that the chronic diseases can be easily predicted. In this paper, the description about the data mining has been depicted. Moreover, the need of the heart disease prediction has also been explained. The SVM algorithm has been implemented for predicting the disease as this algorithm will improve the efficiency and disease can be predicted will less amount of time.

**Keywords:** Support Vector Machine, Data Mining, Machine Learning Algorithms, Heart Disease Prediction.

## I. INTRODUCTION

### 1.1 Data Mining

Nowadays, the complete data is stored in different applications, files and databases as the classified and the private information ought not to be put away anyplace else. Thus, it is critical to find and propose a framework wherein every information and data [1] can be put away securely and safely. Some time it turns out to be exceptionally hard for the clients to concentrate and utilize just applicable information from this huge information. From this circumstance, Data Mining is utilized. It is a strategy for choosing, picking and evacuating just the information which is valuable and important for that specific moment of time. It enables the client to get to their information whenever and from anyplace. Information mining is normally known as "Knowledge- Discovery-in Databases" (KDD). There are enormous measure of significant and unimportant information being put away on the databases and numerous different zones. It has offered to the term Data Mining, which can be further helpful in basic decision process. It is the wonder of separating valuable and significant records from the gigantic amount of information being collected wherever on the web. This demonstrates the most significant motivation behind information mining is to gather and pick important information from the recently put away information. There are tremendous kinds of informational indexes being utilized in data mining.

**It consists of repetitive sequence of steps:**

- 1. Data Cleaning:** It removes the unwanted noises and inconsistent data.
- 2. Data Integration:** It combines multiple data sources.
- 3. Data Selection:** in this step, the information which is appropriate is retrieved plus extracted from the databases.
- 4. Data Transformation:** in this, the data is changed into its most reasonable structure which should be possible by performing activities.
- 5. Data mining:** It is a most significant advance, wherein exceptionally propelled techniques are utilized for the extraction of information.
- 6. Knowledge Presentation:** In this, representation and exhibit strategies are utilized to demonstrate the mined information to the customer.

### "MACHINE LEARNING ALGORITHMS":

The highly used machine learning library is Apache Mahout that allows developers to use the algorithms that are highly optimized. It generally uses the most popular techniques that include the clustering, classification and recommendation. The term machine learning is used that deals with science and includes programming such that the systems without human intervention learns and also improves just by its experience. The term learning here refers to understand the inputs along with this make best decisions that are completely based on the data supplied. Machine learning itself is a very vast domain and the techniques that are used to differ are supervised learning and unsupervised learning.

- **Supervised learning:** this type of learning is used in learning functions that are present from training data. It uses various algorithms that include SVMs, neural networks along with this Naïve bayes' classifiers.
- **Unsupervised learning:** in this learning there is no particular dataset along with this there is no proper training person. This is an exceptionally powerful tool since it helps in analyzing the data and also helps in getting the trends and patterns. The most common usage of this type of learning is that it is used for making clusters in which all the similar inputs are grouped together. This learning methods includes algorithms such as k-means, hierarchical clustering plus self' organized' maps.
- **Convolutional - Neural' -Network':** this is the network that is comprised of more number of convolutional layers which are further followed by the fully connected MNP (Multilayer' neural Network'-) the architecture of CNN- is simply designed so that the benefits of 2D structure can be taken that includes the image inputs. This can simply be attained by using local connections along with the knots tied that can be easily followed by pooling'. The CNN are extremely easy to use since they have very less connections that multilayer networks that have similar hidden units [2].

## 1.2 Heart Diseases Prediction

The major cause of a person's death is a Heart Disease. Smoking, utilization of liquor in huge amount, cholesterol, is the major reasons of heart diseases. The principle functioning of a body is a Heart, if it won't work appropriately, at that point it will straightforwardly affect the working of the other body parts. A portion of the central point prompts the heart sicknesses are hypertension, high rate of cholesterol, age, family ancestry, horrible eating routine and some more. Smoking is one of the real reasons for heart ailments. Various sorts of information digging strategies are utilized for the expectation of an illness like Naïve Bayes, K-Nearest Neighbor (KNN) calculation, Neural Network, Decision tree. In 'KNN' calculation K is utilized to discover the estimations of variables that lead to heart sicknesses. 'Decision tree' is utilized to sort out arranged report on the heart enduring patients. The naïve' bayes' algorithm is utilized for predicting the heart diseases probability. Last yet not the least, the 'neural systems' are utilized to limiting the mistakes happened during prediction. By utilizing every one of these strategies, the records are named well [3] and kept up routinely. With the assistance of every one of these classifiers doctors are skilled to foreseeing the heart ailment at the underlying stage.

## II. LITERATURE REVIEW

**Sarath Babu et .al (2017)** introduced [4] a strategy called information mining, which is the technique of finding new arrangement of data from gigantic measure of information.

It is utilized to examine huge volume of information and the examples were separated to change over the superfluous data into valuable data. This gathered data is fed into a few classifiers like KNN', naïve' bayes, SVM' and decision' trees, every one of them play out some particular undertakings. These systems are utilized to anticipate heart illnesses at their beginning time. It indicates powerful execution so as to accomplish the analyses for the heart related infections. There are several benefits of this methodology, for example, the sicknesses can be anticipated at their extremely starting stages and can be analyzed accurately and appropriately on schedule. Hence, the specialist inferred that, this strategy is exceptionally valuable in anticipating heart related issues.

**Priyanga et.al (2017)** proposed an intelligent and efficient technique called naïve bayes for the prediction of heart related issues [6]. The data is collected from the given attributes and then they are implemented as web based applications. All the methods which are already been discovered and used don not show effective and satisfactory results for the prediction of heart diseases. But all the techniques are introduced for the very same purpose that is for predicting the severe diseases like cardiac crest, cancer, brain tumor. But they were failed to discover the diseases at their initial stage. Therefore, the researcher concludes that the approach classifies that it has low cost and extensively tested by experienced cardiologists. The research mainly focuses on detection of heart disease using UCI dataset.

**Tülay Karayölan et.al (2017)** recommended back propagation algorithm for the forecast of heart illnesses with the assistance of 'artificial neural' systems [5]. It has some clinical highlights where neural systems are utilized as info and is prepared alongside this proposed back propagation method. The prediction can be up to 95%. The methods that have been already proposed were not so enough to predict the diseases. Therefore, the improvement in predicting the heart diseases has emerged the usage of machine learning methods. Along these lines, the specialist reaches the inference that the proposed methodology has practically 100% exactness in forecast heart related infections at their beginning periods. It gives better outcomes in contrast with different systems.

**Priyanka N et.al (2017)** proposed [7] two strategies for information mining that is naïve algorithm and decision' algorithm for predicting the ailment like cerebrum tumor, and heart assault as well as cancer. The outcomes and exactness of both the presented techniques are looked at and afterward chooses which one is better and most broadly utilized. Because of inadequate assets accessible in therapeutic and wellbeing district, it was very hard for predicting the extreme sicknesses on schedule. It is imperative to give great facilities to the emergency clinics and the patients. Data mining is extraordinarily intended to defeat this issue. As a result, the investigation of the paper reasons that the both the techniques were looked at and

results that both are appropriate and best at various fields, likewise gives progressively precise and productive outcome for heart illnesses identification.

### III. OBJECTIVE OF THE WORK

Following are the various objectives of this research work:-

1. To learn and examine different methods for heart disease prediction
2. To Implement the classification approach for the heart disease prediction
3. Analyze the performance of classification approach in terms of certain parameters

### IV. METHODOLOGY

#### 4.1. Research Design

The research design of this work has the various phases and these phases are literature survey, problem identification, research methodology and implementation. The research process is described in the figure 1.

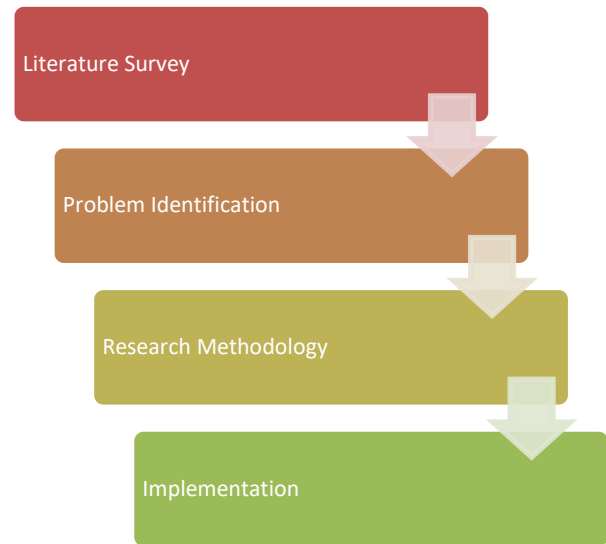
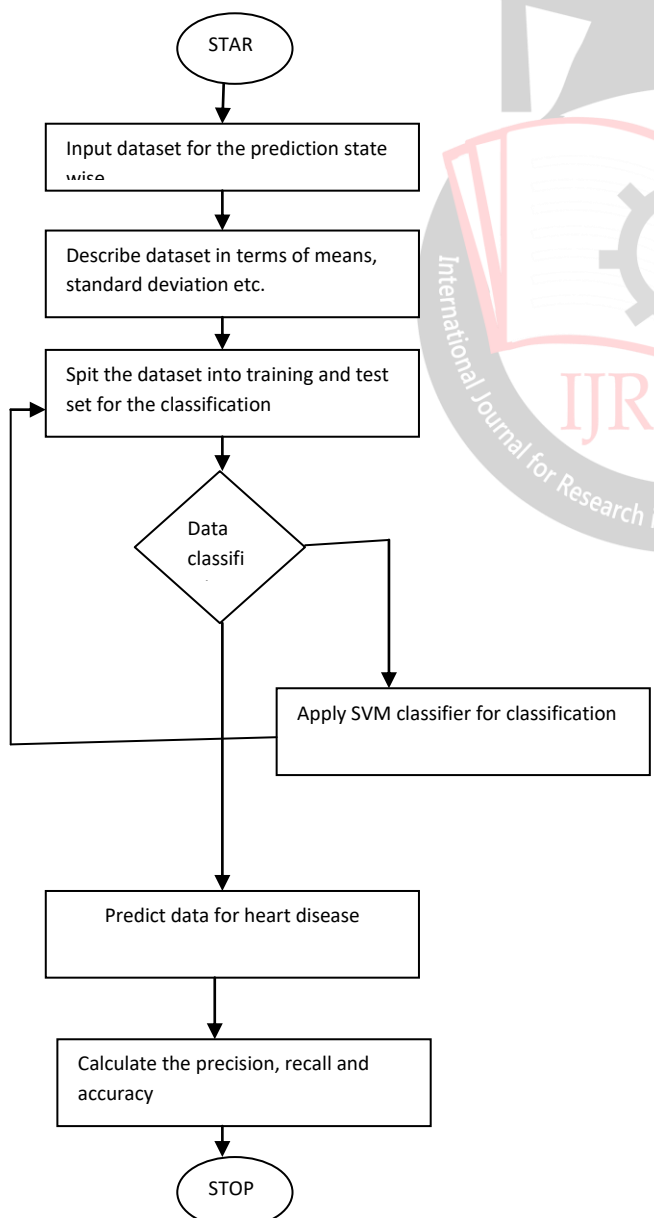


Fig 1. Research Design

The research design of this work has the various phases and these phases are literature survey, problem identification, research methodology and implementation. The research process is described in the figure 1.



#### 4.2. Research Process

Following are the various steps of the research process:-

- 1. Pre-processing:** The pre-processing is the first phase of the heart disease prediction research process. In this phase, the dataset is taken as input and data set is cleared by removing missing values
- 2. Feature Extraction:** In the second phase, the features of the input data are extracted for the classification. In this phase, the relationship between the various attributes are established for the identification of main attributes from the dataset
- 3. Classification:** In the last phase, the relationship between the attributes is taken as input for this phase, the input data is divided into training and test set. The approach of classification is applied for the prediction. The performance of planned approach will be analyzed on the basis of certain parameters like accuracy, precision, f measure as well as recall.

### V. RESULT AND ANALYSIS

#### 5.1. Tools used

A high-level programming language that includes dynamic semantics within it is known as python. The Rapid Application Development finds this tool very attractive due to the high-level built-in data structures that are integrated along with dynamic typing and binding provided in them. Further, the existing components can be linked together using scripting or glue language. Readability is efficient here due to the ease of implementation and easy-to-learn syntax this tool provides. Thus, the cost required for maintenance is minimized here. The program modularity and code reutilization provided by certain modules and packages are supported by python. Without modifying any major platforms, python interpreter and extensive standard library are provided in the source form. It is possible to distribute these platforms freely in the applications. Since

this tool provides increase in overall productivity of the application, Python is highly used by several programmers. There is high speed of performance of edit-test-debug cycle since no compilation step is included here. It is easy to debug python programs. A segmentation failure will never occur in case when a bug or bad input is generated. In fact, an exception is raised by the interpreter when it recognizes an error in the program. A stack trace is printed by the interpreter in case when the exception is not caught by program. The local and global variables are inspected by the source level debugger along with the setting up of breakpoints and evaluation of arbitrary expressions.

### 5.2 Results and screenshots of the work performed

```

age          sex      chest_pain  blood pressure  serum_cholesterol \
count 303.000000 303.000000 303.000000 303.000000 303.000000
mean 54.438944 0.679868 3.158416 131.689769 246.693069
std 9.038662 0.467299 0.960126 17.599748 51.776918
min 29.000000 0.000000 1.000000 94.000000 126.000000
25% 48.000000 0.000000 3.000000 120.000000 211.000000
50% 56.000000 1.000000 3.000000 130.000000 241.000000
75% 61.000000 1.000000 4.000000 140.000000 275.000000
max 77.000000 1.000000 4.000000 200.000000 564.000000

fasting_blood_sugar  electrocardiographic  max_heart_rate \
count 303.000000 303.000000 303.000000
mean 0.148515 0.990099 149.607261
std 0.356198 0.994971 22.875003
min 0.000000 0.000000 71.000000
25% 0.000000 0.000000 133.500000
50% 0.000000 1.000000 153.000000
75% 0.000000 2.000000 166.000000
max 1.000000 2.000000 202.000000

induced_angina  ST_depression  slope  diagnosis
count 303.000000 303.000000 303.000000 303.000000
mean 0.326733 1.039604 1.600660 0.937294
std 0.469794 1.161075 0.616226 1.228536
min 0.000000 0.000000 1.000000 0.000000
25% 0.000000 0.000000 1.000000 0.000000
50% 0.000000 0.800000 2.000000 0.000000
75% 1.000000 1.600000 2.000000 2.000000
  
```

Fig 2: Data Description

As shown in figure 2, the dataset of heart disease is taken from the UCI repository. In this figure, dataset is described which mean value, standard deviation etc.

```

The Accuracy is
80.0
Success Ratio is
0.8
The execution time is 2.024115800857544 seconds ---
  
```

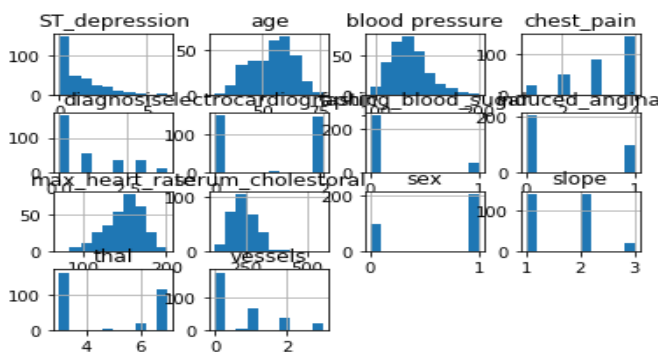


Fig 3: Apply Existing Approach

As shown in figure 3, the dataset for the heart disease prediction is taken from UCI repository. The k-mean algorithm is applied with the SVM classifier and its accuracy is 80 percent. The basis k-mean algorithm is applied which is not able to drive exact relationship between attributes and due to which accuracy is less.

### REFERENCES

[1] Nidhi Bhatla, Kiran Jyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques", IJERT, Vol 1, Issue 8, 2012.

[2] Syed Umar Amin, Kavita Agarwal, Rizwan Beg, "Genetic Neural Network based Data Mining in Prediction of Heart Disease using Risk Factors", IEEE, 2013.

[3] Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", IJCA, Volume 47- No.10, June 2012.

[4] Sarath Babu, Vivek EM, Famina KP, Fida K, Aswathi P, Shanid M, Hena M, "Heart Disease Diagnosis Using Data Mining Technique," 2017, IEEE

[5] Tülay Karayölan, Özkan KÖlÖç, "Prediction of Heart Disease Using Neural Network," 2017, IEEE

[6] Priyanga and Dr. Naveen, "Web Analytics Support System for Prediction of Heart Disease Using Naïve Bayes Weighted Approach (NBwa)," 2017, IEEE

[7] Priyanka N and Dr. Pushpa RaviKumar, "Usage of Data mining techniques in predicting the Heart diseases – Naïve Bayes & Decision tree," 2017, IEEE