MACHINE LEARNING FOR PREDICTION AND DIAGNOSTICS OF CARDIOVASCULAR DISEASE

Supervised by:

Prof. L.A. Patil

Created by:

Shashwat Babhulgaonkar

Pritam Bendkule

Jayesh Suryawanshi

Vivek Sanap

WHY HEART DISEASE

- Heart Disease is one of the most fatal disease.
- Results in 12 million deaths annually.
- Impact of heart disease is so tackle it rarely gives a chance to tackle a situation.
- One person dies every 33 second due to heart disease in India.

LITERATURE SURVEY

1) The healthcare industry collects huge amount of healthcare data which, unfortunately, are not mined to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited.
2) Association rules represent a promising technique to improve heart disease prediction. Unfortunately, when association rules are applied on a medical dataset, they produce an extremely large number of rules.
3) The healthcare environment is still 'information rich' but the 'knowledge is poor'. There is wealth of data is available in the system but the lack of effective analysis tools to discover hidden relationship in the data.

DRAWBACKS OF THE EXISTING STSTEM

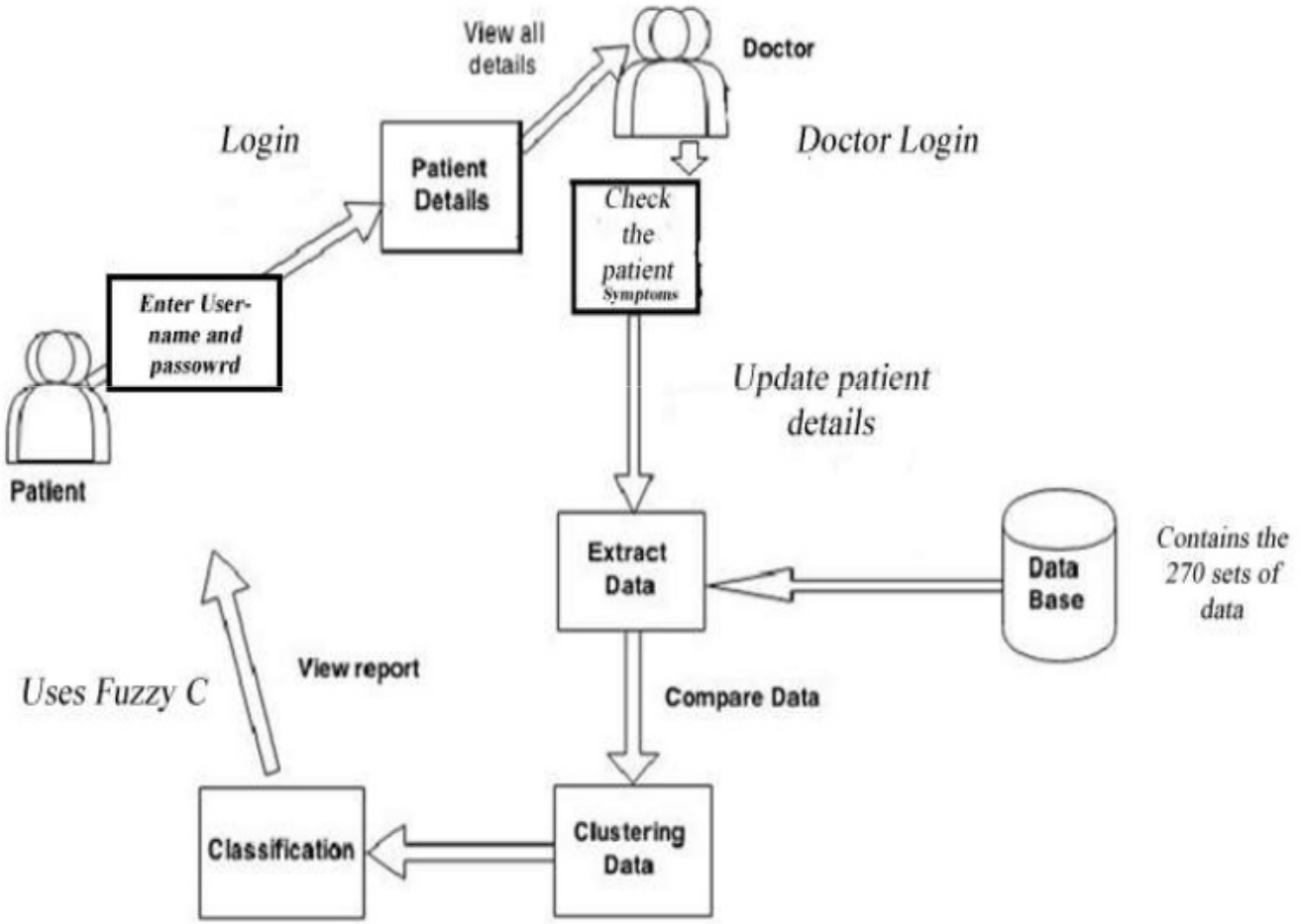
- Very few systems use the available clinical data for prediction purposes and even if they do, they are restricted by large number of association rules that apply.
- Diagnosis of the condition solely depends on the doctor's intuition and patients records.

- **The disadvantage are:**
- Detection is not possible at an earlier stage.
- In the existing system, practical use of various collected data is time consuming.

Problem Statement

- To implement a system 'S' for **Prediction and Diagnosis** of **Cardiovascular Diseases 'C'** using **Machine Learning Algorithm 'A'** with Dataset 'D'.

Block Diagram



MOTIVE:

- The main motive of this research is to develop cost effective prediction of said disease using machine learning techniques.
- As all medical experts don't competence in every field and even in some place there is a shortage of medical experts.
- In that case prediction results are used along with experts advice to reduce the chances of undesirable results.

DATA COLLECTION AND PREPROCESSING:

- Data is collected from different resources.

- Data Preprocessing includes:
- **Data Transformation**: Target variable in raw data contain five different values(0,1,2,3,4). Here all non zero values are target variable is transformed to 1 to make this problem binary class problem.

- **Data Cleansing**: Few instances in raw data are missing.

DATASET USED:

- **Dataset used for prediction include:**

- Age
- Gender
- CP(Chest Pain)
- Blood Pressure
- Cholesterol
- Fasting Blood Sugar
- Restecg

FEATURE SELECTION:

- This is approch of choosing best subset of features which contain least number of dimension and contribute most to improve the performance.

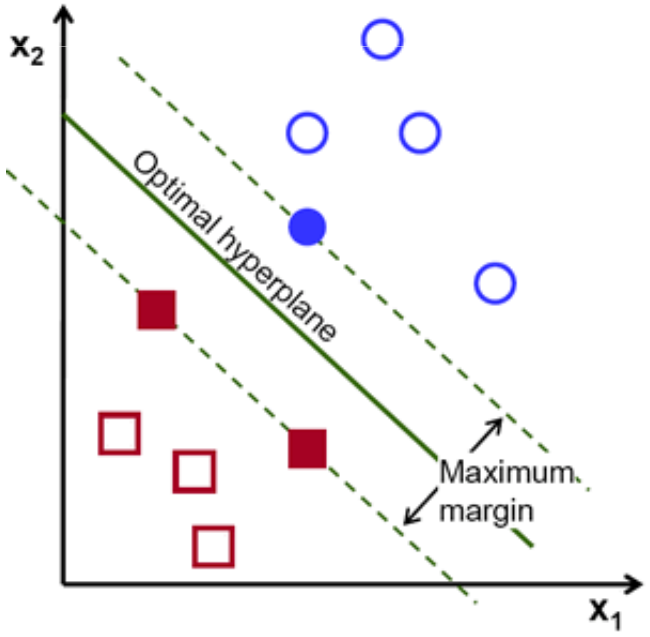
- **Features are categorized**:
  - Relevant Features
  - Redundant Features
  - Irrelevant Features

METHODOLOGY USED FOR SELECTING FEATURES:

- The algorithm Support vector Machine(SVM) used to compute score and rank of each feature individually.
- SVM is used for rank calculation.
- Gain Ratio is used for score calculation.

ALGORITHMS USED:

- SVM : Support Vector Machine recursive function elimination. Basic idea of this algorithms is to find optimal hyperplane.



CLASSIFICATION:

- Naive Byes Algorithm is applied to subset of features which contribute best to accuracy.

- **Naïve Byes**: This is based on Byes theorem.
- Byes theorem is written as:

  P(X/Y)=(P(Y/X)*P(X))/P(Y)
- Naïve Byes have capability to trained even on smaller datasets.
- Easy to implement.

CONCLUSION:

- Recognizing the disease is mainly the purposed approach which can recognize the heart disease with little computational effort.
- The approach can be used for medical applications like detection and classification of heart diseases with suitable classifier.
- The proposed system will describe a possible and feasible approch for diagnosis.
- It will help to detect heart diseases effectively and efficiently at early stages before it lead to disastrous consequences.


Thank
you