International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 12 | Dec -2016 www.irjet.net p-ISSN: 2395-0072

# Hidden pattern analysis for heart disease classification

**Chavan Krushna D,Kale Abhijeet A ,Kulkarni Swapnil P, Sayyed Ajmeer D.**

*Department of Computer Engineering Savitri-Bai Phule Pune University SBPCOE,indapur,Maharashtra,India*

---

**Abstract -** *Prediction of Heart Disease using technique of Data Mining is effective but there is loss of Accuracy by using the image processing as additional processing for more Accuracy . In Proposed System we are using the algorithms like Decision Tree, Nueral Network and the Naive Bayes in the data mining and in the image Processing we are using the popular algorithms like Local Binary Pattern. The research result shows prediction accuracy of 99 Percent. Data mining enable the health sector to predict patterns in the datasets. Here we use image processing for comparing the ECG, CT scan, Angiography, etc, reports and finding the more accurate results.*

***Key Words*: Data mining, Image databases, Clustering, Query formulation, Search process, Selection process, Image Processing, Image displays**

## 1.INTRODUCTION

Prediction of Heart Disease is the difficult task in the Medical Science Field. In this paper, we can use the Data mining classification algorithm for heart disease prediction. Thus there is need arises to develop a decision support system for detecting possibility of heart disease can be occurred to patient. Today medical field takes lots of time to treat patients with various kind of diseases. Among the most threatening one is the Heart disease which cannot be observed with a open eye and comes instantly when its limitations are reached. Bad clinical decisions would cause death of a patient which cannot be affordable by any hospital. To achieve a correct and cost effective treatment computer-based and support Systems can be developed to make good decision. Many hospitals use hospital information systems to manage their healthcare or patient data. The main objective of this paper is to develop a prototype which can determine and extract unknown knowledge (patterns and relations) related with heart disease from a past heart disease database record. It can solve complicated queries for detecting heart disease and thus assist medical practitioners to make smart clinical decisions which traditional decision support systems were not able to. By providing efficient treatments, it can help to reduce costs of treatment.

## 2. RELATED WORK

The Researchers[1] uses the data mining methods like Neural Network, Decision Tree, Naive Bayes for finding connection and pattern between large amount of health care data. As huge amount of data is produced in medical associations(healing facilities, therapeutic focuses).

The Researchers[2] applied the algorithms such as J48, Naïve Bayes, REPTREE, CART, and Bayes Net for predicting the heart attacks. By applying this algorithms the result shows prediction accuracy of 99%.

The Researchers[3] objective is to evaluate different classification techniques in heart disease diagnosis. Classifiers algorithms like J48 Decision Tree, K Nearest Neighbors, Naive Bayes, and SMO are used to classify dataset. After classification, some performance evaluation measures like accuracy, precision, sensitivity, specificity, F-measure and area under ROC curves are evaluated and compared. The comparison results show that J48 Decision tree algorithm is the best classifier algorithm for heart disease diagnosis on the existing datasets.

The Researchers[4] objective is to provide a study of different data mining techniques that can be employed in automated heart disease prediction systems which is used to predict the heart disease from the different data sets. Various techniques and data mining classifier algorithms are defined in this paper which has emerged in recent years for efficient and effective heart disease diagnosis. The analysis shows that Neural Network with 15 attributes has shown the highest accuracy i.e. 100% so far that means it provides exact result. On the other hand, Decision Tree has also performed well with 99.62% accuracy by using 15 attributes i.e.it is also precisely accurate. Moreover, in combination with Genetic Algorithm and 6 attributes, Decision Tree has shown 99.2% efficiency i.e. all algorithms effectively provides the best result.

The Researchers[5] presents a hybrid swarm intelligence optimization algorithm called GA-ACO algorithm that combines the evolution ideas of genetic algorithms and ant colony optimization based on compensation for solving healthcare problem. Thus the GA-ACO model along with the extends definition for identifying heart disease provides a good classification accuracy based models.

The Researchers[6] says that in health concern business, data mining plays a important task for predicting

diseases. Numeral number of tests must be requisite from the patient for detecting a disease . However using data mining technique can reduce the number of test that are required to predict th diseases. Cardiovascular disease is the principal source of deaths wide spread and the prediction of Heart Disease is an important phase. In order to reduce number of deaths from heart diseases there have to be a quick and efficient heart disease prediction technique. The principle of this study is to extract hidden patterns by applying data mining techniques, which are noteworthy to heart diseases, from a data collected together by an International Cardiovascular Hospitals.

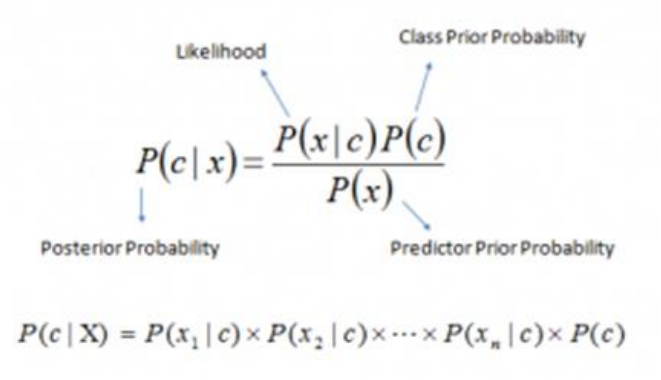
## 3. DATA MINING

Data Mining is the way toward breaking down information from alternate points of view and compressing it into helpful data - data that can be utilized to build income, cuts costs, or both. Information mining programming is one of various systematic instruments for breaking down information. It permits clients to investigate information from a wide range of measurements or edges, sort it, and outline the connections recognized. In fact, information mining is the way toward discovering connections or examples among many fields in substantial social databases.

### 3.1 Naive Bayes Classifier:

It is an order method in light of Bayes' Theorem with a suspicion of autonomy among indicators. In basic terms, a Naive Bayes classifier accept that the nearness of a specific element in a class is disconnected to the nearness of some other element. For instance, an organic product might be thought to be an apple in the event that it is red, round, and around 3 creeps in distance across. Regardless of the possibility that these elements rely on upon each other or upon the presence of alternate components, these properties autonomously add to the likelihood that this natural product is an apple and that is the reason it is known as 'Innocent'

Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.Bayes theorem provides a way of calculating posterior probability P(c|x) from P(c), P(x) and P(x|c). Look at the equation below:




Above,

$P(c|x)$ is the posterior probability of *class* (c, *target*) given *predictor* (x, *attributes*).

$P(c)$ is the prior probability of *class*.

$P(x|c)$ is the likelihood which is the probability of *predictor* given *class*.

$P(x)$ is the prior probability of *predictor*.

### 3.2 Decision Tree:

Decision tree assembles characterization or relapse models as a tree structure. It separates a dataset into littler and littler subsets while in the meantime a related choice tree is incrementally Developed. The last outcome is a tree with choice hubs and leaf hubs. A choice hub (e.g., Outlook) has at least two branches (e.g., Sunny, Overcast and Rainy). Leaf hub (e.g., Play) speaks to a grouping or choice. The highest choice hub in a tree which compares to the best indicator called root node.Decision trees can deal with both clear cut and numerical information.

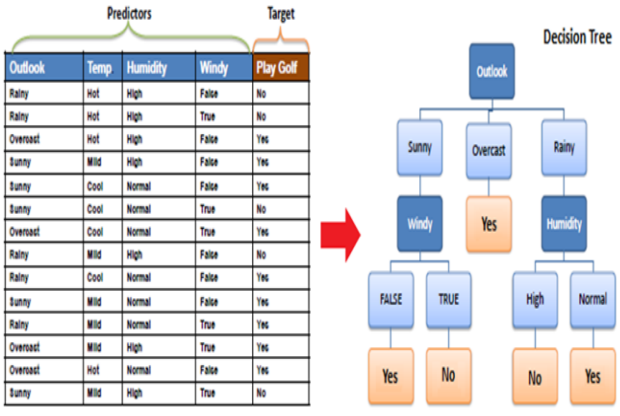



**Fig-1**: Example of Decision Tree. *(figure caption)*

### 3.1 Neural networks*:*

The field of Artificial Neural Networks (ANN) is worried with the examination of computational models propelled by hypotheses and perception of the structure and capacity of organic systems of neural cells in the mind. They are for the most part composed as models for tending to numerical, computational, and building issues. All things considered, there is a great deal of interdisciplinary research in arithmetic, neurobiology and software engineering.

An Artificial Neural Network is for the most part contained an accumulation of simulated neurons that are interconnected with a specific end goal to plays out some calculation on info designs and make yield designs. They are versatile frameworks fit for altering their inward structure, commonly the weights between hubs in the system, permitting them to be utilized for an assortment of capacity estimation issues, for example, arrangement, relapse, highlight extraction and substance addressable memory. Given that the concentration of the field is on performing calculation with systems of discrete registering units, the field is customarily called a "connectionist" worldview of Artificial Intelligence and 'Neural Computation'.

## 4.Image processing

Image processing is the study of any algorithm that takes an image as input and return an images as output Image processing usually refers to digital image processing, but optical and analog image processing also are possible.

Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a sort of flag regulation in which info is picture, similar to video edge or photo and yield might be picture or attributes connected with that picture. Typically Image Processing framework incorporates regarding pictures as two dimensional signs while applying effectively set flag handling techniques to them.

### 4.1 Local Binary Pattern:

The LBP feature vector, in its simplest form, is created in the following manner:

Divide the inspected window into cells (e.g. 16x16 pixels for every cell).

For every pixel in a cell, contrast the pixel with each of its 8 neighbors (to its left side top, left-center, left-base, right-beat, and so on.). Take after the pixels along a circle, i.e. clockwise or counter-clockwise.

Where the middle pixel's esteem is more prominent than the neighbor's esteem, compose "0". Something else, compose "1". This gives a 8-digit double number (which is normally changed over to decimal for comfort).

Compute the histogram, over the phone, of the recurrence of every "number" happening (i.e., every mix of which pixels are littler and which are more noteworthy than the middle). This histogram can be viewed as a 256-dimensional component vector.

Optionally standardize the histogram.

Concatenate (standardized) histograms of all cells. This gives an element vector for the whole window.

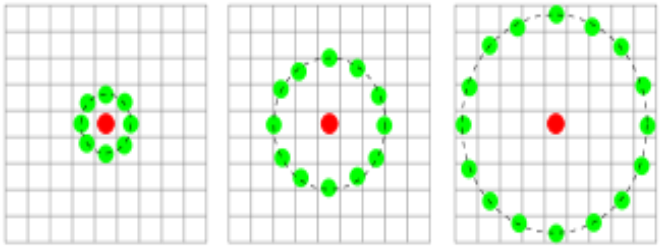


**Fig 3:** Three neighborhood examples used to define a texture and calculate a local binary pattern (LBP)

The component vector can now be handled utilizing the Support vector machine or some other machine-learning calculation to group pictures. Such classifiers can be utilized for face acknowledgment or surface examination.

A helpful augmentation to the first administrator is the purported uniform pattern, which can be utilized to lessen the length of the component vector and actualize a basic turn invariant descriptor. This thought is propelled by the way that some twofold examples happen more regularly in surface pictures than others. A nearby parallel example is called uniform if the paired example contains at most two 0-1 or 1-0 moves. For instance, 00010000(2 moves) is a uniform example, 01010100(6 moves) is most certainly not. In the calculation of the LBP histogram, the histogram has a different receptacle for each uniform example, and all non-uniform examples are alloted to a solitary canister. Utilizing uniform examples, the length of the element vector for a solitary cell lessens from 256 to 59.

## 5. SYSTEM ARCHITECTURE:

In Proposed System Patients Data taken as input and stored in Dataset and this Dataset is Compared with the other Dataset Which contain the patients data who is suffering From the Heart Disease. Here we are using combination of Pathology results and graphical reports like ECG, Angiography reports. In proposed System there are two Techniques first is Data Mining for Handling Pathology results and second technique is image processing which is used to handle graphical reports these techniques are used for more Accurate Prediction of Heart Disease

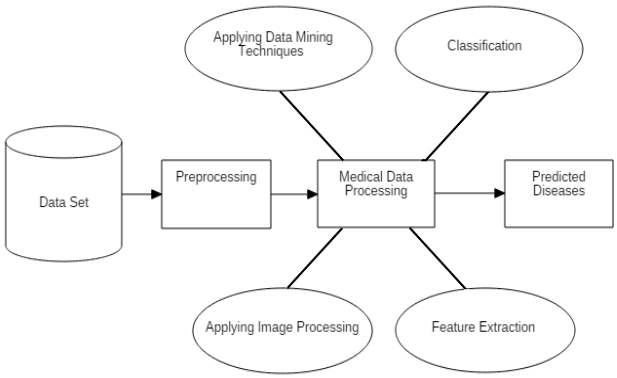



**Fig 4:** System Architecture

## 6. ADVANTAGES:

1. Because of Image processing and Data mining it gives more Accuracy.
2. As compare to existing system it is faster

## 7. CONCLUSIONS

In This Paper, We studied how data mining technique brings with set of techniques to find out hidden patterns for making decision in healthcare organization.We focused on classification methods of data mining used in data discovery.different classification techniques of data mining have merits and demerits for data classification and knowledge extracation.Furthermore, neural network, decision tree or navie Bayes can be studied in more detail to Implement an algorithm that is helpful in healthcare organization.

## REFERENCES

[1] K Raj Mohan, Ilango Paramasivam ,Subhashini, SathyaNarayan " Prediction and Diagnosis of Cardio Vascular Disease – A Critical Survey",2014 World Congress on Computing and communication technology

[2] Ms. Priti V. Wadal, Dr. S. R. Gupta, "Predictive Data Mining For Medical Diagnosis: An OverviewOf Heart Disease Prediction "International Conference on Industrial Automation and Computing (ICIAC- 12-13th April 2014)

[3] Aqueel Ahmed, Shaikh Abdul Hannan, "Data Mining Techniques to Find Out Heart Diseases: An Overview", International Journal of Innovative, Technology and Exploring Engineering (IJITEE), September 2012.

[4] K.Srinivas B.Kavihta Rani Dr. A.Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", (IJCSE) International Journal on Computer Science and Engineering, 2010.

[5] Deepali Chandna, "Diagnosis of Heart Disease Using Data Mining Algorithm", 1678-1680, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014.

[6] K.Sudhakar, Dr. M. Manimekalai, " Study of Heart Disease Prediction using Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 1, January 2014.

[7] Qeethara Kadhim Al-Shayea, "Artificial Neural Networks in Medical Diagnosis", IJCSI International Journal of Computer Science Issues,Vol. 8, Issue 2, March 2011

[8] Ms.Rupali R.Patil," Heart Disease Prediction System using Naive Bayes and elinek-mercer smoothing", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 5, May 2014.