

# Smart Health Prediction System Using Data Mining

Pradnya Suresh Joshi

*Deogiri Institute of Engineering and Management Studies,  
Aurangabad, India*

Asst.Prof. Ashwini Gaikwad

*Deogiri Institute of Engineering and Management Studies,  
Department of Computer Science,  
Aurangabad, India*

**Abstract** - In this paper, we present the techniques and applications of data mining in medicinal and instructive parts of Clinical Predictions. In health care fields, a huge quantity of information is turning into accessible due to availability of computers. Such an large amount of information can't be processed to make health predictions in the early stage and make treatment schedules to diagnose. The aim is to assess the techniques of data processing in the fields of health care to develop correct choices for disease prediction. Data mining technique is a latest powerful technology that is of high interest in the computer world. It uses already existing data in several databases to rework it into new researches and results. From available data sets, to extract new patterns and the knowledge related to these patterns data mining uses machine learning and database management. This system supports an end user and online consultation. Here we propose a framework that enables clients to get moment direction on their medical problems through an smart health prediction system.

**IndexTerms:** DataMining, Healthcare, classification, Disease, Prediction, Prescription.

## I. INTRODUCTION

It might have happened so many times that we or someone ours need doctors help immediately, but they are not available due to some reason. The Health Prediction system is an end user support and online consultation project for correct prediction of illness based on patients input. Here propose system allows users to get instant guidance on their health issues through an intelligent health care system online. The smart health prediction system is fed with various symptoms and the disease/illness associated with those systems. The system allows user to share their symptoms and issues then system processes patients symptoms to check for various illness that could be associated with it.

Here some intelligent data mining techniques to guess the most accurate illness that could be associated with patient's symptoms. If the system is unable to provide suitable results, it informs the user about the type of disease or disorder it feels user's symptoms are associated with. If patients symptoms do not exactly match any disease in our database, it shows the diseases user could probably have judging by his/her symptoms. Disease prediction using patient input symptoms

history and health data by applying data mining and machine learning techniques is ongoing struggle for the past decades.

### Necessity:

Sometimes we need the help of doctors immediately, but due to some reasons they unavailable. In project proposed system is user favourable to get guidance on health issues instantly through online health care system.

The System is helpful in emergency of patients by suggesting the doctors and immediate prescriptions on their disease. patient can get help from anywhere at any time. In medical fields, the foreign students have solved some medical issues that are laborious to be settled in classic statistics by classification of Bayesian. Without an extra information, classification rules are generated by the samples trained by themselves.

## II. LITERATURE SURVEY

### 2.1 Research work

Classification algorithm is one of the greatest significant and applicable data mining techniques used to apply in liver, kidney and heart disease prediction. Classification algorithm is the most common way in several automatic medical health diagnoses. Many of them show a good classification accuracy listed below[1].

The paper "Smart E-Health Prediction System Using Data Mining" applies the data mining process to predict hypertension from patient medical records with eight other diseases [2]. The reference paper "A Health Prediction Using Data Mining" Predict Liver Disorder with methods C4.5, NBC in 2016. The result shows NBC algorithm has the highest accuracy [3]. It analyzed the Liver Disorder Using Data Mining Algorithm with Naïve Bayes algorithms, FT tree algorithm, and K Star in 2010.

A novel approach for Liver disorder Classification using Data Mining Techniques" with methods Fuzzy, K-means classification in 2015[4].Fuzzy based classification gives better performance and achieved above 94% accuracy for each type of liver disorder analyze the performance of Classifier Over Liver Diseases "[5] It have propose Prediction of Liver Disease (Biliary Cirrhosis) Using Data Mining Technique with method FT Tree algorithm in 2015. The classification accuracy is found to be better using FT Tree algorithm Liver Disease Prediction using Data Mining Technique with method Naïve Bayes SVM in 2015. The SVM classifier is considered

as a best classification algorithm [6]. The paper “An approach to devise an Interactive software solution for smart health prediction using data mining” [7] aims in developing a computerized system to check and maintain your health by knowing the symptoms.

The another research have diagnosis the Chronic Kidney Disease Using Machine Learning Algorithms” with methods Random Forest ,Back Propagation ,Radial Basis Function in 2016.The results indicate that Random Forest algorithm outperformed all other techniques with the help of feature selection[8].This is the most effective model to predict patients with heart disease proposed Classification of Heart Disease Using K-Nearest Neighbour and Genetic Algorithm with methods KNN, Genetic algorithm in 2013. KNN is more accurate than genetic algorithm [9].The Intelligent Heart Disease Prediction System Using Data Mining Techniques with methods Decision Trees, Naïve Bayes and Neural Network in 2008[10].It is user friendly, web based, scalable, reliable and expandable Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm with methods Genetic Algorithm in 2008[11]. Medical Knowledge Acquisition through Data Mining with method KNN in 2008. This paper presents a model of medical knowledge acquisition through data mining [12].

### III. PROPOSED SOLUTION

This system supports an end user and online consultation. Here propose a framework that enables clients to get moment direction on their medical problems through an astute social intelligent health care system online. The framework is fed with different symptoms and the disease or illness associated with those systems. Also the system allows user to share their symptoms and issues. Data Mining is a field of research has already well proven capabilities of identifying hidden patterns, analysis and knowledge applied on different research domains, now gaining more popularity among researchers towards generating novel and deep insights of these large biomedical datasets also.

Uncovering new healthcare related knowledge to support clinical decision making, is another dimension of data mining. Through a deep literature survey, it is found that early disease prediction is the most demanded area of research in health care sector. The basic idea behind the project is to propose a system that allows users to get instant guidance on their health issues. This smart health prediction system is fed with various symptoms and the disease/illness associated with those systems. This system allows user/patients to share their symptoms and issues It then processes patients symptoms to check for various illnesses and based on input it predict the disease or disorder it feels user’s symptoms are associated with and also suggest the doctor to whom he or she can contact and also book an appointment.

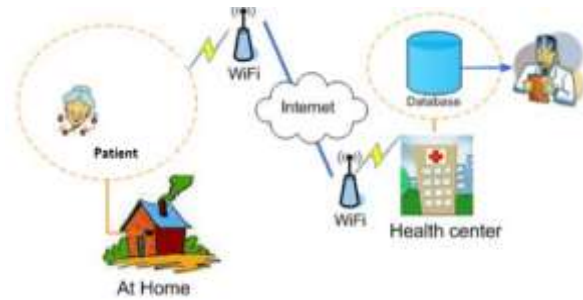


Figure 3.1 System Architecture

#### Objectives:

1. The objective of proposed system is to predict the accurate disease based on users or patients input symptoms and after accurate prediction suggest the prescription or medicines and also suggest nearby doctors with details.
2. Patient can search for doctor’s help at any point of time and can talk about their illness or health issues and get instant diagnosis. doctors may get more patients online.

### IV. DESIGN METHODOLOGY AND ALGORITHM

#### 4.1 Data Mining Technique

Data mining techniques such as association, classification and clustering are used by healthcare organization to increase their capability for building appropriate predictions regarding patient health information from large data. This encompasses a number of technical approaches like clustering, data summarization, classification.

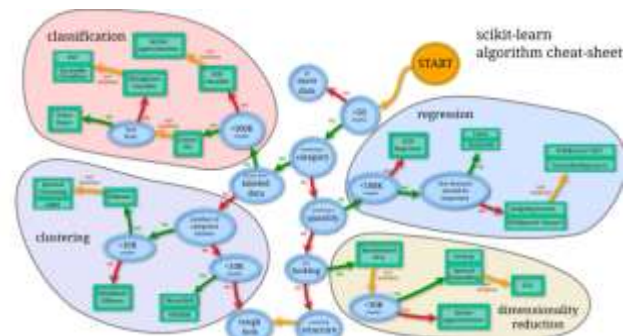


Figure 4.1 Algorithm Analysis

#### Classification:

Classification comprises of two steps: - 1) Training and 2) Testing. Training builds a classification model on the basis of training data collected for generating classification rules. The IF-THEN prediction rule is popular in data mining; they signify facts at a high level of abstraction. The accuracy of classification model based on the degree to which classifying rules are true which is estimated by test data.

#### Prediction:

Prediction in data mining is to identify data points purely on the description of another related data value. It is not necessarily related to future events but the used variables are unknown. Prediction in data mining is to identify data points purely on the description of another related data value.

### The Random Forest Classifier:

Random forest classifier, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the majority of votes becomes our model's prediction. Random forest fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and also control over-fitting.

### Advantages of Random Forest:

Overfitting is one of the critical problem that may make the results worse, but for Random Forest algorithm, if there are enough trees in the forest, the classifier won't overfit the model. The another advantage is the classifier of Random Forest can handle missing values, and the Random Forest classifier can be modelled for categorical data.

There are different ways that Random Forest classifier makes data decisions, and consequently, there are some important related terms to know, Some of these terms include:

1. **Entropy:** It is a measure of randomness or unpredictability in the data set.
2. **Information Gain:** A measure of the decrease in the entropy after the data set is split is the information gain.
3. **Leaf Node:** A leaf node is a node that carries the classification or the decision.
4. **Decision Node:** A node that has two or more branches.
5. **Root Node:** The root node is the topmost decision node, which is where you have all of your data.

### V. CONCLUSION

In this Paper, idea behind the proposed system is to easily predict the disease based on patients symptoms and provide correct prescription online and the patient also get informed about specialist/doctors if they need. Some of the time the circumstance happens when you need the specialist's assistance quickly, yet they are not accessible because of some reason in that case proposed system will be beneficial.

### VI. REFERENCES

- [1] "Smart E-Health Prediction System Using Data Mining" by G.Pooja reddy, M.Trinath basu,K.Bala Sita Ramireddy, Ravi Kumar Tenali, at International Journal of Innovative Technology and Exploring Engineering (IJITEE) , Volume-8 Issue-6, April 2019
- [2] Prof. Krishna Kumar Tripathi, Shubham Jawadwar and Siddhesh Murudkar published paper on "A Health Prediction Using Data Mining" in International Research Journal of

Engineering and Technology (IRJET),Volume: 05 Issue: 04 / Apr-2018

[3] Predicting Disease By Using Data Mining Based on Healthcare Information System" By MIN CHEN, (Senior Member, IEEE), YIXUE HAO, KAI HWANG,ON HEALTHCARE BIG DATA ,2017

[4] "Analyzing the Suitability Of Relevant Classification Techniques On Medical Data Set For Better Prediction" by Ms.M.C.S.Geetha, Dr.I.Elizabeth Shanthi on I-SMAC ,2017

[5] Longhi, Sauro, Freddi, Allesandro," An approach to devise an Interactive software solution for smart health prediction using data mining" , 2017.

[6] Koehler, "Smart Health: What Today's Doctors Aren't Telling You", 2016

[7] "Survey on Data Mining Algorithms in Disease Prediction",by Kirubha V et al. IJCTT, 2016

[8] S. Vijayarani and S. Sudha, "Disease Prediction in Data Mining Technique – A Survey," International Journal of Computer Applications & Information Technology, vol. II, no. I, pp. 17–21,2016

[9]Aditya Tomar, "Disease Prediction System using data mining techniques",in International Journal of Advanced Research in computer and Communication Engineering,ISO 3297, July 2016.

[10] Dr.B.Srinivasan, K.Pavya, "A study on data mining prediction techniques in healthcare sector", in International Research Journal of Engineering and Technology(IRJET), March-2016.

[11]Dr.B.Srinivasan, K.Pavya, "A study on data mining prediction techniques in healthcare sector", in International Research Journal of Engineering and Technology(IRJET), March-2016.

[12]N. A. Sundar, P. P. Latha, and M. R. Chandra, "PERFORMANCE ANALYSIS OF CLASSIFICATION DATA MINING TECHNIQUES OVER HEART DISEASE DATA BASE," International Journal of Engineering Science & Advanced Technology, vol. 2, no. 3, pp. 470– 478, 2015.

[13] S. A. Pattekari and A. Parveen, "PREDICTION SYSTEM FOR HEART DISEASE USING NAIVE BAYES," International journal of Advanced Computer and Mathematical Sciences, vol. 3, no. 3, pp. 290–294, 2015.

[14] M. Jabbar, P. Chandra, and B. Deekshatulu, "CLUSTER BASED ASSOCIATION RULE MINING FOR," Journal of Theoretical & Applied Information Technology, vol. 32, no. 2, pp.196–201, 2012.

[15] P. Chandra, M. . Jabbar, and B. . Deekshatulu, "Prediction of Risk Score for Heart Disease using Associative Classification and Hybrid Feature Subset Selection," in 12th International Conference on Intelligent Systems Design and Applications (ISDA), 2012, pp. 628– 634.