



# **A Survey on Predictive System for Medical Diagnosis with Expertise Analysis**

Pushkaraj R. Bhandari, Sapna P. Yadav, Shyam A. Mote, Devika P. Rankhambe

BE Student, Department of IT, Anantrao Pawar College of Engineering, Pune, India

BE Student, Department of IT, Anantrao Pawar College of Engineering, Pune, India

BE Student, Department of IT, Anantrao Pawar College of Engineering, Pune, India

Head of Department, Department of IT, Anantrao Pawar College of Engineering, Pune, India

**ABSTRACT:** We live in the world of digital data and each day, more and more medical data are collected and stored in databases. This medical data about large patient in different hospitals is analysed to perform medical research. Diagnosis is a very important and sophisticated task that need to be performed accurately and exactly with correct previous information and insights. Large number of studies have shown that the diagnosing of one patient will vary if the patient is examined by totally different physician or perhaps by constant physician at varied times. The amounts of information generated by care transactions are too advanced and are not used for further analysis. To overcome this, we develop predictive medical diagnosis system which helps patient to identify the disease at an early stage and expertise analysis helps them to get detailed prescription and preventive measures in less time. The goal of this work is to develop a web application using data mining technique like Naïve Bayesian algorithm that can be used to predict the occurrence of disease based on user input symptoms.

**KEYWORDS:** Naïve Bayesian algorithm application, Predictive medical diagnosis system, Medical diagnosis using data mining techniques, Naïve Bayesian application in healthcare, Naïve Bayesian algorithm.

## **I. INTRODUCTION**

Predictive systems are the systems which are used to predict some outcome on the basis of some pattern recognition. Diagnosis is the identification of the nature and cause of a certain phenomenon. Diagnosis is used in many different ways in order to find and determine cause and effect of a particular evolution or disorder. Medical diagnosis is the process of determining which disease or condition may involve on the basis of persons signs and symptoms. Diagnosis is challenging and difficult because many signs and symptoms are not specific. For example, fever itself is a symptom of many disorders which doesn't tell the healthcare professional what exactly the disease is. As the results or opinion vary from one physician to another, there is a need to assist a medical physician which will have similar opinion for certain symptoms and disorders. This can be done by analysing the information generated by medical data or medical records. As a result, the new information can be compared with previous records and optimistic diagnosis can be done.

Predictive medical diagnosis is a web application which will predict a particular disorder on the basis of symptoms and provide diagnosis for same disorder which will be detected by algorithm. Healthcare professionals use their previous knowledge and insights to arrive at a particular decision regarding any disease or disorder. In the similar manner, we train our web application through naïve Bayesian algorithm to arrive at a particular decision for diagnosis by using generic disease datasets.

## **II. RELATED WORK**

In [1] authors proposed how data mining techniques like Neural networks, Decision trees and Naïve Bayes can be used to find out hidden patterns for making decision in healthcare organizations. In [2] author showed how data mining methods like Decision trees, CART and ID3 can be used in predicting Heart disease with high accuracy rate and efficiency. These algorithms used historical medical datasets for analysis and performance analysis of different algorithms. In [3] Association rule mining technique was used for diagnosis of diabetes. The authors concluded that the

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

data mining techniques when used appropriately increases the computation and also the classification performance. These rules have the potential to improve the expert system and to make better clinical decision making. In [4] author viewed the comparison between Bayes theorem, Aitken's formula and Logistic model for decision making. The algorithms described could predict different dermatological diseases and Bayes theorem prediction accuracy was greater than other algorithms. In [6] author reviewed that the computer aided heart disease prediction system helps the physician as a tool for heart disease diagnosis. From analysis it is concluded that neural network with offline training is good for heart disease prediction at an early stage and the good performance of the system can be obtained by pre-processed and normalised datasets and the classification accuracy can be improved by reduction in features. In [9] author presented a system which was able to detect the hepatitis virus through symptoms as input at early stage and to support physicians in medical decision making.

### III. PROPOSED WORK

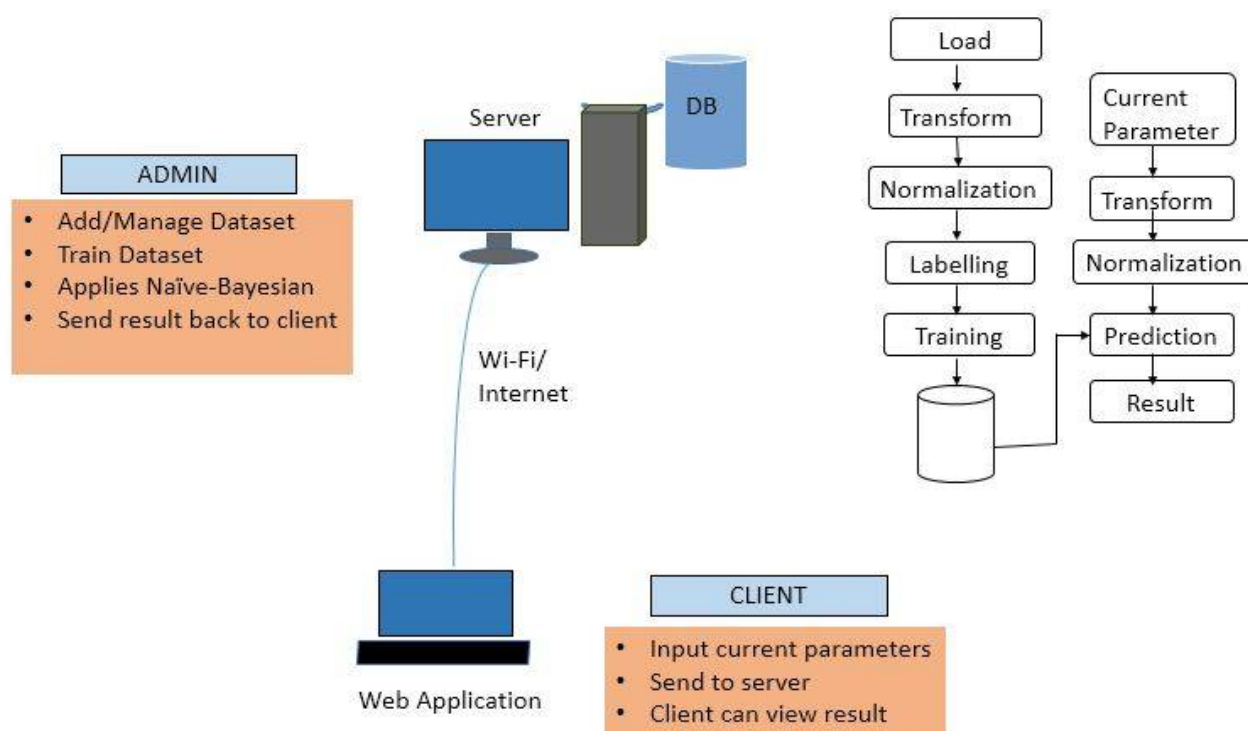


Fig: Block diagram of proposed application

#### A. Design Considerations:

- Client server application.
- Client sends input as symptoms and view results after computation on server.
- Server performs computation of algorithm and results.
- Dataset of diseases is used to train algorithm.
- Only mentioned diseases in dataset can be shown as output after computation.
- Diseases other than one that are mentioned in dataset are not shown as output.

#### B. Description of the Algorithm:

The algorithm used is **Naïve Bayesian classification** algorithm for prediction. The rule (algorithm) is used to create models with predictive capabilities. Aim of algorithm used is to classify individual's signs and symptoms and predict the possible diseases/disorders through computation with the help of dataset.



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Step 1: Calculating Prior probability:

Prior probabilities are the original probabilities that is number of observation divide by total number of observation of an input data, which will be calculated at initial stage of algorithm.

Step 2: Calculating Likelihood probability:

Likelihood probability is the probability of outcomes of every attribute with every other attribute in input data. Also called as Frequency table.

Step 3: Calculating Posterior probability:

Posterior probability is the multiplication of prior and Likelihood probabilities. The output of the posterior probability is used for the result.

Formula for Naïve Bayesian algorithm:

$$P(C|X) = P(X|C) * P(C)$$

**P(X)**

Were,

Posterior probability= P (C|X)

Prior probability= P (C)

Likelihood probability= P (X|C)

Posterior= Likelihood \* Prior

C. Pseudo code of Algorithm:

Step 1: Check for dataset in integer format (0/1).

Step 2: Create Prior probability table.

Step 3: Create Likelihood table.

Step 4: Calculate Posterior probability.

Step 4: Check input from client and calculate Posterior probability. Compare P[x], P[y].

Step 5: if (P[x] > P[y])

Then P[x] =Predicted output.

Else

P[y]=Predicted output.

Step 6: Repeat step 4.

Step 7: End.

D. Description of Application Working:

CLIENT SIDE WORKING:

1] Client will fill basic details like name, gender, age in form through Web application GUI.

2] Selection of symptoms based on individual choice and submit.

3] View output of probable disease on application.

SERVER SIDE WORKING:

1] Create dataset of diseases and their symptoms.

2] Convert the dataset in the integer format by renaming strings to integer (0/1).

2] Write Naïve-Bayes algorithm and differentiate between train data and test data.

3] Train algorithm. Before training algorithm, first load dataset, then apply normalization and manual labelling and then train data.

4] When input is send from client side, the server performs classification using algorithm on input data and sends prediction result back to client through application GUI.

5] Server admin can add, update, and delete records in database.

## IV. CONCLUSION AND FUTURE WORK

Diseases and disorders need to be predicted and diagnosed at an early stage before it get complicated. Several side effects can occur of a disease if not diagnosed at right time. The system helps to predict an occurrence of disorder at an



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

early stage by taking symptoms as input which will help the patient to recover it on time and gives details about the disease/disorder which is predicted by application. Also, the application can work as a supporting hand in decision making for physicians and medical students. Diagnostic tools when well designed and implemented, holds great potential to improve, healthcare quality, increase efficiency and reduce healthcare costs. The diagnostic system should not be viewed as a technology or substitute for clinician but should be considered as a system which gives right outcomes and which can be used by clinicians to get better insights.

The goal of the web application is to assist an electronic medical practitioner in decision making and diagnosis which is available for all at any time of day. The algorithm used in this web application is naïve Bayesian which performs well in multi-class classification problems and helps to improve accuracy in prediction by training algorithm. The application can be extended in future by creating Android application.

## REFERENCES

1. Monika Gandhi, Dr. Shailendra Narayan Singh, "Prediction in Heart Disease using techniques of Data mining", IEEE Transaction on Futuristic trend in Computational Analysis and Knowledge Management, 2015.
2. M.A. Jabbar, B.L. Deekshantalu, Priti Chandra, "Computational Intelligence Technique for early diagnosis of Heart Disease", ", IEEETransaction on Engineering and Technology, 20<sup>th</sup> March, 2015.
3. Nishara Banu, Gomathy, B, "Disease Forecasting System Using Data Mining Methods", IEEE Transaction on Intelligent Computing Applications, 2014.
4. K Raj Mohan, Ilango Paramasivam, Subhashini SathyaNarayan, "Prediction and Diagnosis of Cardio Vascular Disease- A Critical Survey", 2014 World Congress on Computing and Communication Technologies, 2014.
5. Sankaranarayanan S., Dr Pramananda Perumal.T, "Diabetic prognosis through Data Mining Methods and Techniques", IEEE Transaction on Intelligent Computing Applications, 2014.
6. Manjusha K. K, K. Sankaranarayanan, Seena P, "Prediction of Different Dermatological Conditions Using Naïve Bayesian Classification", IJARCS, Volume 4, Issue 1, January 2014.
7. Hlaudi Daniel, Masethe Mosima, Anna Masethe, "Prediction of Heart Disease using Classification Algorithms", Proceedings of the World Congress on Engineering and Computer Science 2014 Vol II , 22-24 October, 2014.
8. R. Chitra, V. Seenivasagam reviewed "Heart Disease Prediction System using Data mining and Hybrid Intelligent Techniques", ictact journal on soft computing, Volume: 03, Issue: 04, July 2013.
9. Abhishek Taneja proposed "Heart Disease Prediction System Using Data Mining Techniques", ORIENTAL JOURNAL OF COMPUTER SCIENCE & TECHNOLOGY, Vol. 6, Issue No. (4): December 2013.
10. Ahmad Ashari, Iman Paryudi, A Min Tjoa, "Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool", IJACSA, Vol. 4, Issue No. 11, 2013.
11. Aqueel Ahmed, Shaikh Abdul Hannan, "Data Mining Techniques to Find Out Heart Diseases: An Overview", International Innovative Technology and Exploring Engineering (IJITEE), September 2012.
12. Shweta Kharya proposed "Using Data mining Techniques for Diagnosis and Prognosis of Cancer Disease", International Journal of Computer Science, Engineering and Information Technology (IJCEIT), Vol.2, Issue No.2, April 2012.
13. K.Srinivas, Dr.G.Raghavendra Rao, Dr. A.Govardhan, "Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques", IEEE Transaction in International Conference on Computer Science and Education , 2010.
14. Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IEEE INSPEC Number 978-1-4244-1968-5, 2008.
15. Doina Drăgulescu, Adriana Albu, "Medical Predictions System", Acta Polytechnica Hungarica, Vol. 4, Issue No. 3, 2007.
16. Darcy A. Davis, Nitesh V. Chawla, Nicholas Blumm, "Predicting Individual Disease Risk Based on Medical History", 2007.
17. Jin Huang Jingjing Lu, Charles X. Ling, "Comparing Naive Bayes, Decision Trees, and SVM with AUC and Accuracy", IEEE Transaction in International Conference on Data Mining, 2003.
18. K.Srinivas, B.Kavihta Rani, Dr. A.Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", (IJCE) International Journal on Computer Science and Engineering, Vol. 02, Issue No. 02, 2010.