

# Disease Prediction System

Utkarsh Srivastava<sup>1</sup>, Abhishek Vikram<sup>2</sup>, Shubham Mishra<sup>3</sup>, Ashutosh Gupta<sup>4</sup>

<sup>1,2,3,4</sup> *Department of Computer Science and Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad*

**Abstract** - Successful adoption of data mining in prominent areas such as market analysis, e-commerce, production control, and scientific discovery has led to applications in other industries. The medical environment continues to be information-rich but knowledge weak. The medical system has a lot of possible knowledge. However, there is still no powerful analytical tool for revealing hidden relationships and trends in data. The disease could be a term that assigns to some health care conditions associated with the body. This medical condition represents an undesirable state of health that directly controls all parts of the body. Medical data processing techniques like association rule mining, classification, clustering is implemented to research the various varieties of general body-based problems. Classification is a very important problem in data processing. Several popular classifiers construct decision trees to get class models. The information classification relies on the ID3 Decision Tree algorithm which ends from inaccuracy, the info is estimated using entropy-based cross-validations and partition techniques, and the results are compared.

**Index Terms** - Data mining, Data processing, Disease prediction, General body diseases, Prediction system.

## I. INTRODUCTION

It is estimated that over 70% of individuals in India are at risk of general body diseases like viral, flu, cough, cold. Etc., in every 2 months. Because many folks do not realize that the general body diseases may well be symptoms to something more harmful, 25% of the total population succumbs to death thanks to ignoring the first general body symptoms. This could be a dangerous situation for the population and might be alarming. Therefore, it is extremely important to recognize or predict the disease as early as possible to avoid unwanted victims. The currently available systems are the systems that are either dedicated to a selected disease or are in research phase for algorithms when it involves generalized disease. The purpose of this method is to supply prediction for the final and

more commonly occurring disease that when unchecked can grow to be fatal disease. The system applies data processing techniques and ID3 decision tree algorithms. This method will predict the foremost possible disease supported the given symptoms and precautionary measures required to avoid the aggression of disease, it will also help the doctors analyze the pattern of presence of diseases within the society. In this project, the disease prediction system will perform data processing in its preliminary stages, the system is trained using machine learning and data.

## II. LITERATURE REVIEW

In the document “Disease Prediction System using data mining techniques” [1] the author has discussed about the data mining techniques like association rule mining, classification, clustering to analyze the different kinds of heart-based problems. The database used contains a collection of records, each with a unique class tag, a classifier provides a concise and clear definition of each class that can be used to classify consecutive records. The data classification is based on the MAFIA algorithm to provide computational accuracy and information is calculative.

The exploitation entropy primarily based on cross validations and partition techniques and the results are compared. Algorithmic rules are used because they indicate the rank of attacks with the selection tree. Heart Disorders Information is a major abuse of the rules governing the K means algorithm to remove heart attack data from the database. Some problems faced by the system like, time complexity is more due to DFS traversal, Time complexity increases as unnecessary branches are searched and, in the end, no precautions are taken.

In the article “A study on predictive data mining technology in the healthcare sector” [2] The field mentioned are, information discovery method (KDD)

is that the method of placing the low-level data into high-level knowledge. Therefore, KDD refers to the unimportant deletion of implicit, previously unknown, and certainly useful data from the information contained in the database.

The repetitious method consists of the subsequent steps: information cleansing, information integration, information choice, information transformation, data processing, pattern recognition. Healthcare Data Mining Prediction supports the following data processing technologies: Neural network, Bayesian Classifiers, Call tree, Support vector machine. The paper states the comparative study of various aid predictions, Study of data mining techniques and tools for predicting cardiovascular disease, many cancers, diabetes, disease, and medicine. Few limitations are that if the attributes of were not relevant, the decision tree predictions would be less accurate, and ANN is computationally intensive to train also it does not lead to precise conclusion.

The paper “Predicting Disease by Using Data Mining Based on Healthcare Information System” [3] applies the information mining process to predict high blood pressure from patient medical records with eight alternative diseases. The data is pulled from real world health system information, including medical records. Under- sampling technique has been applied to come up with coaching knowledge sets, and data processing tool that has been used to generate the Naïve Bayesian and J48 classifiers created to improve the Predictive performance, and fuzzy set tools failed to scale back the ensemble supported the concept of second- order approximation.

Results originated from the experiment showed small improvement of the ensemble approach over pure Naïve Bayesian and J-48 in accuracy, sensitivity, and F-measure. Initially they had a classification and so ensemble the classifiers and so the reduction of Ensemble Classifiers is employed. But the choice trees generated by J-48 often lack within the leveling therefore the overall improvement of victimization ensemble approach is a small amount.

The paper “An approach to devise an Interactive software solution for smart health prediction using data mining” [4] aims in developing a computerized system to check and maintain your health by knowing the symptoms. It has a symptom checker module which defines our body structure and gives us liability to select the affected area and checkout the symptoms.

The technologies implemented in this research paper are: The interface is designed using HTML, JavaScript, and CSS. The back end is designed using MySQL which is used to design the databases.

This document also contains information about tests such as alpha tests performed on the server. Also, we can say at the developer’s end, this is an actual testing done with potential users or as an independent testing process at server end and beta testing is done after executing alpha testing, versions of a system or so-called beta software is available to specific viewers outside the programming team. Only the limitation of this paper is it suggests only the award-winning doctors and not the nearby doctors to the patient.

### III. INFRENCES DRAWN FROM LITERATURE REVIEW

The general disease prediction system predicts the chance of the presence of a disease present in a patient based on their symptoms. It also recommends necessary precautions to treat the predicted disease. The initial system was fed up with data from various sources. i.e., patients, the data will then be pre-processed before the further process is carried out, this is done to get clean data from the raw initial data, as the raw data would be noisy, or flawed. This data will be processed using Data mining algorithms, the system will be trained to predict the disease based on the input data given by the user.

The system is implemented into two parts, the admin part, and the user part. The admin must train the system for the creation of the disease prediction model. The user uses the services provided by the model after logging in as the user, on entering the symptoms into the model, the model returns the necessary predictions and precautions.

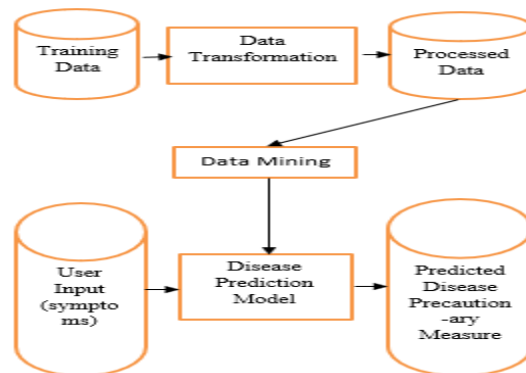


Fig 1: Block diagram for disease prediction system

#### IV. PROPOSED WORK

The General body disease prediction system applies data mining techniques using ID3 algorithm. Decision trees are considered easily understood models because a reasoning the process can be given for each conclusion. The knowledge model that follows this model can be directly transformed into a set of IF-THEN rule, that are one of the most common popular forms of knowledge representation.[2]

##### 1. Admin

The results for the admin of the DPS are as follows.

##### (a). Login

Admin can log in to the system selecting the user type and entering the required details.



Fig 2: Login page for admin

##### (b). System Training

The admin should train the system by uploading the data set into the system.

Experiments were carried out to evaluate the performance and usefulness of different classification algorithms for predicting disease present inpatients.

The performance of the learning techniques is highly dependent on the nature of the training data. Confusion matrices are extremely useful for evaluating classifiers. The columns represent the predictions, and the rows represent the actual class.[3]

##### 2. User

The results for the user in DPS are as follows

##### (a). User login

Pre-registered users should log in to the system to have access to the services.

##### (b). Enter Symptoms

The user will have to select the symptoms here.



Fig 3: Symptom selection form for user

##### (c). Prediction and precaution

The results are calculated by the model based on the rule set will be shown here.



Fig 4: Prediction result for disease

#### V. CONCLUSION

The system has been implemented with an accuracy of 86.67% on the dataset of 120 patient data. The current system covers only the most common diseases or the more commonly occurring disease, the plan is to include disease of higher fatality, like various cancers in the future, so that early prediction and treatment could be done, and the mortality rate of deadly diseases like cancer decreases and chronic diseases could be reduced, with the economic benefit in long sight.

#### REFERENCES

- [1] Aditya Tomar, "Disease Prediction System using data mining techniques", in International Journal of Advanced Research in computer and

Communication Engineering, ISO 3297, July 2016.

- [2] Dr. Srinivasan, K.Pavya, “A study on data mining prediction techniques in healthcare sector”, in International Research Journal of Engineering and Technology (IRJET), March2016.
- [3] Feixiang Huang, Shengyong Wang, and Chien-Chung Chan, “Predicting Disease by Using Data Mining Based on Healthcare Information System”, in IEEE 2012
- [4] M.A. Nishara Banu, B Gomathy, “An approach to devise an Interactive software solution for smart health prediction using data mining, in International Journal of Technical Research and Applications, eISSN, Nov-Dec 2013.