



# A REVIEW ON DISEASE PREDICTION SYSTEM

Anurag Singh , Ashutosh Singh, Arohi Rastogi, Ayushi, Ms. Shivani Agarwal  
Department of CSE  
IMS Engineering College, Ghaziabad, Uttar Pradesh, India

**Abstract**— Medicinal services though providing everything a person needs in sickness still lack knowledge. There is an abundance of realities accessible inside the social insurance structures. Be that as it may, there is a nonappearance of incredible evaluation devices to find concealed connections in insights. This framework plans to make GUI based Interface to enter the patients symptoms which infirmity the influenced individual is having to utilize different machine learning algorithms.

The forecast is practiced from mining the influenced individual's indication realities or records vault. This paper has investigated forecast structures for the illness the utilization of a greater amount of info properties. The framework utilizes clinical expressions comprising of fever, torment, cholesterol-like credits to expect the likelihood of an influenced character getting a selected sickness. As of not long ago, over a hundred traits are utilized for the forecast. The data mining methods, specifically Decision Trees, Naive Bayes, and Random Forest are broke down on the sickness database. In light of the exactness, the exhibition of these strategies is looked at.

**Keywords**— Predictive analysis, Data mining, Machine Learning

## I. INTRODUCTION

Data mining is a technique of extracting useful data from huge amount of raw data. The data is increasing days by day in very huge amount. The processing and investigation of these large databases are almost impossible for human beings so he man need computer-based technologies for analysis of data. The computers are primarily used to do mechanized medical diagnosis. This mechanized diagnosis tool supports the medical team to make a good choice in disease prevention. Data mining is the platform used to process the huge quantity of patient datasets in many ways which include making a feel of complex diagnostic tests, interpreting previous results, and combining the dissimilar data.

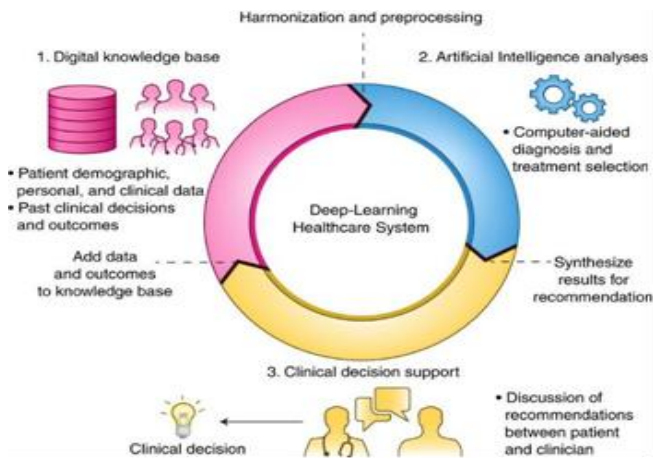
We live in an era where predicting the disease and providing suitable medications through some user friendly application is more appropriate and efficient. This study aims mainly for the health concern and help the patients who want to be their own doctor. It is an interactive, user friendly application which helps the patients to know the disease as per their symptoms provided by them.

## II. EXISTING SYSTEM

The existing disease prediction systems are mostly based on data mining techniques and the have an array of symptoms that the end user selects that he/she are feeling and the system will eventually predict the disease on the basis of the symptoms selected the issue with these systems are that there are no real time analysis and moreover the person might have symptoms that he/she may not know thus making the results false.

## III. PROPOSED SYSTEM

The proposed system will not only identify the illness but also provide the necessary medications required in the illness. The system will be helpful in things such as identifying illness, identifying symptoms, suggesting medications. This system will predict the common-cold, cough, viral, as well as supplementary infection like acidity and weakness. It is more efficient than the previous systems. The user interface is made as user friendly as possible. It will use three algorithms namely Random Forest, Naïve Bayes, and Decision Tree. Many steps are followed in the system which include iterative techniques such as: Data cleaning, Data integration, Data selection, Data transformation, Data mining, Pattern assessment. There is heavy use of statistics and mathematical model to make the system as efficient as possible. The end result is the one with the maximum matches. We have used a reliable platform for sharing information with the physicians to aid the patients as well as doctors. The doctors on the basis of this data would be able to provide adequate consultancy to the patients.



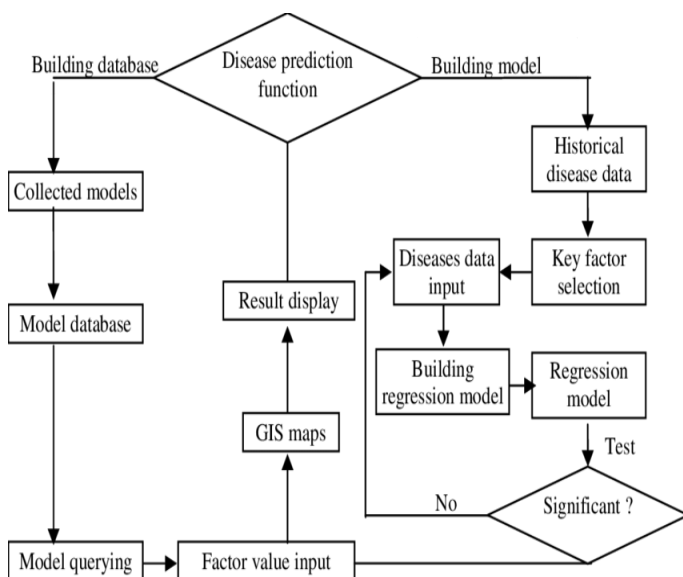
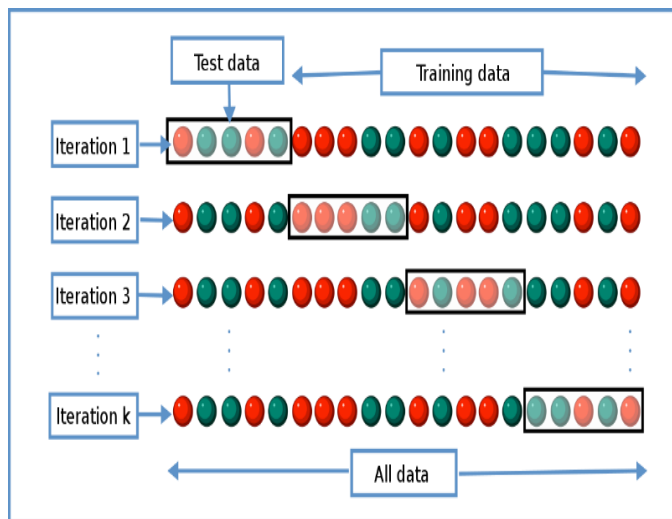
#### IV. LITERATURE SURVEY

In the paper "Disease Prediction System the usage of information mining techniques" techniques like association rule mining, class, clustering are used to analyze the different kinds of diseases. The database used to contain a collection of records, each with a single class label, classifier performs a brief and clear definition for each class that can be used to classify successive records. The data classification is based on MAFIA algorithms which result in inaccuracy, the data estimated the use of entropy-primarily based completely cross-validations and partition techniques, and the consequences are compared. The database is cluster reducing the K-means clustering set of rules, which will remove the data applicable. Some Limitations Are Faced By the system like the time complexity is more due to DFS traversal, C4. Five- Time complexity will increase while attempting to find insignificant branches and lastly no precautions redefined.

In the paper "A study on data mining prediction techniques in the health care sector" the fields discussed are, Knowledge Discovery Process(KDD) is the process of changing low-level data into high-level knowledge. Hence, KDD refers to the nontrivial removal of implicit, previously unknown, and potentially useful information from data in databases. The Knowledge Discovery In Databases process comprises of a few steps leading from raw data collections to some form of new records. The iterative technique includes the subsequent steps: Data cleaning, Data integration, Data selection, Data transformation, Data mining, Pattern assessment, Knowledge.

Healthcare statistics mining prediction based on facts mining techniques are as follows: Neural network, Bayesian Classifiers, Decision tree, Support. Some of the previous related works in basic disease prediction check of diverse healthcare predictions, the Study of statistics mining techniques and equipment for the prediction of coronary heart sickness, several cancers, diabetes, eye disease, and dermatological conditions. The data mining based prediction system reduces the human effects and cost-powerful one. Few limitations are that if attributes aren't associated then Decision bushes prediction is less accurate and ANN is computationally intensive to train so it does not lead to a specific conclusion.

The paper "Predicting Disease By Using Data Mining Based On Healthcare Information System"[4]applies the data mining process to stop redirect hypertension from patient medical records with 8 other diseases. The records became extracted from a real-international Healthcare Information System database containing medical facts. The under-sampling approach has been applied to generate schooling facts sets and records mining tool Weka has been used to generate the Naive Bayesian and J-forty eight classifiers created to improve the prediction performance, and the hard set system had been used to reduce the ensemble-based totally at the concept of second-order approximation. Experimental effects showed a little





improvement of the ensemble method over herbal Naive Bayesian and J-48 in accuracy, sensitivity, and F-measure. Initially, they had a category and then ensemble the classifiers and then the reduction of Ensemble Classifiers Used. But the decision tree generated by J-48 is from time to time lacking in the balancing so the overall development of using the ensemble method is much less. And for the prediction of coronary heart ailment level, no precautions are defined.

The paper "A method to plot an Interactive software solution for clever fitness prediction using facts mining" [5] interests in developing a computerized device to check and keep your fitness by knowing the symptoms. It has a symptom checker module which without a doubt defines our body shape and gives us a legal duty to pick the affected location and test out the symptoms. Technologies applied in this paper are: The front stop is designed with the help of HTML, JavaScript, and CSS. The back quit is designed using MySQL that's used to layout the databases. This paper additionally consists of the records of finding out like Alpha trying out that's done at the server side or we can say on the developer's quit, this is trying out carried out with potential clients or as an independent sorting out technique at the server cease. And Beta attempting out is achieved after performing alpha attempting out, variations of a gadget or software program known as beta variations are given to a specific audience out of doors the programming team. Only the quandary of this paper is it shows handiest the award-winning medical doctors and no longer the close-by medical doctors to the patient.

Author	Year	Data Mining Tool	Techniques used	Accuracy
Abhishek et al.	2013	WEKA 3.6.4	J48	95.56%
			Naïve Bayes	92.42%
			Neural Network	94.85%
Chaitrail et al.	2012	WEKA 3.6.6	Neural Network	100%
Nidhi et al.	2012	WEKA 3.6.6 TANAGRA .NET	Naïve Bayes	99.52%
			Decision Tree	52.33%
			Neural Network	96.5%
Vikas Chaurasia et al.	2013	WEKA	CART	83.49%
			ID3	72.93%
			Decision Table	82.50%
Hlaudi Daniel Masethe et al.	2014	WEKA	J48	99.074%
			REPTREE	99.74%
			Naïve Bayes	97.22%
			Bayes Net	98.14%
			Simple CART	99.74%
Rashedur et al.	2013	WEKA	Neural Network	79.19%
		TANAGRA	Fuzzy logic	83.85%

## V. CONCLUSION

Thus in the following paper, we have successfully analyzed the current architecture and through the help of this architecture we have developed a system that helps in prediction of diseases.. The system is capable in predicting various diseases and suggesting medications accordingly by using multiple machine learning techniques like Classification, Association rule mining. It has plenty of scopes and solves various issues like identifying diseases at an earlier stage and coping with them at an early stage

before they become larger issues. The system can further be expanded in order to identify those diseases. We have also added a chat bot through which the end user i.e patients can take consultancy from doctors and therefore wouldn't require to take appointments or to physically visit a clinic. Having quick data about patients vitals can also ease up doctors work and thus he would be able to attend more patients thus it helps the doctors along with the patients. The advantage of this system is that it incorporates machine learning and thus it is capable of handling huge amounts of data and providing better insights to physicians so that they can do their job better and finally resulting in increased patient satisfaction and in lowering medical charges. In long term the system will eventually cover more diseases and drugs, as we get more and more integrated data. We are ready to use more data in order to increase the use case of the system such that it provides real time analysis both to patients and the doctors.

## VI. REFERENCE

- [1] Tomar Aditya (July 2016) "Disease Prediction System the usage of facts mining strategies", in International Journal of Advanced Research in Computer and Communication Engineering, ISO 3297,
- [2] Dr. Srinivasan B., Pavya K.(March-2016)"A check on statistics mining prediction strategies within the healthcare sector", in International Research Journal of Engineering and Technology (IRJET).
- [3] Rathi Megha, Pareek Vikas(Nov-Dec 2016.) "An integrated hybrid facts mining approach for healthcare", in IRACST - International Journal of Computer Science and Information Technology Security (IJCSITS), ISSN: 2249-9555, Vol.6, No.6.
- [4] Huang Feixiang, Shengyong Wang, and Chien-Chung Chan(Nov-Dec 2016) "Predicting Sickness By Using Data Mining Based on Healthcare Information System", in IEEE 2012.
- [5] Raul Akshay , Patil Atharva , Raheja Prem, Sawant Rupali, "Knowledge Discovery, Analysis And Prediction in tending victimization data processing And Analytics", in IEEE,2nd International Conference on Next Generation Computing Technologies.
- [6] Prince Theresa, R Research Scholar, J.Thomas, "Human Heart Disease Prediction System using Data Mining Techniques", in 2016 International Conference on Circuit, Power and Computing Technologies [ICCPCT], @2016 IEEE. [10] RanganathaS.1, Pooja Raj H.R.2, Anusha C.3, Vinay S.K, "Medical Data Mining and Analysis for Heart Disease Dataset Using Classification Techniques", in IEEE 2009.
- [7] Palaniappan Sellappan, Awang Rafiah, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", in IEEE 2008.
- [8] Banu Nishara M.A.,Gomathy B(2014) "Disease



Forecasting System Using Data Mining Methods”, in  
IEEE, 2014 International Conference on Intelligent  
Computing Applications.