International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Prediction of Diseases in Smart Health Care System using Machine Learning

N. Shabaz Ali, G. Divya

Abstract: ***The main aim of this paper is to discuss about the use of Data mining in the field of Medical health care. These data mining techniques can be used in various fields of Research and Education also. The Smart Health Prediction System Is the Fastest emerging area in the field of medical science. Data mining is one of the fields of computer science that uses the existing data in medical field to predict the occurrence of diseases. By the use of machine learning and database management tools we can extract new patterns from group of large datasets and gain knowledge. In the following paper the survey is made on how the data mining techniques are used along with the machine learning to predict the diseases based on the user symptoms.***

Keywords: ***Data mining, prediction analysis, healthcare, symptoms, machine learning***

I. INTRODUCTION

Data mining is a process of knowledge discovery from unknown or useless datasets. There are various techniques of data mining that are used to process the data and convert them as useful information. The data mining can be used in the various fields such as business analysis, healthcare, stock management etc. Medical field has wide amount of data that can be processed by the help of data mining techniques.

It might have happened before that yourself or someone near you want immediate help of doctor but could not find anyone. By creating a model that can predict the diseases based on user symptoms is quite helpful in getting fast and appropriate medical facilities for patients. The timely analysis of data and gaining accurate prediction of diseases from symptoms can save many lives. Early detection of diseases helps doctor to give accurate medication. In the field of medicine different algorithms of machine learning are used for predicting different diseases and helps the physicians to diagnose fast. Based on the input of data the accuracy of results may vary.

Manuscript published on January 30, 2020.

* Correspondence Author

N. shabaz Ali*, UG Scholar, Department of Computer Science and Engineering, Sabetha school of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India. Email: nayabshabazali@gmail.com

G. Divya, Assistant Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India. E-mail: mailtodivya16@gmail.com

II. LITERATURE SURVEY

In the paper "Smart health prediction system using data mining"[1] the author has discussed many topics related to data mining techniques such as Naive Bayes, KDD(Knowledge discovery in Database). The Bayesian statistics can be applied to economic sociology and other fields. This checks the patients at initial level and automatically suggest the possible diseases. The system uses Naive Bayes classifier for the construction of the prediction system. The advantage of this system is that the initial consultation cost of doctor fees can be avoided. Eclipse IDE is used for creating the front end Graphical User Interface and Navicat Mysql is used for backend database purpose. Here java is used as a programming language to connect the database and GUI purpose. The only disadvantage of the system the efficiency in detecting the symptoms or symptom mapping.

The paper "A Smart Health Prediction Using Data Mining" [2] is explaining the similar topics to the paper [1]. But there is detailed explanation of the internal algorithms used in the system. The Naive Bayes algorithm can be used for developing models that are used to assign class labels of different format. Naive Bayes algorithm is not a single, but a group of algorithm based on common principle. The steps involved in the Naive Bayes algorithm include (i) Division of segments, (ii) Comparing the first character of pattern until match occurs, (iii) Comparing the last character of pattern, (iv) Perform each character comparison. Also the hardware requirements used are processor of 2.0 GHZ and Ram of 2GB. The software requirements are JAVA programming language, Mysql 5.0 database and Tomcat server.

In the paper "Smart E-Health Prediction System Using Data Mining" [3] most of the topics covered are on the system architecture. In this paper the design aspects of the system are primarily focused. In this paper the author has given a detailed framework to beat the downside of existing system. The smart health framework is used to implement the design aspects of the project. This framework asks for uses input and gather the symptoms to predict the disease based on data mining techniques. There are various modules such as login- used for authentication of patient and doctor, Diseases prediction, Doctor Searching, Feedback and Chatting with doctor for clearing patient doubts. There are some advantages such as finding the nearest doctor option to find doctor near to our location.

These features can be used for better implementation of the system to help patients.

The paper "Analysis of Heart Disease Prediction Using Data mining Techniques" [4] various data mining techniques of heart disease prediction are discussed. The proposed of this paper gives more accuracy than the present machine learning algorithms. Generally, Naive Bayes classifier is used for the prediction of heart diseases. The main advantage of Bayes classifier is the short training models is used to predict large datasets.

But the author has divided the data into two class namely 0- Absent and 1- Present. Later the probability of each attribute of different classes are compared and maximum probability is calculated. By this method the paper shows that 97% accuracy is achieved in predicting the heart diseases. This paper fails to explain the in-depth analysis of the prediction process.

In the paper "Heart Diseases Detection Using Naive Bayes Algorithm" [5] some of the machine learning algorithms such as Naive Bayes classifier. This paper is used for analyzing the various data mining techniques that can be used for healthcare services. The author has discussed about the different types of datasets that can be used in various fields of medical and healthcare services. The methodologies for preprocessing of data and probabilities used in the algorithm are explained clearly. The parameters of heart disease are specified and visualization of datasets are shown. The disadvantage is that maximum accuracy is not achieved in prediction.

The paper "Data Mining Technique and Applications"[6] discusses about the various data mining techniques that can be utilized and applied in various field of medical and technical sciences. The logical process are used to search large amount of data in order to extract structured data. The steps involved are exploration, pattern identification and deployment. In the exploration part the data is analyzed and transformed to various forms until we get the prescribed pattern. Later this patterns are deployed by applying data mining techniques. There are various algorithms and techniques such as Classification, Clustering, Regression, Artificial intelligence, neural networks, Association rules and Decision trees. The advantage is various data mining techniques are clearly explained. The real time examples are not mentioned in detail.

In the paper "Smart Health Prediction using Machine Learning"[7] the techniques like association rule mining, clustering, and classification algorithm such as decision tree are used for different heart based problems. The K-means clustering techniques can be used to improve the accuracy of diseases prediction. The primary task for implementing the project is selecting the domain. The target data should be chosen carefully and preprocessing of data should be done. The desired knowledge should be obtained and final evaluation need to be performed. The detailed explanation of how to implement an algorithm is given, where the data is spitted as training set and target set. The formulation of Naive Bayes algorithm and its working is explained clearly.

The paper "GDPS - General Disease Prediction System" [8] discusses about how data mining techniques can be used for prediction of different kinds of diseases based on the symptoms selected by users. Here the system is implemented as two different parts, admin module and user module. The admin takes care of data preprocessing, training the system for creating disease prediction model. A special algorithm called ID3 algorithm is used for training the datasets. ID3 stands for Iterator Dichotomiser 3. The algorithm can be used to generate the decision tree from the given datasets. The ID3 mainly works on entropy of each attribute, information gained and entropy of whole dataset. The attributes having the lower entropy value is selected as root node. The new attributes are discovered with the subsets and decision tree is formed.

In the paper "Heart Disease Prediction using Data Mining with Map reduce Algorithm" [9] the datasets used are obtained from university of California Irvine (UCI) which is a machine learning repository. The structure of RFNN was clearly explained which was used in preparing the datasets. The Recurrent Fuzzy neural network has about 7 hidden layers, 13 input and 1 out layers. But the problem here is that it requires high configuration hardware for smooth functioning. Results are obtained only at hardware configuration having Intel i7 CPU, 16 GB ram and LINUX system with java. The Map reduce algorithm is used along with generic algorithm to increase the efficiency in prediction. There are True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) instances used in prediction process.

The paper "Research of Chronic Kidney Disease based on Data Mining Techniques" [10] the data mining techniques for kidney related disease are discussed. The kidney disease is a major issue in low income countries such as India. 60% of deaths worldwide are because of kidney related issues. The kidney disease may also lead to other chronic diseases such as high blood pressure, diabetes, anemia, weak bones and nerve damage. With the help of data mining in healthcare frauds and abuses can be detected. It helps physicians to identify best treatment for particular disease. It can produce fast analysis report, operational efficiency and reduce operational cost. There are also some of the disadvantages such as data ownership problems, privacy and security related issues for human data administration etc. Various algorithms are used at different stages of analysis and prediction of disease.

III. ANALYSIS OF DATA MINING ALGORITHMS

Data mining is a process of discovering analyzing different data patterns from large raw datasets. The main aim of data mining is to extract the relevant information from comprehensive dataset. The data mining comes with a bundle of packages such as machine learning, statistics and database system. All this factors determine the efficiency in Knowledge Discovery in database process. KDD consist of various process such as data cleaning, data selection, data integration, data transformation, data pattern searching and finally knowledge representation. The data mining techniques that mainly used are Association rule, Clustering, Classification, regression etc.

- The association rule can be used to establish relationship between two variables.
- The clustering is a process of grouping the structures based on similarity between them.
- The classification is assigning items in collection to target datasets.
- The regression tries to estimate the various mode to find the relation between data with least error.

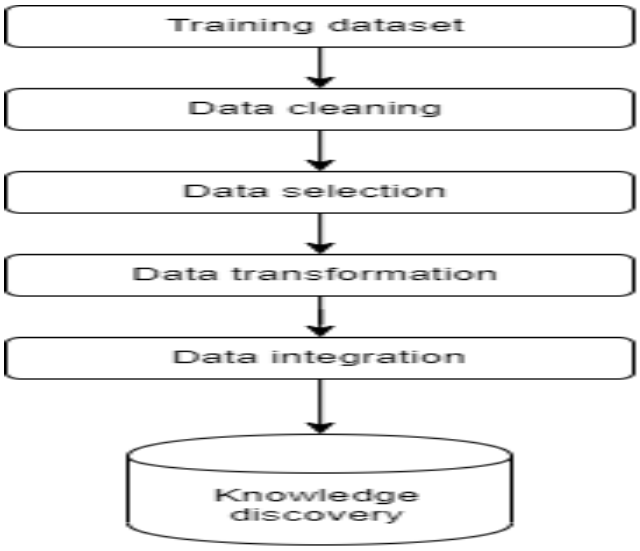

Fig.1 Prediction Process

There are two methodologies of machine learning is used in the data mining process. They are (i) Supervised learning and (ii) Unsupervised learning.

A. **Supervised learning:**

In supervised learning the system trains itself by the given input and learn to generate the result.

B. **Unsupervised learning:**

In unsupervised learning the hidden structure and relation among the dataset is found out.

In healthcare industry, data mining along with machine learning is used for disease prediction. There are various classification models such as Decision trees, Artificial neural networks, Support vector machines and k-nearest neighbors are used.

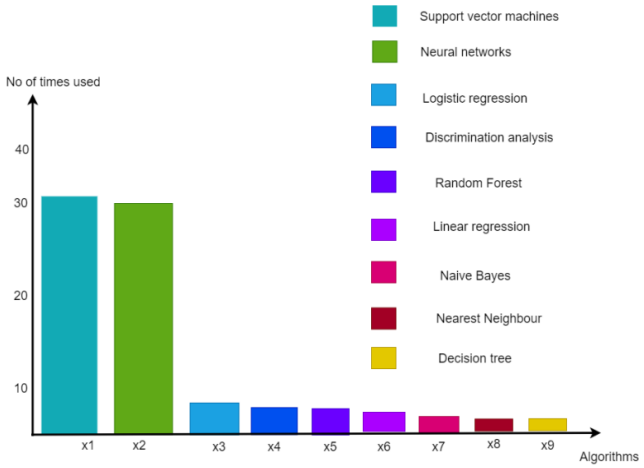

Fig.2 Algorithm analysis in healthcare

As shown in Fig.2, Some of the most used algorithms in healthcare include Support vector machines and Neural networks. Many other algorithms such as Naïve Bayes classifier, linear regression, Logistic Regression are also used.

IV. RESULTS AND DISCUSSION

The results are obtained by the analysis of different algorithms in the healthcare prediction. The major algorithms include Support vector machines, Neural networks, logistic regression, Random forest etc. Among this the accuracy is high in the neural networks if proper training is given by the datasets.

Table 1: Algorithms used in different diseases

Algorithms used	% in use	Diseases	Accuracy
SVM	35	Heart, kidney, liver, diabetes	92%
Neural networks	34	Heart, Kidney	97%
Logistic regression	6	Diabetes	77%
Discrimination analysis	6	Symptoms, Heart, Diabetes	72%
Random forest	5	Heart, kidney	94%
Linear regression	5	liver	72%
Naïve Bayes	4	Heart, cancer	82%
Nearest Neighbor	3	Heart, kidney, liver	92%
Decision tree	2	Diabetes, cancer	94%

In the table 1, the algorithm that are used in healthcare in different fields of medicine along with its accuracy obtained are discussed. This gives us a general idea of the percentage of algorithms which are used in different disease prediction.

V. CONCLUSION

Data mining has greatest importance in the area of medical and technical sciences. Data mining along with the help of machine learning algorithm can create some wonders in the field of medical science. The diagnosis of the disease made easy for doctors and medication can be provided on time. The stages of various diseases can be calculated accurately and according to the patients can be treated. The knowledge gained from the data mining can be helpful to take accurate decisions. In the future by the advancement in the field of IT sector, the data mining will be much more advanced and cam mine different knowledge hidden in medical data.

REFERENCES

1. Nikita Kamble, International Journal of Scientific Research in Computer Science Engineering and Information Technology, Vol. 2, Issue 5, 2017, “Smart Health Prediction System Using Data Mining”.
2. Prof. Krishna Kumar Tripathi, International Research Journal of Engineering and Technology (IRJET) , Vol.5 Issue:4 , Apr-2018, “ A Smart Health Prediction Using Data Mining”.
3. G.Pooja reddy, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol-8 Issue-6, April 2019, “Smart E-Health Prediction System Using Data Mining”.
4. S.SHARMILA, International Journal of Advanced Networking & Applications (IJANA), Vol: 08, Issue: 05, 2017, “Analysis of Heart Disease Prediction Using Data mining Techniques”.
5. K.Vembandasamy, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 9, September 2015, “Heart Diseases Detection Using Naive Bayes Algorithm”.
6. Bharati M. Ramageri , Indian Journal of Computer Science and Engineering, Vol. 1 No. 4 301-305, “Data Mining Technique and Applications”.
7. Vidya Zope1 ,Pooja Ghatge2, Aaron Cherian3, Piyush Mantri4 ,Kartik Jadhav, *IJSRD - International Journal for Scientific Research & Development| Vol. 4, Issue 12, 2017, “*Smart Health Prediction using Machine Learning”.
8. Shratik J. Mishra 1, Albar M. Vasi 2, Vinay S. Menon3, Prof. K. Jayamalini4, International Research Journal of Engineering and Technology (IRJET) ,Volume: 05 ,Issue: 03 | Mar-2018, “GDPS - General Disease Prediction System”.
9. T.Nagamani, S.Logeswari, B.Gomathy, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8 Issue-3, January 2019, “Heart Disease Prediction using Data Mining with Mapreduce Algorithm”.
10. M. Thiyagaraj, G. Suseendran, International Journal of Recent Technology and Engineering (IJRTE), Volume-8, Issue-2S11, September 2019 , “Research of Chronic Kidney Disease based on Data Mining Techniques”.
11. Obenshain, Mary K. "Application of data mining techniques to healthcare data." Infection Control & Hospital Epidemiology, Vol.25, no. 8 2004: 690-695.
12. Sinha, Parul, and Poonam Sinha. "Comparative study of chronic kidney disease prediction using KNN and SVM." International Journal of Engineering Research and Technology, Vol 4, no. 12 pp. 608-12, 2015.
13. Kumar, Manish. "Prediction of chronic kidney disease using random forest machine learning algorithm." International Journal of Computer Science and Mobile Computing, Vol 5, no. 2, pp. 24-33, 2016.

AUTHORS PROFILE

N. shabaz Ali*, Department of Computer Science and Engineering, Saveetha school of engineering , Saveetha Institute of Medical and Technical Sciences, Chennai, India. Email: nayabshabazali@gmail.com.
This publication is based on my final year project in B.E

G. Divya, Assistant Professor, Department of Computer Science and Engineering, Saveetha school of engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India. Email: mailtodivya16@gmail.com. My guide for the project.