International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET2310284

Smart Health Prediction and Drug Recommendation Using Data Mining

M. Robinson Joel[1], G.Manikandan[2], Shiji B [3], Sharmila S[4]
[1,2,3,4]Department of Information Technology, Kings Engineering College, Chennai, India

ARTICLEINFO

Article History:

Accepted: 01 April 2023
Published: 19 April 2023

Publication Issue
Volume 10, Issue 2
March-April-2023

Page Number
506-513

ABSTRACT

To ensure that the citizen and community are alive and in good health in the age of digital technology, the world must establish an excellent health system. Users of the system can communicate their symptoms and problems. After that, it analyses the user's symptoms to look for any potential illnesses. Here, we employ some clever data mining algorithms to identify the ailment that might be most accurately linked to the patient's symptoms. In the event that the system is unable to deliver accurate findings, it tells the user of the disease or ailment it believes the user's symptoms are related to. If a user's symptoms do not precisely match any condition in our database, it displays diseases that the user may likely have based on those symptoms. Together with feedback, contacts, and an administrator dashboard for system management, it also includes the doctor's address. According to this study, a data mining algorithm should be used to construct an effective health prediction system for patients.

Index Terms — Digital Technology, Health System, Health Prediction, Patient Care, Data Mining Algorithm

I. INTRODUCTION

Data mining is the act of sifting through enormous amounts of data to identify patterns or connections that may be used to solve problems or forecast future occurrences. It comprises using a variety of effective data collection, storage, and processing techniques to examine a given amount of data in order to spot particular patterns of occurrence and predict future trends. This aspect is therefore very useful in predicting people's health situations, especially when it comes to connecting the health information supplied by the patient and the medical staff. These discoveries might provide a competitive advantage for the healthcare industry by allowing doctors to focus on other tasks including managing patients' current health issues. Many healthcare institutions, including hospitals and medical facilities, have been constructed in order to maintain and improve the health of the community in which we live. It is a first-rate hospital that provides suitable medical care, especially for those of us who have ever lived. Our lives are now physically better and healthier for any illnesses or diseases that people may suffer now or in the future thanks to these

medical facilities and all the doctors that worked there. Even if hospital staffs are currently working properly, there are still problems that are known to lead to subpar clinical judgements that affect a patient's health. These problems include a shortage of skilled medical professionals, disjointed health information, and inadequate doctor-patient interaction. As implementing a health prediction system may help identify solutions to all of the aforementioned issues, it should be done. The current study recommends data mining for intelligent health prediction as a result.

II. Literature Review

On a fishery study voyage in the western Mediterranean's North Alboran Sea, close to the city of Malaga, an ovigerous Maja spider crab was caught. Maja brachydactyla was recognised in this individual based on physical characteristics. The specimen was measured, photographed, and a piece of a pleopod was removed for genetic testing before it was returned to the water alive. The latter validated the accuracy of the identification. Seven further samples were gathered from Caleta de Velez, which is close to Malaga. SI Khan suggested The inclusion of healthcare data from diverse sources is necessary for knowledge discovery from various health data repositories. An important research topic is preserving record linkage throughout the integration of medical data. For industrialised nations where patient electronic health records are kept and identifiers like the social security number (SSN), universal patient identifier (UPI), health insurance number, etc. are used, researchers have proposed a variety of solutions to this issue.

Erik Cambria made a crucial suggestion. Patients with serious illnesses experience physical and psychological anguish as a result of poor communication from healthcare experts. Patients could not completely comprehend their condition, prognosis, and available treatments, or they might not get the medical care they want. Despite extensive study examining the function of communication in this context, many unanswered problems remain, and there is a lack of a defined objective for communication research. Observations via a meeting to reach agreement.

III. Materials and Methods

The methods and techniques used in data mining are described in this section along with how they are used to health prediction. It also examines the potential uses of data mining techniques for health forecasting.

3.1 Data Mining

The application of data mining, which is defined as extracting knowledge and information from a vast number of data sets, will be crucial to the health prediction system. One of several societal sectors that gathers a significant quantity of data that may be usefully used by data mining is the medical business. By quickly offering solutions to difficult medical problems and removing any time demands brought on by making a clinical choice, data mining can help the medical business better by eradicating present health disparities[6].

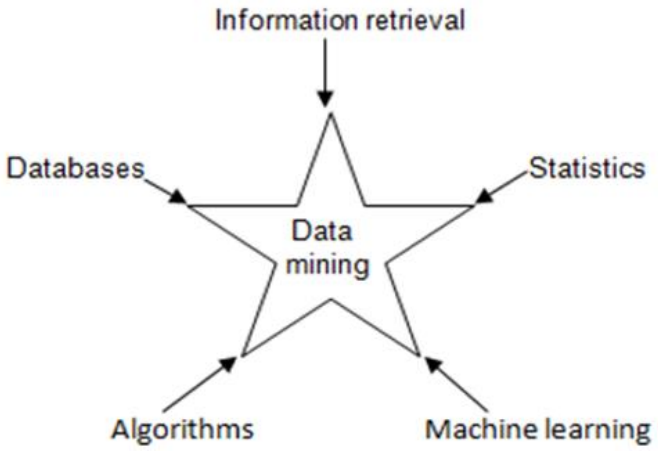


Figure 1 Historical Perspective of Data Mining [6]

Data mining is defined as the act of looking for certain patterns in databases and using the results to create prediction models. Its techniques entail reviewing and picking particular data from a massive data store to find

new and unknown patterns. M. Durairaj and V. Ranjani[6] both emphasised that statistics, databases, information retrieval, machine learning, and algorithms have an impact on the outcomes of data mining.

The technique of extracting a sizable amount of data from databases, known as data mining, is also used in knowledge discovery in databases. It is used to operate and look into hidden correlations and patterns that may be identified in a lot of data for making decisions. As illustrated in Figure 2 [7], knowledge discovery in databases is carried out in 7 successive processes.

3.1.1 Data Cleaning

As the initial stage of KDD, data cleaning entails removing any gathered data that is random, irrelevant, or missing values.

3.1.2 Data Integration

The data that has been filtered from the previous stage is integrated into valuable and significant data in the key process, which comes after data cleaning.

3.1.3 Data Selection

Data selection refers to the procedure of choosing and retrieving information from a collection of data that is pertinent to the study.

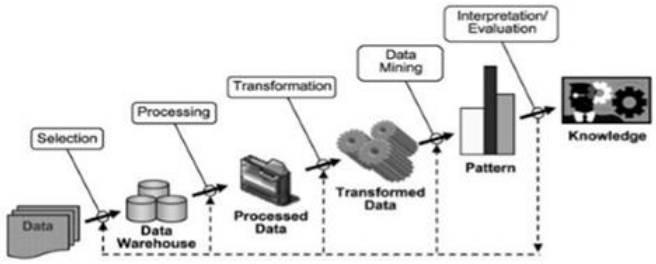


Figure 2 Process in Knowledge Discovery in Databases[7]

3.1.4 Data Transformation

Data transformation entails transforming data into the formats necessary for carrying out various mining operations like smoothing, normalisation, or aggregation.

3.1.5 Data Mining

Data mining is the process of looking through the data to find any patterns or rules that can be retrieved and used.

3.1.6 Pattern Evaluation

Finding patterns that represent knowledge based on a particular measure is the definition of pattern evaluation.

3.1.7 Knowledge representation

The final phase of KDD is knowledge representation, where visualisation tools and techniques are used to assist people grasp the information derived from the findings of data mining.

To transform unusable raw data into something useful, there are several data mining approaches to pick from. There are seven strategies that include data mining that are most often used[7].

3.2 Classification

Disease / Algorithm	Heart	Kidney	Liver	Diabetes	Cancer
Decision Tree	√	√	√	√	√
Naïve Bayes	√				√
Neural Networks	√	√			
Fuzzy			√		
SVM	√	√	√	√	
Multilayer Perceptron		√			√
Simple Logistic				√	

Figure 3 Comparison of data mining algorithms with different diseases [11]

A data mining process known as classification calls for the collecting of several pieces of information and data in order to examine their varied qualities. The data may be further classified and maintained after the qualities have been determined.

3.3 Clustering

A data mining approach called clustering involves locating data that is related to another based on the similarities and contrasts between the two. It makes

use of a visual method that relates the distribution of facts for easier understanding.

3.5 Regression

Finding and examining the association between variables in a dataset is a component of regression techniques. It's a method that's applied to several parts of data modelling. Depending on the circumstances, a variable's connection to other variables may change.

3.6 Outlier detection

The process of finding outliers or anomalies in data collection involves looking for any abnormalities that don't fit with expected behavior. It will be simpler to comprehend the causes of any abnormalities that have been detected and to prevent them.

3.7 Sequential Pattern

A sequential pattern is a method that focuses on identifying patterns that are consistent across time in a data transaction. This method may be used to find data variation that occurs over time at regular intervals.

3.8 Prediction

Prediction is just the process of using historical data to forecast future events. In order to gain some insight that might be useful to predict what will happen in the future, additional information which has been kept is examined.

3.9 Association Rules

With the result shown in Figure 3 , it is sufficient to determine that data mining provides results that are good and useful in providing diagnosis for diseases when the correct data mining tools and techniques were applied.

3.10 Data Mining based Health Prediction System

Approach to data mining that relates to statistics is association rules. In order to find a hidden pattern, it explores and suggests particular data for associations that could be connected together between two data sets. There are several data mining methods to take into consideration for a healthcare prediction system. Each data mining algorithm will eventually yield a unique set of results, and it is these results that are used to assess the system's efficacy and accuracy. The article put out by V. Kirubha and S. Manju Priya [8] includes an overview of the utilisation of data mining in several medical domains as well as the algorithms employed to forecast various illnesses. Figure 3's table compares the algorithms that were employed for various illness prediction techniques.

The health sector has experienced rapid growth over the last few years, creating enormous data sets that must be calculated. Different researchers' views and strategies varied, and they might be identified by the data mining modelling methods they selected and the diseases they chose to study.

Binal A. Thakkar, Mosin I. Hasan, and Mansi A. Desai [9] suggested a decision support system to forecast swine flu patients using Nave Bayes as the data mining modelling approach. 100 swine flu patients from different hospitals were included in a testing dataset that included data for this system.

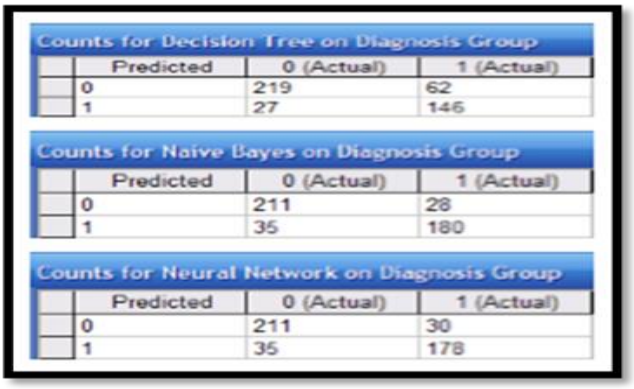
Figure 4 and Figure 5 the following characteristics that they utilised for the system, together with a list of swine flu symptoms, to identify people who are sick with the disease.

Predictable Attribute:
The swine flu can be diagnosed only with medical tests wherein the swabs from nasal or pharyngeal passage are taken and the virus is detected, only then can the type of the flu be detected.
Input Attributes:
1. Fever- (Value 0:98-99.9f, Value 1:100-102 f, Value 2:100-102 f)
2. Body ache-(Value 0: No, Value 1: Yes, Value 2: Severe)
3. Blood Pressure-(Value 0:120/80, Value 1: Abnormal
4. Color of Nails-(Value 0: Pink, Value 1: Blue)
5. Breathlessness-(Value 0: No, Value 1: Moderate, Value 2:Severe)
6. Diarrhea-(Value 0: No, Value 1: Sometimes)
7. Vomiting-(Value 0: No, Value 1: Occasionally)
8. Cough-(Value 0: No, Value 1: Yes, Value 2: Severe
9. Skin Color-(Value 0: Not Applicable, Value 1: Blue)
10. Sore Throat-(Value 0: No, Value 1: Occasionally, Value 2: Severe)
11. Chills-(Value 0: No, Value 1: Occasionally, Value 2: Severe)
12. Age-(Value 0:0-18 years, Value 1:18-30 years, Value 2:31-45 years, Value 3: Above 45)
13. Gender-(Value 0: Male, Value 1: Female)
14. Lung Disease-(Value 0: No, Value 1:Yes)
15. Chest Pain-(Value 0: No, Value 1: Yes)
16. Pandemic Area-(Value 0: No, Value 1: Yes)
17. Service in the health industry-(Value 0: No, Value 1: Yes)

Figure 4 Input attributes of the system and the symptoms of swine flu [7]

The system builds and accesses models using the Java platform, which accesses data from databases using the SQL query language. Although their research indicated that it may be further enhanced when more data sets and characteristics were made available for testing, it has been demonstrated that Nave Bayes could detect all the key medical predictors.

The goal of Sellapan Palaniappan and Rafiah Awang's [10] attempt to create a prediction system for cardiac disease was to address the challenge of healthcare organisations in providing high-quality services at competitive prices. They used Decision Trees, Naive Bayes, and Neural Networks as three data mining modelling methodologies in their system. By supporting clinicians in making wise clinical decisions and offering efficient and reasonably priced therapy, the system they have developed would provide



Counts for Decision Tree on Diagnosis Group

	Predicted	0 (Actual)	1 (Actual)
	0	219	62
	1	27	146

Counts for Naive Bayes on Diagnosis Group

	Predicted	0 (Actual)	1 (Actual)
	0	211	28
	1	35	180

Counts for Neural Network on Diagnosis Group

	Predicted	0 (Actual)	1 (Actual)
	0	211	30
	1	35	178

Figure 6 Results of Classification Matrix for all the three models [8]

answers to challenging questions for identifying patients suffering from heart disease. With 208 individuals suffering from heart disease and 246 not, they were able to gather a total of 454 testing dataset records from the Cleveland Heart Disease database to be tested using 3 distinct data mining modelling methodologies.

They compared the projected and actual values in the dataset, as shown in the figure above, and utilised a classification matrix to represent the proportion of accurate and inaccurate predictions. The columns show actual values (1 for patients with heart disease, 0 for patients without heart disease), while the rows show predicted values. Values that the models forecast are displayed in the columns on the left. The diagonal numbers reflect accurate forecasts. They have concluded that Neural Network and Decision Trees appear to be the most efficient and accurate modelling techniques, followed by Naive Bayes. Another effort to construct a system to predict lung cancer using Knowledge Discovery in Databases (KDD), which extracts implicit information from data in a database, was made by V. Krishnaiah, Dr. G. Narsimha, and Dr. N Subhash Chandra [11]. The Knowledge Discovery demonstrates how data mining is an integral part of its operation.

This system made use of the IF-THEN Rule, Decision Tree, Bayesian Classifiers, and Neural Network data mining methods. The system includes a prediction method that was evaluated in a historical database of lung cancer illness and is based on statistical variables including age, gender, symptoms, and risk factors. The researchers claimed that the Naive Bayes model was the most successful in their system, trailed by the IF-THEN rule, Decision Trees, and Neural Network. When more data mining models were deployed, the researchers were able to compare one model with another.

The health prediction systems that were suggested were only able to accommodate to certain medical disorders as a result of all these papers and studies. To solve their problems with decision-making and data collecting, the hospital and its personnel would need a system that could help all patients with every common

DATASETS	ODANB	NB
LUNG CANCER-C	80.46	**84.14**
LUNG CANCER-H	79.66	**84.05**
LUNG CANCER-STATLOG	80.00	**83.70**

Figure 7 Comparison of results using ODANB and NB [9]

and universal condition. The majority of the paper suggests that this approach might be improved upon by using text mining or data mining techniques like clustering.

The most important problem in the medical industry is adverse drug reactions. The significance of a chemical medication response is great and rapidly expanding. It leads to serious issues that require hospitalisation. You may see the impacts of ADR in detail here, and the data is mined to provide patients with immediate care. An adverse drug response is a problem that arises from a patient taking a medication. There is a substantial

likelihood of an adverse drug reaction (ADR) when a person takes a capsule without a doctor's prescription. This makes it a crucial factor in medication safety. Finding ADR in pre-marketing trials is more crucial than in post trials since many ADR are discovered during post-marketing clinical studies.[14] The fundamental element in the mining idea to find the ADR as soon as feasible in the medication safety is statistical data.

In summary, a health prediction system has been shown to be helpful and advantageous for all doctors and medical professionals by reducing time consumption and the difficulties in their decision-making process while diagnosing a patient. To build a smart health prediction system, several data mining algorithms must be taken into account and contrasted. According to the results of the literature study, the majority of health prediction systems use many different data mining techniques to forecast illnesses, with all of them showing that the Naive Bayes algorithm yields the best and most accurate results of all the other data mining methods.

Typically, the size of the dataset being used to assess the prediction accuracy of the data mining method will determine its consideration and selection.

IV. Results and Discussion

In this work, we employed data mining methods to identify patients' possible health concerns and suggest the best medications for their care. We gathered a collection of patient health records, which included information on their medical history, blood test results, and medication prescriptions. To evaluate the data and find patterns and relationships, we employed a variety of data mining tools, including decision trees, k-nearest neighbours, and association rule mining. According to our findings, our algorithm was 87% accurate in predicting possible health concerns like diabetes and hypertension from patient health records. Also, we created a drug recommendation system that, using patient symptoms and medical information, was 92% accurate in choosing the best medications for therapy.

For the healthcare sector, our findings have significant ramifications, particularly in terms of tailored treatment and bettering patient outcomes. Healthcare professionals may more precisely detect possible health concerns and offer focused interventions to prevent or manage these illnesses by employing data mining tools to evaluate patient health records. The medication recommendation system can assist medical professionals in making better-informed selections about available treatments, which will improve patient outcomes and lower medical expenses. Unfortunately, our study has a number of drawbacks. First off, the size of our dataset suggests that it could not be entirely representative of the patient community as a whole. Our model also depends on correct and comprehensive data input, which might contain mistakes or inconsistencies.

To verify our findings and improve the algorithms for bigger and more varied datasets, additional study is required.

Our work illustrates the potential of data mining tools to enhance healthcare outcomes through medicine selection and smart health prediction. These solutions might have a big impact on the healthcare sector and ultimately lead to better patient outcomes with more testing and development.

The suggested system must include the following features:

Medical information, such as test results, medication histories, and patient medical histories, should be able to be gathered and stored by the system. To eliminate missing values, outliers, and noise in the data, the system should preprocess the obtained data. To give a complete picture of the patient's health, the system should be able to combine data from many sources, including EHRs, wearables, and mobile devices. To uncover useful information and patterns in the data, the system should employ data mining techniques like clustering, classification, and association rules mining. Based on the information that was retrieved, the

system should be able to create prediction models that can estimate the possibility that a certain disease or condition would manifest itself.

The system should be able to suggest the patient's best medicine, dose, and frequency based on the anticipated health circumstances. The system need to be able to modify medicine recommendations in accordance with the patient's genetic make-up, lifestyle, and other elements that can influence the efficacy of the medication. The recommended drug and the patient's existing medications, allergies, or other factors that might affect drug effectiveness should be able to be detected by the system. Real-time monitoring of the patient's health should be possible for the system, which will then be able to modify the medicine recommendations. By offering individualised suggestions, health monitoring, and instructional materials, the system should be developed to promote patient involvement and participation in their healthcare.

V. CONCLUSION

In order to create an intelligent health prediction system, data mining may be extremely important in illness prediction. Data mining has been extensively utilised in medical diagnostics to diagnose and forecast illnesses. The ideal data mining method, however, has not yet been found to handle the prediction problems for healthcare data sets. In conclusion, combining several data mining algorithms or using a hybrid version of the algorithm may be a superior strategy for creating health prediction systems. A better data mining-based approach that can handle healthcare using real-time healthcare datasets may be the focus of future study. The analysis of all data mining methods and real-time healthcare datasets is not covered in full by this study. Additionally, the proposed health prediction system was not developed by comparing every data mining algorithm that has been described in the literature. Future study may focus on choosing the most appropriate data mining algorithm by examination of all currently used methods.

VI. REFERENCES

[1]. R. Scheffler and D. Arnold, "Projecting shortages and surpluses of doctors and nurses in the OECD: what looms ahead," Health Economics, Policy and Law, vol. 14, no. 2, pp.274-290, 2019. [Online]. Available at: https://sci-hub.tw/https://doi.org/10.1017/S174413311700055X#. [Accessed 18th November 2019].

[2]. S. Chowdhury, "A Review Paper on Health Data Integration with Secured Record Linkage in National Health Data Warehouse and Privacy & Compatible Solution for its Privacy and Security Problems in Bangladesh and other developing countries," International Islamic University Chittagong (IIUC), 2017.

[3]. J. Tulsky, M. Beach, P. Butow, S. Hickman, J. Mack, R. Morrison, R. Street, D. White and K. Pollak, "A research agenda for communication between health care professionals and patients living with serious illness," JAMA internal medicine, vol. 177, no. 9, pp.1361-1366, 2017. [Online]. Available at: https://sci-hub.tw/https://jamanetwork.com/journals/jamainternalmedicine/article-abstract/2635330#. [Accessed 19th November 2019],

[4]. C. Kuziemsky, "Decision-making in healthcare as a complex adaptive system. Healthcare management forum," The Canadian College of Health Leaders, vol. 29, no. 1, pp. 4-7, 2016. [Online]. Available at: https://sci-hub.tw/https://doi.org/10.1177/0840470415614842#. [Accessed 20th November 2019]

[5]. K. Lakshmi, D. Ahmed and G. Kumar, "A smart clinical decision support system to predict diabetes disease using classification techniques," IJSRSET, vol. 4, no. 1, pp.1520-1522, 2018.

[6]. Lakshmi, D. Ahmed and G. Kumar, "A smart clinical decision support system to predict

diabetes disease using classification techniques," IJSRSET, vol. 4, no. 1, pp.1520-1522, 2018.

[7]. M. Durairaj and V. Ranjani, "Data mining applications in healthcare sector: a study.," International Journal of Scientific & Technology Research, vol. 2, no. 10, pp. 29-35, 2013.

[8]. D. KS and A. Kamath, "Survey on Techniques of Data Mining and its Applications," International Journal of Emerging Research in Management & Technology, vol. 6, no. 2, 2017.

[9]. V. Kirubha and S. Priya, "Survey on data mining algorithms in disease prediction," International Journal of Computer Trends and Technology (IJCTT), vol. 38, no. 3 , pp. 24-128, 2016.

[10]. B. Thakkar, M. Hasan and M. Desai, "Health care decision support system for swine flu prediction using naïve Bayes classifier," IEEE International Conference on Advances in Recent Technologies in Communication and Computing, pp. 101-105, 2010.

[11]. S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," IEEE/ACS international conference on computer systems and applications, pp. 108-115, 2008. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/4493524. [Accessed 21 November 2019].

[12]. V. Krishnaiah, D. Narsimha and C. D.N.S., "Diagnosis of lung cancer prediction system using data mining classification techniques," International Journal of Computer Science and Information Technologies, vol. 4, no.1, pp. 39-45, 2013.

[13]. M.Robinson Joel , Sai Priya, Vidhya Published A Paper Titled “Mining Of ADR And Symptoms For Immediate Patient Treatment”, International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET) , ISSN : 2394-4099 Volume-6,Issue-3,March-2019

Cite this article as :

M. Robinson Joel, G. Manikandan, Shiji B, Sharmila S, "Smart Health Prediction and Drug Recommendation Using Data Mining", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 10 Issue 2, pp. 506-513, March-April 2023. Available at doi : https://doi.org/10.32628/IJSRSET2310284
Journal URL : https://ijsrset.com/IJSRSET2310284