International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 11, November 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

INTELLIGENT MEDICAL SUPPORT SYSTEM USING PREDICTIVE ALGORITHMS

Pinkesh Wala[1], Rinkle Anuwadia[2], Sachin Gavhane[3]

Supriya Mandhare[4]

[1,2]STUDENT INFT Department, Atharva College of Engineering, Mumbai University, (India)

[3,4]PROF INFT Department, Atharva College of Engineering, Mumbai University, (India)

ABSTRACT

The healthcare industry collects a huge amount of data which is not properly mined and not put to the optimum use. Discovery of these hidden patterns and relationships often goes unexploited but Medical diagnosis through Machine learning techniques of decision tree algorithm by learning pattern through the collected data of various diseases will help to develop intelligent medical decision support systems. In this article, we proposed disease prediction system using K-means, Apriori and ID3 predictive algorithms. Training dataset is used for prediction of particular disease. The main aim of this article is that to predict the disease which input symptoms is taken from patient or user. Recommend treatment and specialized doctors, so that patients can book an appointment if the results are positive.

Keywords : Machine Learning, Data mining, ID3 Algorithm, Apriori Algorithm, K-means

I. INTRODUCTION

1.1 Aims and Objectives

The health care domain have a lot of challenges and difficult task its one of the main difficult challenge is in disease diagnosis. Data mining can deliver an assessment of which courses of action prove effective by comparing and evaluating causes, symptoms, and courses of treatments. The detection of a disease from several factors or symptoms is a multi-layered problem and might lead to false assumptions frequently associated with erratic effects. Therefore it appears reasonable to try utilizing the knowledge and experience of several specialists collected in databases towards assisting the diagnosis process. The researchers in the medical field identify and predict the diseases besides proffering effective care for patients with the aid of data mining techniques.S

The goal of this technique is to find patterns that were previously unknown. Once we have found these patterns, we can use them to solve a number of problems. The goal of the person who uses data mining is, he/she should be able to predict certain behaviors or patterns. Once the user is able to predict the behavior of something which he analyzing, he will be able to make strategic decisions that can allow him to achieve certain goals. The main objective of this project is to create a fast, easy and an efficient mode for disease prediction, with less error rate and can apply with even large data sets and show reasonable patterns with dependent variables along with their

corresponding treatments. For disease identification and prediction in data mining appropriate algorithm should be used in order to maximize the accuracy rate. [1] [2]

1.2 Problem Statement

As mentioned in the objective, we believe there is a need for a classification of data, which take care of all the different aspects in analyzing the frequently occurring diseases. The information can be converted into knowledge about historical patterns and future trends. Our project devices a simple system, using various data mining algorithms to obtain better statistics from the available data. Health care industry today generates large amounts of complex data about patients, hospitals resources, diseases, diagnosis methods, electronic patient's records, etc. The data mining techniques are very useful to make medicinal decisions in curing diseases. The system planned to be developed will help in drawing effective conclusions. The discovered knowledge can be used by the health care administrators to improve the quality of service. In this project, we will develop a method to identify frequency of diseases in particular geographical area at given time period with the help of data mining tools. [1]

1.3 Scope of Project

A large population needs a great demand of doctor. But their deficiency create problem so console plays very important role for some extent. It will facilitate the users to predict the disease with the heip of the proposed system even if they are at remote location and very hard to reach doctors regularly provide they are within the range of the network coverage of internet. Less cost and time saving if proposed system integrates it to web portals.

II. REVIEW OF LITERATURE

TABLE 2.1 Papers referred for project selection

Sr. No.	Title	Authors	Approach	Publication
1	Prediction of Dengue, Diabetes and Swine Flu Using Random Forest Classification Algorithm, June 2017	Amit Tate, Ujwala Gavhane, Jayanand Pawar, Bajrang Rajpurohit, Gopal B. Deshmukh.	Prediction of Dengue, Diabetes and Swine Flu Using Random Forest Algorithm.	2017, International Research journal of engineering and Technology
2	Decision Support System for Medical Diagnosis Using Data Mining	D.Senthil Kumar, G.Sathyadevi, S.Sivanesh	Intelligent Medical decision support system for medical diagnosis of Hepatitis, Diabetes and Heart Diseases	2011, International journal of computer science issues
3	Application of training dataset using naïve bayes	Dr. Vani perumal, Shibu Samuel, Dr. P.	Prediction of stomach cancer in female	2017, International Journal of Scientific

	classifier for prediction of stomach cancer in female population	Indra, Muthu meena	population	Engineering and Technology and Research
4	Medical Diagnosis System using Machine Learning	Dhaval Raval, Dvijesh Bhatt, Malaram K Kumhar, Vishal Parekh, Daiwat Vyas	Prediction of Swine Flu using Dynamic Node Creation, feed-forward neural network and cascade correlation algorithms	2016, International journal of computer science and communication
5	Survey on Decision Support System for Medical Diagnosis Using Data Mining	Huzaifa Shabbir Dhorajiwala1 , Er. Asadullah Shaikh	Prediction of heart diseases, diabetes	2015, International journal of engineering research and general science.

2.1 Extracted Information

2.1.1 PREDICTION OF DENGUE, DIABETES AND SWINE FLU USING RANDOM FOREST CLASSIFICATION ALGORITHM

The aforementioned paper attempted to figure out the best approach to perform prediction of diseases such as Dengue, Diabetes and Swine Flu using RFA chief among them. The proposed system benefitted from this paper by observing the use of Classification algorithm such as RFA. (But in the proposed system we have used Apriori and ID3 algorithm for classification,) It assisted in the development of the premise of the proposed system as well as a tentative implementation of basic sentiment analysis. [1]

2.1.2 DECISION SUPPORT SYSTEM FOR MEDICAL DIAGNOSIS USING DATA MINING

The proposed system consists of 2 approaches, that is to say, data mining and machine learning. This particular article/paper helped provide insight into the process of how data can be mined from numerical data. It also assisted the proposed system in learning the kind of decision making when dealing with the medical data. In this paper, we propose the use of decision trees C4.5 algorithm, ID3 algorithm and CART algorithm to classify these diseases and compare the effectiveness, correction rate among them. [2]

2.1.3 APPLICATION OF TRAINING DATASET USING NAÏVE BAYES CLASSIFIER FOR PREDICTION OF STOMACH CANCER IN FEMALE POPULATION

The proposed system consists of 2 approaches, that is to say, data mining and machine learning. This particular article/paper helped provide insight into the process of how conventional methods of data mining go about predicting diseases as well as assisted the proposed system in learning where information extracted using machine learning could be helpful in aid of prediction accuracy to the traditional methods. [3]

2.1.4 MEDICAL DIAGNOSIS SYSTEM USING MACHINE LEARNING

This paper was studied so as to provide a more, well-rounded context as to the techniques related to machine learning used in the medical diagnosis. It helped the proposed system ascertain diagnosis process as the better companion to the traditional machine learning approach. This paper data mining techniques like SVM, Naïve Bayes, KNN on swine flu. [4]

2.1.5 SURVEY ON DECISION SUPPORT SYSTEM FOR MEDICAL DIAGNOSIS USING DATA MINING

As was the case with the previous paper, this paper was studied so as to provide a more well-rounded context as to the techniques related to data mining used in the medical diagnosis. It helped the proposed system ascertain mining patterns as the better companion to the traditional data mining approach and uses CART, ID3, LS-SVM and Genetic algorithms for classifying heart diseases and Diabetes and compares the usefulness and how effective it is. [5].

In conclusion, the proposed system sets out to implement diagnosis of maximum number of diseases as possible that are very much commonly diagnosed among the people along with their risk factors, their respective treatment and specialist doctor or physician that the patient should consult..

TABLE 2.2.2 Comparison Of Association Algorithm

Algorithm	Technique	Runtime	Memory usage	Parallelizability
Apriori	Generate singletons, pairs, triplets, etc.	Candidate generation is extremely slow. Runtime increases exponentially depending on the number of different items.	Saves singletons, pairs, triplets, etc.	Candidate generation is very parallelizable
FP-Growth	Insert sorted items by frequency into a pattern tree	Runtime increases linearly, depending on the number of transactions and items	Stores a compact version of the database.	Data are very inter dependent, each node needs the root.

In proposed project Apriori, ID3 and K-means clustering algorithms will be much more effective. Apriori is generally considered an unsupervised learning approach, since it's often used to discover or mine for interesting patterns and relationships. Apriori is well understood, easy to implement and has many derivatives.

On the other hand, the algorithm can be quite memory, space and time intensive when generating itemsets. The Apriori algorithm learns association rules and is applied to a database containing a large number of transactions. [6]

After analyzing the results of testing clustering algorithms, the following conclusions are obtained:

- As the number of clusters K becomes greater, the performance of SOM algorithm becomes lower.
- Performance of K-Means and EM is better than Hierarchical clustering algorithm.All algorithms have some noisy data which lowers the quality of hierarchical algorithms. The quality of K-Means and EM algorithm becomes very good when using huge dataset.
- K-Means and EM are very sensitive for noisy datasets.
- If using random dataset, hierarchical algorithms and SOM gives better results.
- Runing clustering algorithms using any a software gives almost same results even when changing any of the factors because most software use same procedures and ideas in any algorithm implemented by them.

K-Means algorithm produces more dense clusters than the hierarchical method especially when clusters are spherical. In the case of large number of variables, this algorithm has higher computing rate than hierarchical approach (If k is small). [7]

III. PROPOSED SYSTEM

Data mining based on association rule and generates the frequency of diseases affected by patients and also the number of patients affected by these diseases .Based on various geographical areas and at various time periods the study is made. Existing electronic medical details obtained from hospitals are utilized as training data set for analysis. The analysis and study concluded that the patients are affected frequently by different diseases during a particular year

3.1 Block Diagram

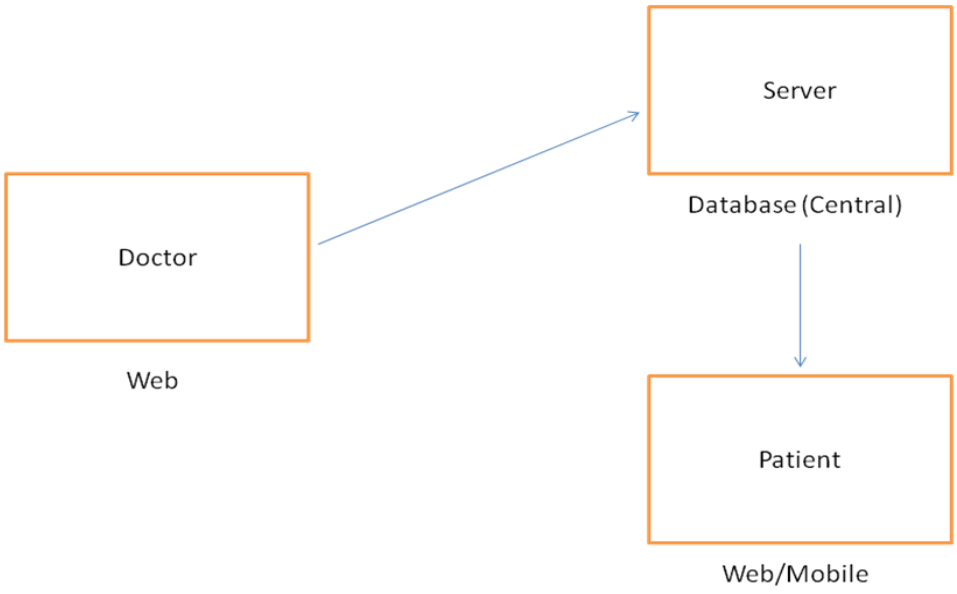

Fig 3.1 block diagram

International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 11, November 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

IV. DATASET

Main Entity in the project is Doctor. This User (Doctor) is responsible for entire system function. Doctor will create his own userID and password. Patient books an appointment with Doctor, Doctor will generate a Case Paper history for patient where all the medical record of the patient will be maintained. Details of Patient such as PatientID, Name, Address, Mobile No, Email ID, Gender, Date of Birth etc, where in PateintID is primary Key.

A single patient can visit doctor multiple times regarding multiple issues so it is necessary to maintain Patient's Visit record which includes VisitID, PatientID, Date of visit, Time of Visit, here VisitID is primary key and PatientID becomes foreign key.

To generate case history doctor has to enter Patient's medical data such as CaseID, VisitID, Case Description, Prescription provided from clinic or outside, Test recommended and Test reports. Here CaseID is primary key and VisitID becomes foreign key. Doctor will then enter data and symptoms into particular disease section. Where in our algorithm will run and diagnose repeated patterns. Proposed project algorithm will then execute and help doctor take some medical decisions based on the output of Algorithm. Disease related data is saved in particular data record.

TABLE 4.1 Dataset

ID	FIRST_NAME	LAST_NAME	GENDER	SUGAR LEVEL	DIABETES	DATE_OF_BIRTH
1	Manoj	Shah	Male	92	NO	22/06/1985
2	Pinkesh	Wala	Male	120	YES	08/02/1996
3	Manish	Jain	Male	105	NO	15/05/1995
4	Rinkle	Anuwadia	Female	150	YES	11/02/1996
5	Anita	Joshi	Female	95	NO	16/10/1980
6	Mahendra	Rai	Male	175	YES	14/05/1972
7	Priti	Shah	Female	160	YES	18/9/1990
8	Gopal	Pandey	Male	100	NO	01/09/1962

- This dataset primarily will contain the ID of the user that registers on the system. The user can be a Doctor or a Patient.
- The scope of the dataset can be set accordingly to any date,
- This dataset will be initially entered manually in an excel sheet. Then in future as our system will proceed ,data will be automatically generated due to the user login and functioning.

- As the dataset is of the actual values of the diseases, there is absolutely no need of cleaning the data. As more is the data, more will be the accuracy of our system

V. METHODOLOGY

Data owner (Doctor) will be able to create an account with unique id and can store data records into database server. And data consumer is able to access server data entered by data owner and to access the data they have the authority to access the server. Main Entity in the project is Doctor. This User (Doctor) is responsible for entire system function. Patient books an appointment with Doctor, Doctor will generate a Case Paper history for patient where all the medical record of the patient will be maintained. Our project algorithm will then execute and help doctor take some medical decisions based on the output of Algorithm. Patient will create account with unique id and password. Patient will then enter report values in our portal such as age, gender, weight, height, etc. Patient will start entering symptoms from which our algorithm will guide the patient about medical problems if any. If major medical problems exist then will book an appointment to visit doctor. Patient will first book an appointment with doctor, after which doctor will make case paper of patient. Based on study from our algorithm doctor will suggest medicines to patient. A database server is a computer program that provides database services. Users access a database server either through a "front end" running on the user's computer – which displays requested data or through the "back end", which runs on the server and handles tasks such as data analysis and storage. In the proposed project, the Water fall model is chosen for the development.

5.1 Flow Chart

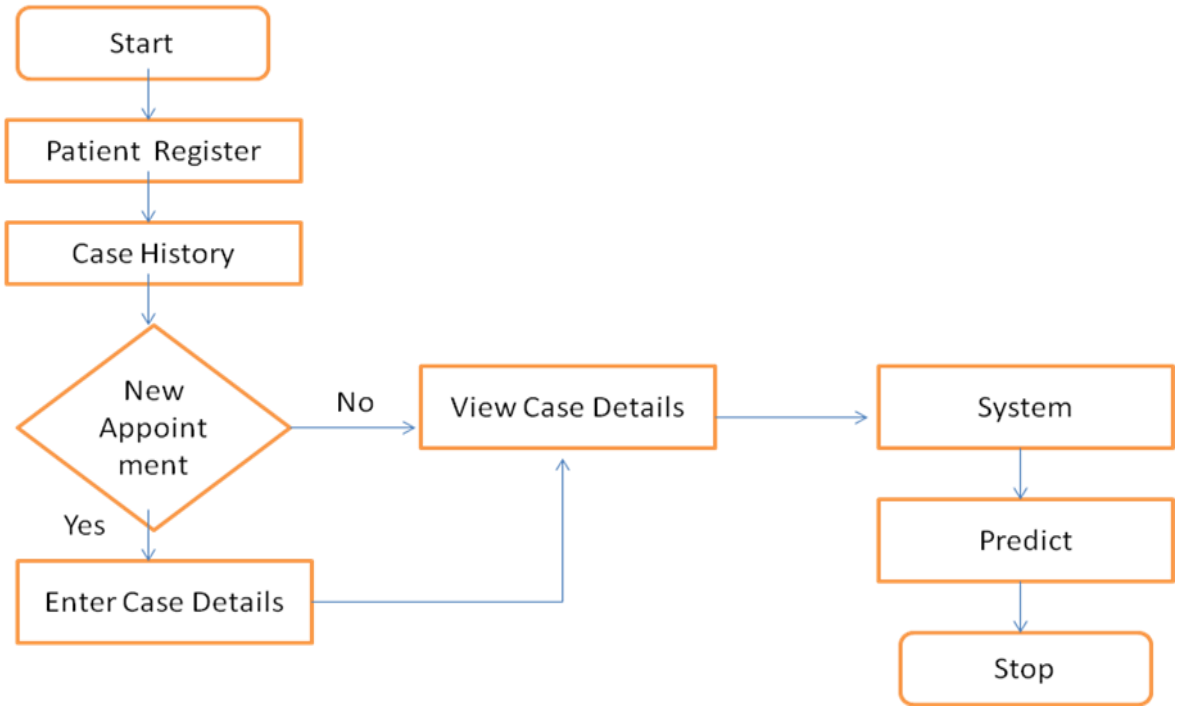

Fig 5.1 : flow chart

This diagram shows the entire flow of the system. Here patient first performs the login process. If the patient is new to the system, then he needs to sign in and enter the rest of the case details. If their exist any case history then depending on those reports, the doctor will generate the reports. It also allows the user to predict the disease with their corresponding treatment.

VI. HARDWARE AND SOFTWARE DETAILS

Table 6.1 Hardware and Software Details

Component	Minimum	Recommended
Processor	2.5 gigahertz (GHz)	Dual processors that are each 3 GHz or faster
RAM	1 gigabyte (GB)	2 GB
Disk	NTFS file system–formatted partition with a minimum of 3 GB of free space	NTFS file system–formatted partition with 3 GB of free space plus adequate free space for your Web sites
Drive	DVD drive	DVD drive or the source copied to a local or network-accessible drive
Display	1024 × 768	1024 × 768 or higher resolution monitor
Network	56 kilobits per second (Kbps) connection between client computers and server	56 Kbps or faster connection between client computers and server

Frond End and Back End Used:

Front End : Microsoft ASP.Net

Back End : Microsoft SQL Server

VII. CONCLUSION

The increasing ability of institutions to collect electronic data, facilitated by advanced in computer processing, means that the desire to mine" data is likely to expend. Proposed system is unique and different from the commonly used other system in data mining. The Proposed system can be proved efficient in terms of time and space complexity and proved to be accurate when compared with other system and can overcome the disadvantages of existing methods. This Proposed system can be enhanced by considering and incorporating many more parameters. For disease identification and prediction on the basis of various parameters such as age, sex, geographical areas, same algorithm can be applied. Algorithms are accurate according to the given scenario. The user can enter the symptoms to check the disease which is likely to affect him and can take preventive measures accordingly. The system can be used by researchers in order to predict future diseases. This algorithm can be enhanced by considering and incorporating many more parameters and creating a new hybrid algorithm which should be more feasible according to the given environment.

REFERENCES

Journal Papers:

[1] Amit Tate, Ujwala Gavhane, Jayanand Pawar, Bajrang Rajpurohit, Gopal B. Deshmukh, Prediction of Dengue, Diabetes and Swine Flu Using Random Forest Classification Algorithm,

International Research Journal of Engineering and Technology, ISSN: 2395-0072, Vol. 04 Issue 06| June 2017

[2] D.Senthil Kumar, G.Sathyadevi, S.Sivanesh, Decision Support System for Medical Diagnosis Using Data Mining, *IJCSI International Journal of Computer Science Issues, ISSN (Online): 1694-0814 Vol. 8, Issue 3*, No. 1| May 2011

[3] Dr. Vani perumal, Shibu Samuel, Dr. P. Indra, Muthu meena , Application of training dataset using naïve bayes classifier for prediction of stomach cancer in female population, *International journal of scientific engineering and technology research, ISSN 2319-8885, Vol 05, Issue. 45*| Nov 2016

[4] Dhaval Raval, Dvijesh Bhatt, Malaram K Kumhar, Vishal Parekh, Daiwat Vyas, *Medical Diagnosis System using Machine Learning. International journal of computer science and communication, DOI: 10.090592/IJCSC.2016.026, Vol. 7, No 1, pp.177-182*| Mar 2016

[5] Huzaifa Shabbir Dhorajiwala, Er.Asadullah Shaikh, *Survey on Decision Support System for Medical Diagnosis Using Data Mining. International journal of engineering research and general science, ISSN 2091-2730, Vol. 3*, Issue 6| Nov.2015

Books:

[6] Jiawei Han and Micheline Kamber, , *Data Mining Concepts and techniques* Morgan Kaufmann Publishers, San Francisco, CA, 2007.

Theses:

[7] D.S. Chan, Theory and implementation of multidimensional discrete systems for signal processing, doctoral diss., Massachusetts Institute of Technology, Cambridge, MA, 1978. (10, Times New Roman)

[8] Ahmad Shahin, Walid Moudani, Fadi Chakik, Mohamad Khalil, *Data Mining in Healthcare Information Systems*, Case Studies in Northern Lebanon Doctoral School for Science and Technology The Lebanese University Tripoli, Lebaon

Proceedings Papers:

[9] C-H. Weng, T. C-K. Huang. And R-P Han, *Disease predicton with different types of neural network classifiers, Telematics and informatics*, vol. 33, no. 2, pp. 277-292, 2016.

ACKNOWLEDGEMENT

It gives us immense pleasure in presenting this project synopsis report titled: “Intelligent Medical Support System using Predictive Algorithms”.

We would like to deeply express our sincere gratitude to our respected principal **Dr. Shrikant Kallurkar** and the management of Atharva College of Engineering for providing such an ideal atmosphere to build up this

project with well-equipped library with all the utmost necessary reference materials and up to date IT Laboratories.

We are extremely thankful to all staff and the management of the college for providing us all the facilities and resources required.

Submitted by:

Pinkesh Wala

Rinkle Anuwadia