A Survey on Smart Health Care Using Data Mining

Pinky Saikia Dutta, Sunayana Dutta, Tridisha Das, Sweety Buragohain, SusmitaSarma

CSE Department, Girijananda Chowdhury Institute Of Management & Technology
Hathkhowapara, Azara, Guwahati-17, India

Abstract— In this data-rich world, people are running out of information. This can be a matter of risk for the person who needs immediate remedies regarding their poor health. To unfold this hurdle, the concept of data mining is the best suited. Here, the traditional approaches have been replaced by smart technologies. The main purpose of data mining application in healthcare system is to develop an automated tool for identifying and disseminating relevant healthcare information. In this paper, we present a methodology for Prediction of diseases based on user input symptoms. It predicts probable diseases by mining data sets and provides suggested doctors and remedial solutions for Effective Treatment. It will also guide the users by giving tips to live a healthy life. We have discussed in detail on the existing Data mining techniques with data mining algorithms and its application tools which are more valuable for healthcare services.

Keywords—: Data Mining, E-health care, Prediction, Effective remedies.

I. INTRODUCTION

Health is one of the most important assets of our life which directly reflects in any form of progress or development. In today's hustle and bustle of life, most of the people neglect this asset which may be due to lack of time and complexity in the vast data available over the Internet.
Data mining has introduced various techniques which would overcome this problem and assist us to emphasize on both health and work simultaneously.
In present era, Data Mining is becoming popular in healthcare field because there is a need of efficient analytical methodology for detecting unknown and valuable information in health data. In health industry, Data Mining provides several benefits such as detection of the fraud in health insurance, availability of medical solution to the patients at lower cost, detection of causes of diseases and identification of medical treatment methods. It also helps the healthcare researchers for making efficient health care policies, constructing drug recommendation systems, developing health profiles of individuals etc. [1]

II. CONCEPT OF DATA MINING

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information.
Data mining holds great potential for the health care industry to enable health system to systematically use data and analytics to identify in efficiencies and least practices that improve care and reduce costs.
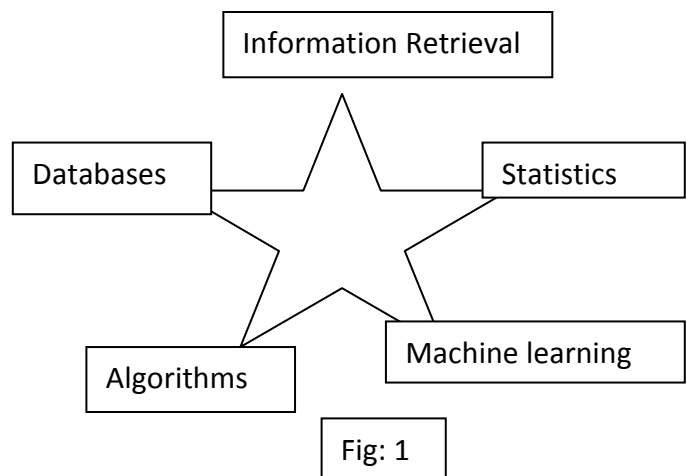
Fig: 1

The current evaluation of data mining functions and products is the results of influence from many disciplines, including databases, information retrieval statistics, algorithms and machine learning.

Fig.1 shows the historical perspectives of data mining. Data mining can answer questions that cannot be proven by other techniques. The key properties of datamining are:

1.Pattern Discovery:
Mining of data can be done by making new models. It uses a set of procedure to handle information. The concept of pattern discovery is the prototype of this model analysis. They can be used to mine the data on which they are built, but most types of models prefer new data.

2. Probable Outcomes:
Whatever data is mined, will surely give some results. These results may not be precise and they showcase all the possibilities that can occur with the data items.
3.Useful Results:
From whatever is obtained, we need to segregate the most useful results for the solution. This technique provided the same by applying the algorithm on data items.
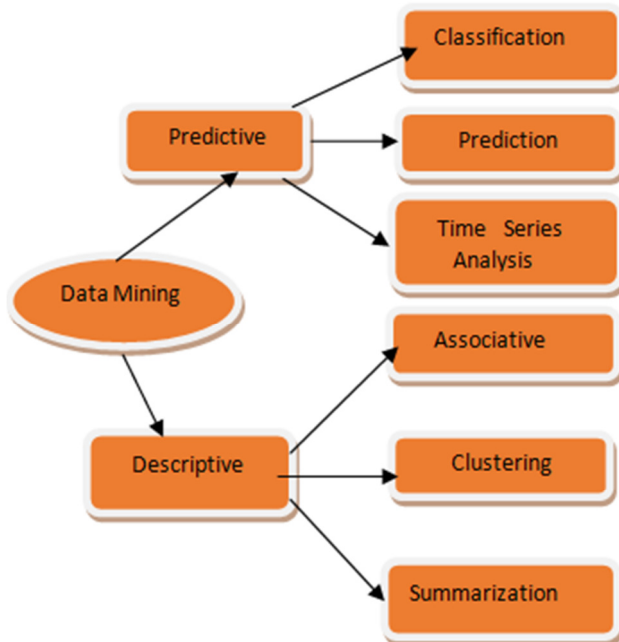**4. Data mining tasks:**

Fig.2: Data mining task

III. APPLICATIONS OF DATAMINING IN HEALTHCARE

Generally, all the healthcare organizations across the world, stored the healthcare data in electronic format. Healthcare data mainly contains all the information regarding patients as well as the parties involved in healthcare industries. Due to continuous increasing size of the electronic healthcare data, a type of complexity exists in it. The advancement in field of statistics, mathematics and very other disciplines it is now possible to extract the meaningful patterns from it that might indicate an attack by bio-terrorists. Data mining is beneficial in such a situation where large collections of healthcare data are available.

  Data Mining mainly extracts the meaningful patterns which were previously not known. These patterns can be then integrated into the knowledge and with the help of this knowledge essential decisions can be taken.

Data Mining Application Areas

Data mining has two broad categories of
Application areas which require new capabilities that are not currently being supplied by today's technology:
1. Business and E-Commerce

2. Scientific, Engineering and Health Care Data

Data mining applications in healthcare:

- **Treatment effectiveness**

→Provides better medical treatments at reasonable price

→Detection of diseases at early stages and chronic disease states of high-risk patients, etc.

- **Healthcare management**

→Number of days of stay in a hospital gets less

→Ranking of hospitals

→Estimation of medical staff

→Decision regarding health insurance policy

- **Customer relationship management**

Customer interactions may occur through-

→Call centers,

→Physicians' offices,

→Billing departments,

→In patient settings,

- **Fraud and abuse**

→Detect fraud and abuses & establish norms

→Identify unusual patterns of claims by physician's clinics.

→Inappropriate prescriptions or referrals

→Fraudulent insurance.

- **Medical Device Industry**

Healthcare system's one important point is medical device. For best communication work this one is mostly used.
→Mobile communications

75

→Low-cost of wireless biosensors have paved the way for development of mobile healthcare applications that supply a convenient way of monitoring of vital signs of patients.

→Ubiquitous Data Stream Mining (UDM) techniques such as light weight, one-pass data stream mining algorithms can perform real-time analysis on-board mobile devices while considering available resources such as battery charge and available memory.

- **Pharmaceutical Industry**

→Manage their inventories

→Develop new product and services.

→Deep understanding of the Pharma data knowledge helping firm's competitive position

- **HospitalManagement**

 Modern hospitals generate huge amount of data. Three layers of hospital management:

→Services for hospital management

→Services for medical staff

→ Services for patients

- **SystemBiology**

Biological databases contain a wide variety of data types with rich relational structure. So multi-relational data mining techniques are frequently applied to biological data [9]. Systems biology is at least as demanding as, and perhaps more demanding than, the genomic challenge that has fired international science and gained public attention.

In this study, the following list of medical problems has been analyzed.

- Heart Disease
- Cancer
- HIV/AIDS
- Blood
- Brain Cancer
- Tuberculosis
- Diabetes Mellitus
- Kidney dialysis
- Dengue
- IVF
- Hepatitis C

TABLE1. DATA MINING APPLICATIONS IN HEALTHCARE [6]

Types of diseases	Technique	Algorithm	Traditional method	Accuracy level(%) from DM Application
Heart Diseases	Classification	Naïve	Probability	60
Cancer	Classification	Rules, Decision Table		97.77
HIV/AIDS	Classification, Association Rule Mining	J48	Statistics	81.8
Blood Bank Sector	Classification	J48		89.9
Brain Cancer	Clustering	MAFIA		85
Tuberculosis	Naïve Bayes Classifier	KNN	Probability, Statistics	78
Diabetes Melitus	Classification	C4.5 Algorithm	Neural Network	82.6
Kidney Dialysis	Classification	Decision Making	Statistics	75.97
Dengue		C5.0	Statistics	80
IVF	Classification			91
Hepatitis C	Information Gain	Decision Rule		73.26

IV SURVEY OF RELATED WORK

K. Srinivas, B. Kavitha Rani and Dr. A. Goverdhan discusses mainly examine the potential use of classification based data mining techniques such as Rule Based, Decision tree, Naïve Bayes and Artificial healthcare data. Using an age, sex, blood pressure and blood sugar medical profiles it can predict the likelihood of patients getting a heart disease [2].

ShwetaKharyadiscussed various data mining approaches that have been utilized for breast cancer diagnosis and prognosis Decision tree is found to be the best predictor with 93.62% accuracy [3]

Arvind Sharma and P.C. Gupta discussedData mining can contribute with important benefits to the blood bank sector. J48 algorithm and WEKA tool have been used for the complete research work. Classification rules performed well

in the classification of blood donors, whose accuracy rate reached 89.9% [4].

 M. Durairaj, V. Ranjani in their study analyzed and evaluated the following list of medical problems- Heart Disease, Cancer, HIV/AIDS, Blood, Brain Cancer, Tuberculosis, Diabetes Mellitus, Kidney Dialysis, Dengue, IVF, Hepatitis C.They compared the results using Traditional and Data Mining approach and found the Data Mining approach to give better accuracy. Among the Data Mining Techniques, the Classification Technique resulted in highest accuracy of 97.77% in predicting Cancer[5]

GitanjaliJproposed study of huge datasets from various angles and obtaining gist of useful information. These methods are useful in detecting diseases and providing proper remedy for the same. [6]

S. Nagaparameshwara aims to evaluate various methods of data mining in applications to develop precise decisions and also provides a detailed discussion of medical data mining techniques can improve various angles of clinical predictions [7]

Priyanka Vijay Pawar, MeghaSakharamWalunj, PallaviChitte discussed how Apriori algorithm is used to find frequent data items and compared them with the existing algorithms, how data mining techniques can be applied on medical data which has abundant scope to improve health solutions,how electronic health records and other historical medical data can prove miracles if used for a right purpose,how huge amounts of complex data generated by health care sector includes details about diseases, patients, diagnosis methods, electronic patient's details hospitals resources[8]

V. SYSTEM DESIGN

This system tends to replace the existing system for going to the doctor for getting diagnosis on illness you are suffering from with a smart solution where you get instant diagnosis on entering symptoms in the system. The main features of this system will be giving instant diagnosis on the user entered symptoms and getting home remedies for emergency purpose. The diagnosis will be done here with the help of data mining techniques.

Figure 3 shows the overview and flow of the software model activities. The left section describes the user or patient activities which can be viewed by the user and the right section describes the server zone where the entire activities such as serving the queries of user will be done by the software itself.

Figure 3: software model
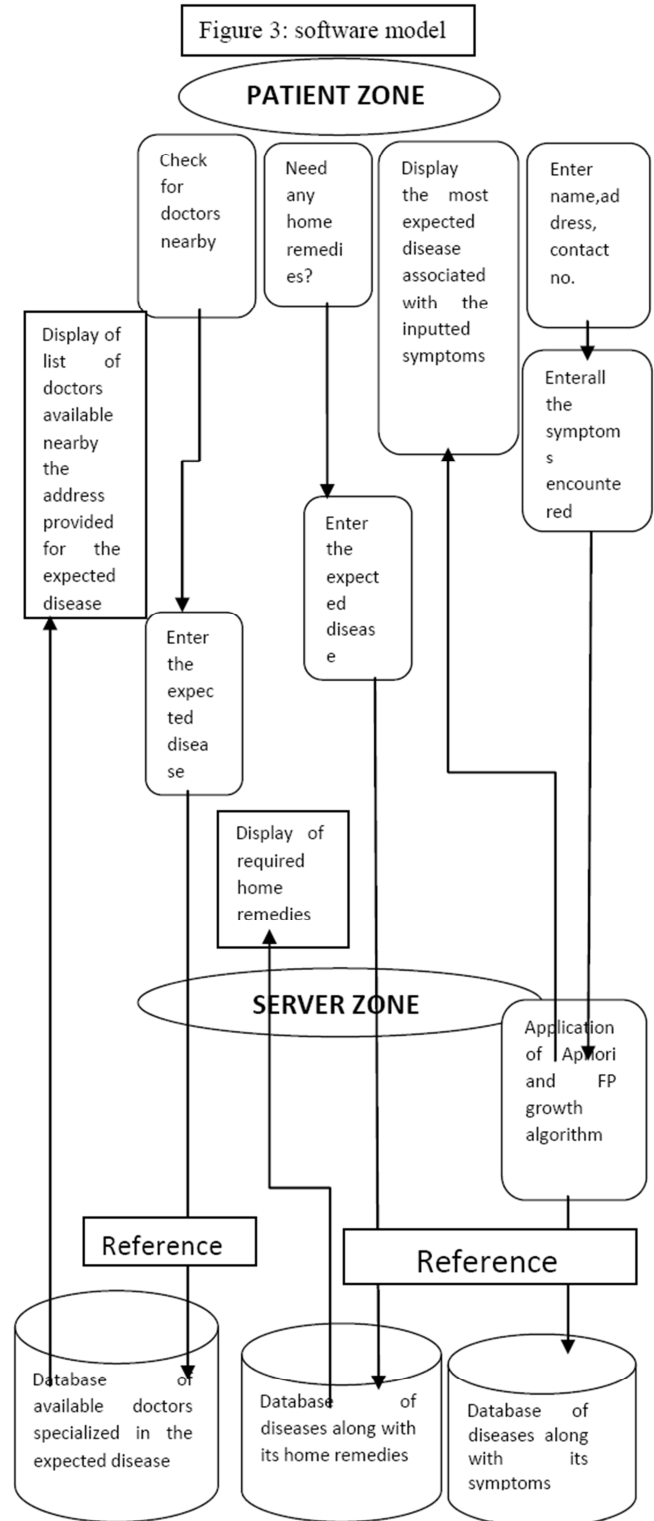
Following are the data mining Algorithms that we are going to use here:

• **Association Rule and mining frequent patterns:** The central task of association rule mining is to find sets of

binary variables that co-occur together frequently in a transaction database [8], while the goal of feature selection problem is to identify groups of that are strongly correlated with each other with a specific target variable. Association rule has the several algorithms like: Apriori, FP growth.

- [1]Apriori is mainly used for common diseases and FP growth for acute and chronic diseases.

- If the inputted symptoms by a user searches for a common disease, then apriori algorithm is best suited for it but if the software does the analysis on the basis of huge data large number of candidate sets then FP growth algorithm is preferred.

Here basically we are going to use three databases:

- Database for various common and chronic diseases along with their symptoms

- Database for remedial solutions needed to be taken for all those diseases constituting the previous database.

- Database for various specialized doctors in a particular disease available at nearby places in accordance to the address information provided by the user

Following are the, features we are going to include internally in the software:

- **Patient Login:**Patient Login to the system using his ID and Password.

- **Patient Registration:**If Patient is a new user he will enter his personal details and he will user Id and password through which he can login to the system.

- **My Details:**Patient can view his personal details.

- **Disease Prediction:**Patient will specify the symptoms caused due to his illness. System will ask certain question regarding his illness and system predict the disease based on the symptoms specified by the patient and system will also suggest doctors based on the disease.

- **Search Doctor:**Patient can search for doctor by specifying name, address or type.

- **Feedback:**Patient will give feedback this will be reported to the admin.

- **Doctor Login:**Doctor will access the system using his User ID and Password.

- **Patient Details:**Doctor can view patient's personal details.

- **Notification:** Doctor will get notification how many people had accessed the system and what all are the diseases predicted by the system.

- **Admin Login:** Admin can login to the system using his ID and Password.

- **Add Doctor:** Admin can add new doctor details into the database.

- **Add Disease:** Admin can add disease details along with symptoms and type.

- **View Doctor:** Admin can view various Doctors along with their personal details.

- **View Disease:** Admin can view various diseases details stored in database.

- **View Patient:** Admin can view various patient details that had accessed the system.

- **View Feedback:** Admin can view feedback provided by various users.

VI. CONCLUSION AND FUTURE WORK

In this system, hidden knowledge will be extracted from the historical data by preparing datasets by applying Apriori algorithm and FP growth algorithm.
Predicting smart health can be done only if system responds that way. These datasets will be compared with the incoming queries and the final report will be generated using Association Rule Mining. Since this proposed methodology will work on real historical data, it will provide accurate and efficient results, which will help patients, get diagnosis instantly. This system will also guide the users of how to take remedial actions stay cautious.

The further enhancements that can be done would be integrating this web application in an Android app. This will be available to users on mobile basis and its use can be further increased. Also feature like getting the doctor online on chat so that patients can directly talk to the concerned doctors. This will make this web application predictable in true sense.

References

[1] Han, J, Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006

[2] K. Srinivas, B. Kavitha Rani and Dr. A. Govardhan, ―Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks International Journal on Computer Science and Engineering (2010).

[3] ShwetaKharya, ―Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease‖, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2, April 2012.

[4] Arvind Sharma and P.C. Gupta ―Predicting the Number of Blood Donors through their Age and Blood Group by using Data Mining Tool‖ International Journal of Communication and Computer Technologies Volume 01 – No.6, Issue: 02 September 2012.

[5] M. Durairaj, V. Ranjani- Data Mining Applications in Healthcare Sector: A Study International Journal of Scientific& Technology Research Volume 2, Issue 10, October 2013

[6] Gitanjali J., C.Ranichandra, M. Pounambal, ―APROIRI algorithm based medical data mining for frequent disease identification, ‖ Univ. VIT, Vellore, 4, April 4, 2014, Vol 2.

[7] S. Nagaparameshwara Chary, B.Ram, "Analysis of Classification Technique Algorithms in Data mining- A Review", International Journal of Computer Sciences and Engineering, Volume-04, Issue-06, Page No (180-185), Jun -2016,

[8] Priyanka Vijay Pawar, MeghaSakharamWalunj, PallaviChitte, Estimation based on Data Mining Approach for Health Analysis,International Journal on Recent and Innovation Trends in Computing and Communication

[9]. David Page and Mark Craven,―Biological Applications of Multi Relational Data Mining‖.

AUTHORS PROFILE:

Mrs. Pinky SaikiaDutta, BE from Jorhat Engineering College, Assam and ME from Guwahati University in 2002 and 2014 respectively. She is currently working in Girijananda Chowdhury Institute of Management and Technology, Guwahati, Assam, India. She is a member of IEEE since from 2015. Her main research work focuses on Data mining, its algorithm. She has 13 years of teaching Experience.

Sweety Buragohain, pursuing B.Tech in Computer Science and Engineering deparment from Girijananda Chowdhury Institute of Management and Technology, Guwahati, Assam, India. Her area of interest is in Data Mining and Database Management System.

Tridisha Das, pursuing B.Tech in Computer Science and Engineering deparment from Girijananda Chowdhury Institute of Management and Technology, Guwahati, Assam, India. Her area of interest is in Data Mining.

Susmita Sarma, pursuing B.Tech in Computer Science and Engineering deparment from Girijananda Chowdhury Institute of Management and Technology, Guwahati, Assam, India. Her area of interest is in Data Mining.

Sunayana Dutta, pursuing B.Tech in Computer Science and Engineering deparment from Girijananda Chowdhury Institute of Management and Technology, Guwahati, Assam, India. Her area of interest is in Data Mining.