

# Improving the Quality of Healthcare Using Big Data

Dr. Ashfaq Amir Shaikh, Singh Nikhil Chandrashekhar, Ayshan Maredia,  
Khan Faimin Shakeel, Saif Sayyed

**Abstract**— In India there is a lack of doctor's availability in rural areas compare to urban areas because of which the number of deaths is increasing in the rural areas. To solve this issue we are building an android application (Healtho) which will recommend the disease based on the symptoms given by the end user. Basically, a recommended system will be used by using Hadoop with mahout that is a Big Data concept. By using android as a platform we can provide higher availability of the system to the end user and provide some emergency services like location of nearby Hospitals and blood bank. The system also provides the medicine time (Meditime) in which the end user may come to know at what time the medicine is to be taken. This system could mostly be used by the people who live in rural area because there is lack of doctor's availability and hospitals.

**Index Terms**— Big Data Analytics, Recommendation Engine, NoSQL, Apache Hadoop, Apache Mahout, MongoDB.

## I. INTRODUCTION

In Today's world the big data playing a very important role in various field from E-commerce to Education so why not in healthcare. Every second internet, IoT devices and mobile generate large amount of data many of this data will never use for anything i.e. wastage. But this data can be store and can be analyze to answer many real world question and solving many real world problem. Suppose a person walk 2km in a day very first time it look like this data is not so useful and if it is useful then it's use is limited to that specific user only but if this data is collected and analyze we can conclude the average distance people walk in a day. Our proposed system is Big Data based system with Android as an end device which will recommend all possible disease based on symptoms provided by user.

## II. PROBLEM STATEMENT

In India as the growth rate of the people is increasing, the diseases and also the death due to them is also increases. So there must be a proper mechanism of finding the diseases, by its symptoms, at a very early stage before it grows into a major problem. In India the 62% of death is due to

**Dr. Ashfaq Amir Shaikh**, Information Technology, M. H. Saboo Siddik College of Engineering, Mumbai, India

**Singh Nikhil Chandrashekhar**, Information Technology, M. H. Saboo Siddik College of Engineering, Mumbai, India

**Ayshan Maredia**, Information Technology, M. H. Saboo Siddik College of Engineering, Mumbai, India

**Khan Faimin Shakeel**, Information Technology, M. H. Saboo Siddik College of Engineering, Mumbai, India

**Saif Sayyed**, Information Technology, M. H. Saboo Siddik College of Engineering, Mumbai, India

preventable diseases which can be prevented if it is known earlier [1]. At

the same time India suffering from high shortage of doctor According to WHO the recommended Doctor to Patient ratio is 1:1000 i.e. One doctor to every 1000 patient but in India the doctor to patient ratio is 1:30000 i.e. there is one doctor to every 30000 patient in rural area [2] which is worst. This gap can be minimize by using big data and recommendation system user can get the all possible disease based on their symptoms and can take preventive measure to cure it at early stage and on right time.

## DEATHS FROM DISEASES ACROSS COUNTRIES

Age-standardized mortality rates by causes  
(per 100 000 population)

	Communicable diseases	Non communicable diseases
Germany	22	365
UK	29	359
US	31	413
Japan	34	244
China	41	576
Turkey	44	555
Mexico	58	468
Maldives	59	487
Russian Federation	74	790
Sri Lanka	75	501
Brazil	93	514
Malaysia	117	563
Indonesia	162	680
Global average	178	539
Bangladesh	235	549
Nepal	252	678
India	253	682
Pakistan	296	669
Myanmar	316	709
Afghanistan	363	846
South Africa	612	711

Source: NSSO, WHO

Fig 1: Death from Diseases [3].

Second problem is that after getting know about disease next stage is consult doctor and take medicines and exercises but people are so busy that they forget sometime to take medicine and to do exercises this will put large impact on their health condition. India has highest TB case in the world total 25% of world TB cases are in India [4]. More than 1000 Indians die due to TB every day [5]. This can be solve if user get time to time alert or notification for medicine and exercises.

Third problem is according to WHO about 1.25 million people die each year as a result of road traffic crashes [6]. And many of this death is due lack of emergency services and

help from local people. This problem can be solve by using GPS based assistance providing all hospital and emergency centers nearby. And other solution to this problem is that if user ask a help from friend using short message and sharing their location

### III. EXISTING SYSTEM

#### 1. WebMD Symptoms Checker:

WebMD is web and mobile based symptoms checker that take symptoms as an input from user and provide all possible disease. WebMD also provide the virtual interaction to the user. But the language that WebMD use to recommend disease is very complex and sometime it is not understand by the common people. WebMD mostly communicate through high medical term which is not understandable to the common people other than doctors.

#### 2. Mayo Clinic:

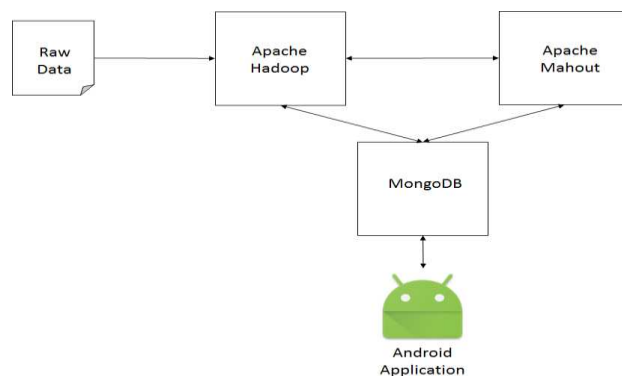
Mayo Clinic is web application which provide symptom checker services and language that it use is also complex and sometime common people don't understand. Especially in India where every region has their own common language. Also mayo clinic is web application so it is not always accessible to user.

### IV. PROPOSED SYSTEM

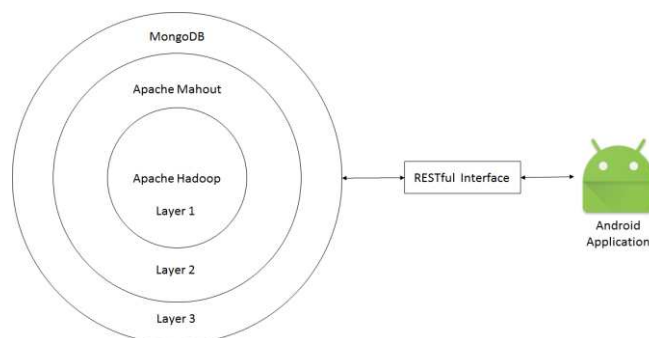
The proposed system is Big Data based system with android as an end device. This system will not only provide the possible disease but also provides the lifesaving feature in the case of emergency or accident. Healtho uses is not only limited to specific user but it can be used to maintain all the medical information of family. Healtho also provide medicine taking alert as Meditime. Meditime help user to remember to take medicine. This system store the emergency contact so that in the case of emergency user can communicate with it. This emergency contact is Emergency Friend it is a contact of one of its friend. In emergency user just need to click on button and his/her friend will receive short message like Need Help. User can also check hospital and Emergency center including blood bank nearby to its location with help of hospital locator this will also help user in case of emergency.

Our system is working on the base of past user record this. User Recommendation System For Example There is two user has same symptoms there is the possibility that they have same Disease. Symptom checker is developed using Apache Mahout on top of Apache Hadoop. While Hadoop is for storing and processing and mahout is for recommendation system. Also MongoDB to store unstructured data.

In our System the data is collected in the form of raw or log data then the useful information like disease and there symptom are extracted and store in apache hadoop and then to MongoDB. And During retrieval process Apache Mahout will access the data from MongoDB and generate recommendation result this result is store back to MongoDB. After this the result will available on android application.



**Fig 2: Simple Architecture**



**Fig 3: Layered Architecture**

#### A. What is Big Data?

A report delivered to the U.S. Congress in August 2012 defines Big Data as “a term that describes large volumes of high velocity, complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.”[7].

Big Data Work on the basis of 5 Dimension which is commonly referred as 5V's: Volume, Velocity, Variety, Veracity and Value [8], [9].

##### 1. Volume:

The healthcare data are mostly in form of Images, Personal Information and Biometric System Readings this data require large data storage and processing capacity. The KPMG report [10] stated that the healthcare data volume has reached 150 exabytes in 2013, and it is continuously increasing at the rate of 1.2 – 2.4 exabytes a year. Volume of data plays important role while making decision whether to use big data or not.

##### 2. Variety:

All data which are available are not always in structured form some of them are unstructured and some are in semi-structured form. Big data combine all the form of data and give some closer report and result.

##### 3. Velocity:

The data storing and updating process require large amount of time. During this process system remain ideal. This problem can be solve by using big data. Big data can store, update and analyze the data in considerably in very short amount time.

4. Value:

The Value of data depends upon the data we are looking for. Many people may have different value of data on the same database based on their point of view and requirement.

5. Veracity:

The veracity of data is deal with quality of data. This quality get differ from structured data to unstructured. In unstructured data the quality is big challenge this problem can be solve using big data.

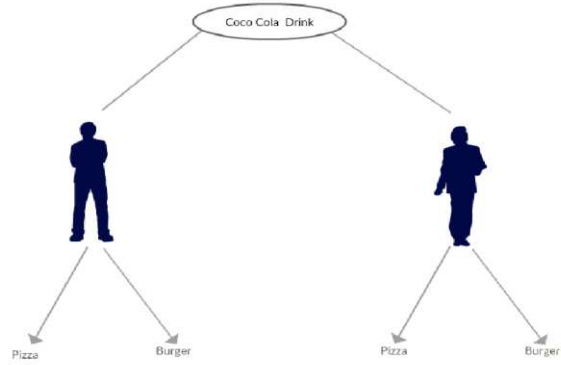


Fig 4: Collaborative Filtering

Table 1: Comparison between Traditional Database and Big Data

PARAMETERS	TRADITIONAL DATABASE	BIG DATA
Data	Suitable for Structured Data	Suitable for Structured, unstructured and Semi-Structured data
Processing	OLTP	Analytical, Big Data Processing
Choice	When the data needed Consistent Relationship	Does not require any consistent relationship between data
Processor	Needs Expensive Hardware	Does not require any expensive hardware

2. Content – Based Filtering

It is also known as cognitive filtering. This approach work on the basis of comparing item content and user profile. Content based filtering work on the similarities between items rather than similarity between users.

D. What is mahout?

Apache mahout is an open source project that is for creating scalable machine learning algorithm. Some popular machine learning technique in apache mahout are:

1. Recommendation System
2. Classification
3. Clustering

E. What is NoSql?

NoSql is non-relational, open source, distributed and horizontally scalable next generation database. It is also known as Non Sql and Not only Sql database.

Table 2: Comparison between SQL and NoSQL

SQL	NoSQL
Rational Database	Non- rational, Distributed Database
Vertically Scalable	Horizontally Scalable
Table Based Database	Document Based, Graph Based or Key- value Pair
Pre-Define Schema	Dynamic schema
Uses SQL	Uses UnQL (Unstructured Query Language)
Not Preferred for Large Datasets	Largely Preferred for large Datasets

A. What is Apache hadoop?

Apache Hadoop is solution to big data problem. It is open source framework which support distributed processing of large set of data. The main component of Apache hadoop is

1. Hadoop Distributed File system (HDFS)
2. MapReduce

C. What is Recommendation Engine?

Recommendation engine is fast growing area in today’s world. Many E-commerce company use recommendation system to target customer. It is a subpart of information filtering system.

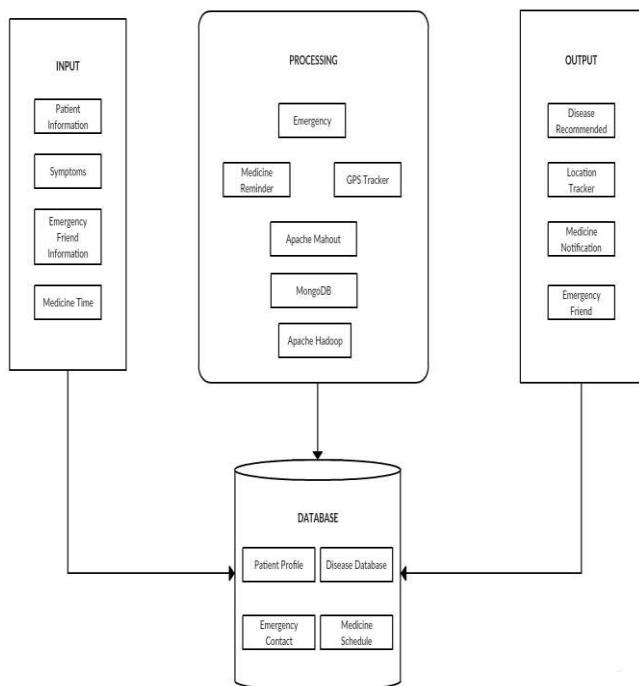
There are Two Approach to build recommendation system:

1. Collaborative Filtering:

Collaborative filtering is based on collecting and Analyzing behavior, activity and test of one user and predicting what other user with same behavior, activity and test will like.

F. What is MongoDB?

MongoDB is open source document oriented NoSql Database.



**Fig 5: Block Diagram of System**

## V. FEATURE OF THE SYSTEM

### A. Symptoms Recommender:-

The Symptoms recommender will use to recommend disease based on entered symptoms which will help user to start preventive measure.

Symptom recommender not only recommend the symptom also it will recommend the preventive measure to be taken to overcome those disease.

### B. Meditime:-

The Medicine Time will help user to remember his/her medicine. It is type of notification cum alarm that notify user or patient to take there or their family member medicine.

### C. Emergency Friend.

Emergency friend is a user friends (Minimum 2) to which user can send short message and share their location and ask for help from them.

### D. Hospital and Blood Bank Locator

In case of emergency the user can search all the emergency centers and hospital nearby.

### E. Feedback

As our system is building on the past user/patient response, Where user/patient can share their disease, symptom and there undergone treatment this will help other user or patient.

## II. FUTURE SCOPE

- An Addition to above mention functionality NLP (Natural Language Processing) can make system more flexible and user friendly.
- Using IoT and getting real time data will allow user to track their health information.

## REFERENCES

- [1] The Lancet "Causes of neonatal and child mortality in India: a nationally representative mortality survey"
- [2] Sudhir Anand, Victoria Fan. "The Health Workforce In India," Human Resources for Health Observer Series No. 16. WHO
- [3] <http://www.livemint.com/Opinion/DSH1OnDr2LG0zAcHh129XJ/The-growing-burden-of-healthcare-costs.html>
- [4] WHO. Tuberculosis control in South-East Asia Region. Regional Report. 2012
- [5] USAID-India. Health overview. 2011
- [6] WHO. Road Traffic Injuries. Fact Sheet. 2016
- [7] Institute for Health Technology Transformation, "Transforming Health Care Through Big Data", Strategies for leveraging big data in the health care industry
- [8] M. A. U. D. Khan, M. F. Uddin, and N. Gupta, "Seven V's of Big Data understanding Big Data to extract value," *Proc. 2014 Zo. 1 Conf. Am. Soc. Eng. Educ. - "Engineering Educ. Ind. Involv. Interdiscip. Trends"*, *ASEE Zo. 1 2014*, 2014
- [9] Hiba Asri, Hajar Mousannif, Hassan Al Moatassime, Thomas Noel, "Big Data in healthcare: Challenges and Opportunities",
- [10] V. Galloro, "Prime numbers." *Mod. Healthc.*, vol. 38, pp. 14-16, 2008.

**Dr. Ashfaq Amir Shaikh** Asst. Professor at Department of Information Technology of M.H. Saboo Siddik College of Engineering, Mumbai, India

**Singh Nikhil Chandrashekhar** is a student pursuing Bachelor of Engineering in Information Technology from M.H. Saboo Siddik College of Engineering, Mumbai, India.

**Ayshan Maredia** is a student pursuing Bachelor of Engineering in Information Technology from M.H. Saboo Siddik College of Engineering, Mumbai, India.

**Khan Faimin Shakeel** is a student pursuing Bachelor of Engineering in Information Technology from M.H. Saboo Siddik College of Engineering Mumbai, India.

**Saif Sayyed** is a student pursuing Bachelor of Engineering in Information Technology from M.H. Saboo Siddik College of Engineering Mumbai, India.