Role of Big Data Analytics in Rural Health Care - A Step Towards Svasth Bharath

Muni Kumar N*1, Manjula R*2
*Research Scholar, SCSE, VIT University, Vellore, Tamil Nadu, India.
*Associate Professor, SCSE, VIT University, Vellore, Tamil Nadu, India

Abstract — Health care is one of the greatest concerns in India. While, those living in cities and big towns have access to high end health services, the millions of people living in rural India, particularly in the remote parts of the country face problems of inadequate facilities and poor access to healthcare. Many experts, including researchers, policy makers and practitioners identified that, there is a big gap in the knowledge about innovations in public and private health financing and delivery. The inefficiencies and inequities in the public health care access in India have pushed forward the need for creative thinking and innovative solutions to strengthen the same. The problems existing in the health care scenario provides apparent calls for the need to change the existing structure of the present health care services by applying big data analytics. This paper identifies the massive shortage of proper health care facilities and addresses how to provide greater access to primary health care services in rural India. Further, it also addresses the critical computing and analytical ability of Big Data in processing huge volumes of transactional data in real time situations to turn the dream of Svasth Bharath (Healthy India) into reality. The objective of this paper is to present the reforms in the health care sector and boosts the discussions on how government can harness innovations in the big data analytics to improve the rural health care system.

Keywords— Big Data Analytics, HealthCare, Rural Health Care, e-Health Care, Tele Medicine, Svasth Bharath.

I. INTRODUCTION

India is a strong country with billion plus people, one of the world’s fastest growing economy, 29.5% of population are below poverty line, 46% of children are malnourished yet. Increasing population in India over-burdens the health care structure in the country. Health care in our country is government financed and government run. But, for many people living in the remote parts of the country, accessing primary health care is still a challenge. For the developing countries like India, the health care is one of the most important areas to emphasis, so as to provide better health care access to the priceless human resources, which in turn can make the India healthier too. The exponential growth [1] of data over the last decade has introduced a new domain in the field of information technology and data science called Big Data. The term Big data is often used to describe a massive volume of data (both structured and unstructured) that is so large and difficult to process using traditional database management techniques. As the health care industry is flooded with huge volumes of data which needs validation and analysis, Big Data Analytics can be applied. Big data has the potential to perform critical computing and analytical ability towards the processing of the huge volumes of transactional data.

As per the statistics available in the Global Health Observatory Data Repository [2], a World Health Organization (WHO) repository, the per capita government expenditure on health care in India during 2011 is at an average of $44, compared to that of $4047 in the USA. The outcome in USA is long lives (increase in the life expectancy), full of sophisticated facilities for health care system, efficient clinical staff, round the clock emergency services and world-class doctors. To provide better e-health care services to the massive Indian people with right care for the right disease at the right time, the big data analytics can be applied. The term, Telemedicine uses the electronic communication technology to exchange patient information among doctors / hospitals and provide health care services at remote locations equivalent to the services rendered by the city hospitals. This innovative technology is gaining increasing attention as a way to improve the performance of health care systems by linking various systems via a data and communications platform to reduce redundant medical tests, improve and expedite clinical decision making, and enable access to all levels of healthcare for a wide range of conditions. With telemedicine, hospitals hope to lower the cost of patient care and increase the effectiveness of chronic disease management. It collects all possible patient information to create thorough electronic health records (EHR’s) for each patient.

Historically, the health care industry has generated huge amounts of data, driven by record keeping, compliance and regulatory requirements and patient care [5], which is considered as big data. In detail, the Big data [5] in health care refers to the Electronic Health Records (EHR’s) which includes patients data, clinical data, doctor’s prescriptions, medical imaging, diagnostic reports, pharmacy information, past medication data, insurance related information and social media posts including tweets, blogs, status updates on Facebook, new feeds and articles in medical journal. The job of the big data scientist is to mine the big data and discover the associations, understand patterns and trends to improve health care, increase life expectancy and lower costs involved by proper diagnosing during the early stages of the disease.
Now, the digital health care solutions has transformed the health care system to become much more efficient, less expensive and achieve better quality than before with the introduction of Big data analytics in health care. To offer better services to the people, the health care system needs to evolve and innovate continuously. The following are the key points that make the performance of health centre better and people live healthier.

1. Measure, store and analyze the data to improve treatment quality
2. Manage revenue costs by reducing un-necessary tests
3. Improve preventive care and increase patient & physicians satisfaction

Adoption of big data in healthcare [7] significantly increases security and patients primary concerns and recommends the patient information to be stored in data centre’s with varying levels of security. As per the health insurance acts, EHR security must be taken as a high priority to ensure patient safety. As per the above discussions, the e-Health service data can be considered as a Big Data because of its variety of data with huge volumes flowing with high velocity. Some of the solutions to the e-Health service big data include the predominant current technologies [6] like HDFS, MapReduce, Hadoop, STORM etc.

At present, with the increase in number of hospitals and rural health care centres, the health care system seems to have improved. But, the recent Mc Kinskey report [19] states that the health care indicators continue to lag behind for India. Also, it is identified that, the health care budget allotments has not grown as that of GDP resulting in the shortage of infrastructure requirements. Also, it is observed that e-health care doctors, technicians, nurses and other administration staff are not in good pace with the kind of growth that the health care is exhibiting, the report says that, US spends 18% of its GDP on health care, where as India spends only 4% of GDP. D. Peter Augustine [1] has noticed the need for big data analytics in India’s health care system and concludes how to overcome the realistic challenges involving the government and its policies. The doctors, medical professionals and others using this technology suggest how to achieve better outcomes at lower costs.

This paper is organized as follows: in section II, we describe the existing health care in India, section III discusses about big data in general, section IV focuses on the problems in the existing healthcare system, section V provides the big ideas to provide better health care to the rural population of India, section VI presents the big data analytics in health care system and section VII concludes the work.

II. EXISTING HEALTHCARE IN RURAL INDIA

Rural India contains over 68% of India’s total population, and half of all residents of rural areas live below the poverty line, struggling for better and easy access to health care and services. Accessing primary health care to the population residing in rural India is still a challenge. Health issues being experienced by rural people are many and diverse – from severe malaria to uncontrolled diabetes, from a badly infected wound to cancer.

A structural outline of Rural Public Health Care system in India

Mr. Arnab Mukherji [3] has pointed out that, the main issue in the Indian health care system is inequity in Healthcare. To overcome the inequity in healthcare system, the Government of India has launched the National Rural Health Mission (NRHM) [4] in 2005. The aim of this mission is to provide effective healthcare to India's rural population. The thrust of this mission is on establishing a fully functional, community owned, decentralized health delivery system with inter-sectoral convergence at all levels, to ensure simultaneous action on a wide range of determinants of health such as water, sanitation, education, nutrition, social and gender equality. Institutional integration within the fragmented health sector was expected to provide a focus on outcomes, measured against Indian Public Health Standards for all health facilities. Though it is hard to believe, the fact is, compared to government health sector the private health sector has highly skilled doctors and the facilities are world-class using the latest laboratory equipment’s and innovative technology. Though, the aim of National Rural Health Mission (NRHM) is to provide effective health care to rural population, the hospitals located in remote villages are not equipped with well qualified doctors and sophisticated equipment due to several unknown reasons like the doctors may not be interested to reside in village to serve the rural population, lack of funds to procure and supply required infrastructure to all the remote hospitals. To access the high-end private sector facilities, people are spending large amounts of money on treatment which in-turn affects their livelihoods and slowly poor people are becoming poorer. In such cases, the concept of Telemedicine along with big data analytics is the better alternative to treat the patients in rural areas.

The Rural public health care system in India has three different levels of health care access. At the lowest level, we have Sub Centre (SC) and on top of that there will be a Primary Health Centre (PHC). This PHC is the referral unit for 6 sub centres and PHC contains 4 to 6 beds and Medical
officer in-charge along with 14 subordinate paramedical staff. After the Primary health centre, a 30 bedded hospital called Community Health Centre (CHC) which will be a referral unit for 4 PHC’s. To access primary health care services, the rural patient’s visits the Sub centre. The Health workers at the sub centre first take note of the vitals and symptoms of the patient, record them using the proposed system and if possible, they will try to resolve the issues themselves. If the case is too complex to handle, they send the corresponding criticality level and escalate it to the next level i.e. primary health centre, this level consists of a pool of medical officers connected to the network via our proposed method. Any medical officer available online at this level will try to resolve this issue by requesting for more information or by providing suitable recommendations. But, if a medical officer feels that the case cannot be resolved at this level, He / She suggests an immediate short time remedy, modifies criticality level if necessary and escalates it to the next level. i.e., the community health centre. At this level, a similar pool of specialist doctors available online will try to resolve the case with their medical expertise. If, however the case is too critical for them as well, they provide their immediate suggestions and escalate the case to the city hospital level. Who, being the final point of contact will resolve the issue of the patient. At any point in this process, if a doctor feels that a physical consultation is necessary, he/she can suggest the same through our system to the patients and diagnose the patient physically. Electronic Health Records (EHR’s) in the proposed system are sorted first by criticality and then by timestamp and also the proposed system should not allow two doctors at the same level to handle the case simultaneously.

III. BIG DATA IN GENERAL

**Definition:** Big Data is a collection of large and complex data sets which are difficult to process using common database management tools or traditional data processing applications. According to zdnet.com, "Big data refers to the tools, processes and procedures allowing an organization to create, manipulate and manage very large data sets and storage facilities".

Big data is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big data is arriving from multiple sources at an alarming velocity, volume and variety. To extract meaningful value from big data, you need optimal processing power, analytics capabilities and skills. Big data is changing the way people within organizations work together. It is creating a culture in which business and IT leaders must join forces to realize value from all data. Insights from big data can enable all employees to make better decisions—deepening customer engagement, optimizing operations, preventing threats and fraud, and capitalizing on new sources of revenue. But escalating demand for insights requires a fundamentally new approach to architecture, tools and practices.

The big data are generated from online transactions, emails, videos, audios, images, click streams, logs, posts, search queries, health records, social networking interactions, science data, sensors, mobile phones and their applications [8].

Big data is a platform for importing, storing and analysing data to uncover information not previously known. This explosion of the data, changing the way people think about everything. From the cutting edge scientific research to the monetization of social media and exchanging the way people think about healthcare analytics too. However, the health care has not kept pace with big data.

The large Indian health care system needs to harness healthcare’s “big data” and analyse a complex set of data, including electronic medical records and sensor data. This enables clinicians to access and analyse healthcare big data to ascertain quality, determine best practice, assess treatment strategies and identify patients at risk. The promises and potential of Big Data in transforming digital government services, governments, and the interaction between governments, citizens, and the business sector, are substantial. From “smart” government to transformational government, Big Data can foster collaboration; create real-time solutions to challenges in agriculture, health, transportation, and more; and usher in a new era of policy-and decision-making. Big Data raise a large number of information management issues, primarily in the areas of privacy, security, accuracy, and archiving, spanning major issues such as personally identifiable information, security of government data and information, and the accuracy of publicly available data. By fostering collaborations and economic development through private-public partnerships, government agencies appear to be tacitly endorsing the privacy, security, and other policies employed by those private sector entities.

The objective of this paper is to provide high quality health care without any discrimination on basis of gender, case, social status and economic status and also aims to ensure better health care to rural people and overcome the problems in the health care like expired drugs not administered to patients, fraud management in the health care. The proposed concept enables doctors, patients and staff to have role-based access to information on electronic health records.
**Characteristics of Big Data**

Big data generally refers to the social network data from the micro-blogging sites like Twitter, LinkedIn and social media platforms like Facebook. Traditional enterprise including transactional data, web store transactions etc. and machine generated / sensor data like call data records, smart meters, manufacturing sensors, trading systems, traffic data, air data etc. which keeps on increasing without the human intervention. Big data is not only driven by the exponential growth of data but also by changing user behaviour and globalization[10]. Globalization provides competition among the participants in the market. As a result, organizations are constantly looking for opportunities to increase their competitive advantage by using better analytical models.

The typical characteristics of the Big data are:

**Volume:** Data volume has been increasing exponentially: up to 2.5 Exabytes of data is already generated and stored every day. This is expected to double by 2015 [10]. The Big data generated vast amounts of data being by organizations or individuals from Terabytes to Exabytes and Zettabytes of data.

**Velocity:** Big data grows rapidly, which generated unprecedented quantities need to be stored, transmitted, and processed quickly [11]. Velocity is the speed at which the data is being generated like streamed data from various smart devices into social media and also camera streamed data which stores the data in motion from huge number of closed circuit cameras.

**Variety:** In Big data, the variety and heterogeneity of data sources and storage has increased, fuelled by the use of cloud, web and online computing [10]. Variety makes big data really big. Big data comes from a great variety of sources and generally has three types: structured, semi-structured and unstructured. Structured data inserts a data warehouse already tagged and easily sorted but unstructured data is random and difficult to analyse. Semi structured data does not conform to fixed fields but contains tags to separate data elements [8].

**Veracity:** Big Data Veracity refers to the biases, noise and abnormality in data. Is the data that is being stored, and mined meaningful to the problem being analyzed[12]. Veracity in data analysis is the biggest challenge when compared to other characteristics like volume and velocity.

**IV. PROBLEMS WITH THE EXISTING HEALTHCARE SYSTEM**

The Indian health system [18] includes public and private hospitals as well as specialized Ayurvedic hospitals. All major cities and medium-sized urban centers have private hospitals that provide an excellent standard of care. Health insurance only covers hospitalization and emergency costs. Other care must be paid for upfront, but even privately it is extremely reasonable compared to other countries, so medical costs should not be a significant expense.

India is the second most populous country in the world, the healthcare infrastructure that is over-burdened with this ever increasing population, a set of the following challenges that are unique to India arise.

- India faces the twin epidemic of emerging infectious diseases as well as chronic degenerative diseases.
- The reason for the infectious diseases is the poor implementation of the public health programs, and the latter to demographic transition with increase in life expectancy.
- Economic deprivation in a large segment of population results in poor access to health care.
- Poor educational status in India, leads to non-utilization of primary health services and results in the increase of risk factors.
- Lack of education, gender inequality and explosive growth of population contribute to increasing burden of disease resulting in the challenge to the health sector.
- The most important from the government side is, Expenditure on health by the Government continues to be low. It is not viewed as an investment but rather as a dead loss.
- Some of the States under financial constraints will cut expenditure on health results in the poor health care.
- Growth in national income by itself is not enough, if the benefits do not manifest themselves in the form of more food, better access to health and education”.

The various types of data can be anticipated from the healthcare system from different health science data sources include data from drug research, social media, patient records, gene sequencing, test results, claims, home monitoring mobile apps etc.,

- Clinical data including unstructured documents, prescriptions and images.
- Day to day research publications and medical references
- Huge amounts of genomic data for analyzing the behaviour of various
- Web and social networking data on healthcare issues
- Streamed data from home monitoring, Tele-Medicine, hand held and sensor based wireless device data.
• Other organization data including billing, scheduling, inventory of medicine
• Health claim data from the payers towards the charges laid on the patient

As the health care data growing drastically day by day, the infrastructural facilities must be enhanced in the health care center’s to meet the needs of the rural people and necessary steps must be taken to store the vast amount data from the above said different sources for the future analysis. Thus the need for BDA (Big Data Analytics) arises, which provides clinical decision support through large amounts of data, personalized care by early detection and diagnosis before a patient develops disease symptoms, clinical operations with great accuracy, fraud management in the health sector.

V. BIG IDEAS TO FIX RURAL HEALTHCARE IN INDIA

Now, we propose the following seven big ideas from [9] to fix rural health care in India and bridge the gap between quality and affordability in government hospitals. These ideas will enable us to access the services on par with the private super speciality hospitals. Further, the implementation of these ideas will provide cheaper, better and easier health care facilities to the citizens of India.

1. e-Health File: The creation of a e-Health care file for each patient, where all health care providers and patients themselves were able to submit information (with the consent of the patient). Both subjective data, symptom diaries, lab data, image diagnostics, pathology reports etc., could be filed. To overcome the information overload from the massive amounts of data, Big Data Analytics could be employed for the processing of the data and obtain the desired results with great accuracy in reasonable time.

2. Creating awareness with chronic diseases: The system must identify and create awareness among the people with the common chronic diseases at particular areas, through which we can prevent diseases. These chronic diseases are responsible for the 75% of health care spending due to lack of awareness and prior care.

3. e-Prescribe: Paper based prescriptions are archaic and lead to several miseries each year due to errors in prescription. But if every doctor is provided with an electronic prescription system, it would improve safety by making prescriptions easier to read and providing instant checks on drug interactions, dosages, and a patient’s medication history.

4. Electronic Medical Records: Medical Experts agree that electronic medical records (EMRs) are a must for the better health care in India. But, at present only few hospital are maintaining EMR’s, mainly because of cost, privacy issues, and the lack of one compatible, easy-to-use infrastructure.

5. Stop Unnecessary Treatments: Doctors should avoid trial and error type of medication. The problem must be examined thoroughly by performing the required diagnostic tests during the preliminary days of disease. The right treatment should be suggested at the first visit only which avoids the disease to become more critical. Most of the issues are arising with the misdiagnosis and wrong treatment during the early stages.

6. Reduce Infant Mortality: The Infant Mortality rate in our country is substantially large compared to other countries. Though, the government has several schemes for the pregnant women, those were not yielding better results due to the lack of proper medical care. If the proper care is taken towards the pregnant women, definitely the new-born baby will be healthy in all aspects to make India healthy.

7. Tele Medicine: Doctors can often diagnose or prescribe without seeing the patient. The patient has to physically appear before the nearby health center, where the nurses or health workers will diagnose at first level, note the symptoms and informs the high level specialist doctors about the case. After examining the reports, the specialist doctor suggests the treatment through health worker which reduces costs and creates satisfaction by virtual communication of patients and doctors to discuss medication changes and test results through an online system.

Many people in the health care business are cognizant off that no-patient remains untreated irrespective of his geographical location. Also, no matter, where the patient is, the patients must be accessible to

- Organized primary care, network of medical centers, network connecting hospitals and doctors, round the clock with emergency services.
- Network of secondary care hospitals in small towns.
- Critical care must be available in emergencies like Gynecology, Pediatrician, General Medicine and General Surgery.
VI. APPLYING BIG DATA ANALYTICS IN HEALTH CARE

We live in the age of big data. The amount of data created in the world up to and including 2005 is now created every two days. Big data is a platform for importing, storing and analyzing data to uncover information not previously known. This explosion of the data changing the way people think about everything. From the cutting edge scientific research to the monetization of social media and exchanging the way people think about healthcare analytics too. However, the health care has not kept pace with big data. Big Data Healthcare is the drive to capitalise on growing patient and health system data availability to generate healthcare innovation. By making smart use of the ever-increasing amount of data available, we find new insights by re-examining the data or combining it with other information. In healthcare this means not just mining patient records, medical images, diagnostic reports etc., for insights, diagnoses and decision support device, but also continuous analysis of the data streams produced for and by every patient in a hospital, at home and even while on the move via mobile devices [14].

Even today the majority of health care analytics is performed by doing monthly data refreshes in relational databases that produce pre-processed reports. A fair gap is often missing lab test is often 45 days old, as the data flow move from batched data fields to real time fields from transactional systems and streaming data from analytical modelling devices. This old model of analytics will fail. Analysis will need to be done on that spot moment not in the pre-processed form. Data refreshes need to be done in real-time not once in a month. The data analysis tools of today are likely yellow pages phone book in the era of Internet Search Engine. They are becoming more obsolete with each passing day. The traditional health care analytic tools are built on tools developed by IBM in 1970, more than 40 years ago.

If all the three parties (payer, provider, pharmaceutical company) [13] work collaboratively and share data/insight, disease management programs will become cost-effective and deliver improved patient outcomes at a scale that will further optimize overall health care cost structure.

The term “e-health” [15] defined by WHO: “ a new term used to describe the combined use of electronic communication and information technology in the health sector”. e-health is the main driver for three significant changes within the health care environment[15]:
1. Patients to become better informed
2. Patients to become more active and empowered in their health care
3. Healthcare to become more efficient.

It is referred in the Cognizant 20-20 insights by Cognizant, [16] Big data solutions attempt to cost-effectively solve the challenges of large and fast-growing data volumes realize its potential analytical value. For instance, trend analytics allow you to figure out what happened, while root cause and predictive analytics enable understanding of why it happened and what it is likely to happen in future. All healthcare constituents – patients, payers, providers, groups, researchers, governments etc. – will be impacted by big data, which can predict how these players are likely to behave, encourage desirable behaviour. These applications of big data can be tested, refined and optimized quickly and inexpensively and will radically change healthcare delivery and research. The healthcare domain has been an easy target for people who seek easy money by using fraud methods [17]. Healthcare fraud is expected to continue to rise as people live longer. The white paper by trend analytics [17] reveals that healthcare fraud prevention has resulted savings of nearly $4.1 billion in 2011. A big data platform has ability to sift through a huge amount of historical data in relatively shorter amount of time, so that the business transactions can use fraud detection on real time. Though, the big data analytics in healthcare plays a crucial role to provide better health care services, provide analysis on the historical data to uncover hidden information, the big data analytics has the challenges like Heterogeneity and Incompleteness of data, scale, timeliness, privacy and Human Collaboration[17]. The future research is all about to overcome the challenges and use big data analytics in healthcare to uncover the knowledge from the raw unstructured data.

VII. CONCLUSIONS

Big data analytics in healthcare is evolving into a promising field for providing insight from very large data sets and improving outcomes while reducing costs. In this paper, an overview of the issues faced by the rural population residing in remote parts of the country in accessing the primary health care is addressed in detail. Also, discussed how the big data analytics are beneficial to transform rural healthcare by gaining insight from their clinical and other data repositories and make informed decisions.

This paper discusses about the big data and its characteristics, methods and challenges and suggests how to overcome the underlying problems being faced by the health care industry. Also presents the big ideas to fix the healthcare system in India. The implementation part of this paper can be done using HDFS (Hadoop File System) for the huge data storage and Hadoop Map Reduce with Amazon Web Services. The use of big data analytics across the healthcare organization and healthcare industry will mine the doctor’s lab transcript’s using text mining and correlation to patient outcomes and location aware application analytics for enhancing customer experience. Achieving better outcomes at lower costs has become very important for health care which can be achieved through the implementation of this paper using Hadoop HDFS and MapReduce to uncover the information lying in big health data sets.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my profound gratitude and deep regard to my guide, Prof. R. Manjula, SCSE, VIT University, for her exemplary guidance, valuable feedback and constant encouragement in completing this paper. Her valuable suggestions were of
immense help in getting this work done. Her perceptive criticism kept me working to make this paper in a much better way. Working under her, was an extremely knowledgeable experience. Also, I would like to extend my sincere gratitude to my parents, my wife and my lovable daughter for their constant support and encouragement in completing this paper.

REFERENCES


www.iwcsit.com  7178