

# **Implementation of Big Data in Health Information Systems: Sample Approaches in Saudi Hospital**

**G. Rasitha Banu, PhD**  
Department of Health Information  
Management and Technology,  
Jazan University, KSA

**Prakash Kuppaswamy, PhD**  
Department of Computer  
Engineering & Networks,  
Jazan University, KSA

**N. Sasikala, PhD**  
Department of Computer Science  
Mohamed.sathak college of Arts  
and Science, India

## **ABSTRACT**

Big data concept provides opportunity to exchange patient's medical information to the different healthcare providers. Health Information System (HIS) has created the ability to electronically store, maintain and move data across the world in a matter of seconds and has the potential to provide healthcare with tremendous increasing productivity and quality of services. Big data analytics is a growth area with the potential to provide useful insight in health information system. Big Data can unify all patient related data to get more option to view patient records to analyze and predict early disease detection. Big data supports and improve clinical practices, new drug development and health care financing process. Implementation of Health Information system (HIS) continues to expand infrastructure in Medical field due to enormous number of patient comes across to store medical data. In this paper we focus the Big data concept to increase and store patients details in Saudi public hospitals with maximum utilization. Most of the Saudi public and private hospitals Health information system locally connected and maintained by own hospital admin. There is no system implemented to share the patient health record, treatment details and medical prescription data to other hospital. The main problem in the Saudi hospital, Health information is not centralized due to unstructured, semi structured data maintain by the Saudi hospital. Proper Health information system is able to offer correct and complete personal health and medical summary through the Big data methods. This paper introduces the Big Data concept and characteristics, health care data and some major issues of Big Data. Big Data methods and challenges in medical applications and health information system are also discussed in this study. This study provides a base model to increase the use of big data in health information system and can assist to understand the breadth of big data applications.

## **Keywords**

Big Data, Health information system, Medical record, Diagnosis, Centralized record, Hadoop, Saudi hospitals

## **1. INTRODUCTION**

Big data refers to huge volume of data exists in different file format such as structured, unstructured and streamed data which is placed in a server to mine the useful information for business profits. Different types of analysis can be taken to get different results. Big data is characterized by 7V's Such as Volume, velocity, Variety, Veracity, Value, visualization and Volatility. Big data contains enormous volume of data [3]. Velocity refers to the speed of dataflow from or to sources like network, social media sites and mobile devices and so on. Variety means data in various file formats. Big data volatility means the validity of data and how long the data will be stored. Veracity means accuracy and correctness of

information. Value refers to the quality of data. Normally data from EMR's and EHR's are recognized as high value data which can lead to good quality. Visualization means charts and graphs which are used to visualize large amounts of complex data [1][3].

The Big data is needed to increase the storage capacity, to increase the processing speed and availability of data. Various tools are used in big data such as NOSQL, Hadoop Map Reduce, and EC2 server and so on. The Big data is used in some application areas such as telecommunication, healthcare, social network and so on. Now a day's Big data makes changes dramatically in health care. While Big data is applied to health care which reduces the cost for treatment, prevent the people from disease through predictions and life span of human life to be improved. One of the biggest issue in healthcare is how medical data is spread across many sources governed by different states, hospitals and administrative departments. Integration of these data will need new infrastructure where all data providers collaborate with each other.

Health information systems refer to any system that captures, stores, manages or transmits information related to the health of individuals or the activities of organizations that work within the health sector in a fraction of seconds which is used to increase productivity and quality of services. Health data sets are too huge and complex .It is very difficult to use traditional software to manage health data. Big data is playing a vital role in health care in terms of storing huge volume of data, different file formats and accessing data in high speed. EHR is the most widespread application of Big Data in healthcare. Electronic health record (EHR), or Electronic medical record (EMR), refers to collect patient information electronically in a digital format [2]. Every patient has his own digital record which includes demographics, medical history, allergies, laboratory test results etc. Records are shared via secure information systems and are available for healthcare providers from both public and private sector. Every record is comprised of one modifiable file, which means that doctors can implement changes over time with no paperwork and no danger of data replication [3]. EHRs can also trigger warnings and reminders when a patient should get a new lab test or track prescriptions to see if a patient has been following doctors' orders. It stores data accurately and maintain up to date information.

The electronic health record (EHR) is considered as "big data. In worldwide, there is an increase in Electronic health record adoption rates [4][5]. Every year, one billion patient visits documented in EHR systems in Saudi hospital. In addition to this data about medical conditions, medications, and treatment approaches are also increased. Thus, Health Information system is needed to organize, interpret, and recognize patterns from these data are needed [6]. The EHR adoption for

healthcare improves quality of patient care and reduces the health care cost. Previous studies shows that EMR systems saves \$77 billion per year at the 90% level of adoption; if we added value for safety and health, it will double these savings [2][3].

In Saudi Arabia, One issue in EMR systems is that they are not highly centralized; each Healthcare Provider (HP) has its own local EMR system. Cloud computing paradigm is one of the popular Health Information Technology infrastructures for facilitating Electronic Health Record (EHR) sharing and EHR integration. Healthcare clouds offer new possibilities, such as easy and ubiquitous access to medical data, and opportunities for new business models. In this paper we are discussing about some approaches to utilize big data in Saudi hospitals.

## **2. LITERATURE REVIEW**

**Priyanka K Prof NagarathnaKulennavar (2014)** This paper gives a brief introduction about how we can uncover additional value from health information used in health care centers using a new information management approach called as big data analytics. This paper defines big data analytics and its characteristics, comments on its advantages and challenges in health care. Big data analytics has the potential to transform the way healthcare providers use sophisticated technologies to gain insight from their clinical and other data repositories and make informed decisions. To that end, the several challenges must be addressed. As big data analytics becomes more mainstream, issues such as guaranteeing privacy, safeguarding security, establishing standards and governance, and continually improving the tools and technologies will garner attention. Big data analytics and applications in healthcare are at a nascent stage of development, but rapid advances in platforms and tools can accelerate their maturing process [7].

**Suzhi Bi, Rui Zhang, Zhi Ding, Shuguang Cui (2015)** in this article, authors discuss the challenges and opportunities in the design of scalable wireless systems to embrace such a “big data” era. On one hand, we review the state-of-the-art networking architectures and signal processing techniques adaptable for managing the big data traffic in wireless networks. On the other hand, instead of viewing mobile big data as a unwanted burden, they introduce methods to capitalize from the vast data traffic, for building a big data-aware wireless network with better wireless service quality and new mobile applications. This article addresses challenges and opportunities that we face in the era of wireless big data. They outlined the major obstacles of big data signal processing and network design with respect to the scale of problem size and the complex problem structures. Nevertheless, research on big data for wireless communications and networking is not only promising but also inevitable in light of the continuing data volume explosion [9].

**Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, and Guang-Zhong Yang (2015)**, This paper provides outlines the key characteristics of big data and how medical and health informatics, translational bio informatics ,sensor informatics. This paper discusses some of the existing activities and future opportunities related to big data for health, outlining some of the key underlying issues that need to be tackled. A better use of medical resources by means of personalization can lead to well-managed health services that can overcome the challenges of a rapidly increasing and aging population. Thus, advances in big data processing for health informatics, bioinformatics, sensing, and

imaging will have a great impact on future clinical research. Another important factor to consider is rapid and seamless health data acquisition, which will contribute to the success of big data in medicine. Specifically, sensing provides a solid set of solutions to fill this gap [10].

**Lidong Wang, Cheryl Ann Alexander (2015)** authors introduces the Big Data concept and characteristics, health care data and some major issues of Big Data. These issues include Big Data benefits, its applications and opportunities in medical areas and health care. Methods and technology progress about Big Data are presented in this study. Big Data challenges in medical applications and health care are also discussed. Big Data is based on data obtained from the whole process of diagnosis and treatment of each case. Big Data has challenges in medical applications and healthcare. The authors of the paper will focus on Big Data in medical sensor data and streaming data processing, privacy-preserving data mining in healthcare, sentiment analysis of medical big data and personalization and behavioral modeling [11].

**Jasleen Kaur Bains(2016)**This paper gives a wide insight and know how about the various Big Data analytics (BDA) initiatives taken to improve healthcare worldwide. It also explains the various phases involved in BDA process and depicts its benefits and challenges with focus on healthcare industry. As has been seen in existing studies, the BDA has shown remarkable outcomes in many healthcare organizations. In the future, with even more advancements in the BDA processes we expect that healthcare cost will come down drastically, life expectancy will increase, and we will see much healthier population as compared to now with people taking more accountability and charge of their health using technological advancements. The future of healthcare is promising [3].

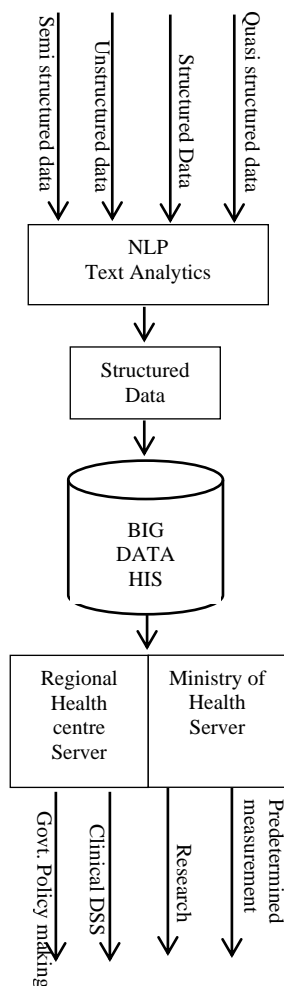
## **3. PROBLEM STATEMENT**

In Saudi public sector, storing patient health information process uses data from many sources. Most of the data is unstructured processed data such as biological samples, medical images, patient claims, medical prescriptions, clinical notes, status updates, comments and diet advices etc., As we stated above almost all aspects of healthcare data including public health record, Electronic health data delivery and research become more dependent on efficient data storage (i.e.) Big data storage significantly required. We need to generate right metadata for this data and transform it into a structured format. The image and video data should be structured for semantic content and search. Practically, patients have a habit of hide some of the personal facts and their lifestyle while filling up data sheet and during consultation with physicians. Digital data need more accuracy of the patients information otherwise system will not predict the cause of disease and it will lead to wrong treatment. We need to ensure that data from sources is a valid data or not and is of good quality. Determining the validation and quality of data in case of patient’s relative, third party or other media is more risk. Existing data storage system in Saudi hospital is not integrating patient’s activity and his treatment from health center and other private hospital. It can be very hard by developing certain standard database design practices meant for a specific domain like private hospital, health centre and other clinical activities. Most of the patients health record handled by non-technical person, they don’t have practice to normalized data according to the data storage format. For instance, big data holds the promise of advanced analytical methods that can help medical researchers and drug manufacturers connect large collection of genomic and

clinical sample data with streaming data from the Web and government censuses in order to understand better how inherited genetic variants contribute to certain genetic diseases or predispositions to diseases, and perfectly implement drug assessments.

#### 4. PROPOSED METHODOLOGY

Nowadays there is a enormous growth of data due to the accumulation of unstructured text data (i.e) up to 80% of medical related data is unstructured text data. Natural Language Processing (NLP) is the scientific discipline used for making natural language accessible to machines. It is necessary to facilitate text analytics by establishing structure in unstructured text for further analysis. Text analytics is a process of extracting useful information from text sources. Text Analytics tools provides enormous services in healthcare sector for constructing structured data from unstructured data. In healthcare, it is used for medical record content extraction, drug interaction discovery from pubmed articles, Disease outbreak monitoring, and control from social media data.



**Fig 1. Structure of the big data in Health care**

In the era of big data, the right platform enables businesses to fully utilize their data lake and take advantage of the latest parallel text analytics and NLP algorithms. In the above diagram, all types of data such as electronic health record (EHR), patient monitoring systems, laboratory systems, imaging systems and operational support systems in the form of semi structures data, unstructured data, structured data and

quasi structured data from saudi government hospitals and private hospitals are given as the input of NLP text analytics. Our proposed model will facilitate the integration of unstructured text data with structured data. BIG data analytical method is applied to structured data. It helps data specialists to find, compile, manage and analyze large volumes of structured. Our new proposed structure make it easier for medical data specialists to combine different kinds of data from many different sources, process high volume of data very quickly and accurately and get different types of data technology to work smoothly together such as image processing and signal processing. The useful information such as patient information is stored in centralized servers such as regional health center server and ministry of health server for supporting research, government policy making and clinical decision support systems and predetermined measurements. The big data allows healthcare professional to access the centralized server from anywhere.

#### 5. BENEFITS

For public and private sector, big data could mean the key to a new era of data analytical and stranded data service distribution. Storing the patient medical information in one centralized source using big data server has good advantage instead of storing the medical information in health centre, private hospital and Saudi government hospital in different individual server. Big data helps to storing the patient medical information in one centralized server the medical practitioner can quickly access and share medical information about a patient across the various departments and organizations. Big data computing is a highly simple to use technology to add to medical organization. Big data based storage system of medical records is much better, faster and easier to access, as well as boasting lower downtime percentages. Big data provide the facility to access the medical information from the centralized data server from anywhere.

Depend on big data tools, healthcare experts may be able in the future to combine epidemiologic approaches with disease and mortality statistics that they have accumulated over the years in order to gain a better understanding of disease propagation patterns and reassess their disaster recovery plan activities. Furthermore, by applying predictive analytics and simulation to healthcare data, healthcare experts may gain insights into or predict the demographic distribution of certain diseases with regards to ethnicity, gender, and geography, and be able to accurately quantify the interplay between the quality of healthcare services accessible in different geographic areas and the Saudi government's investment in health care. Sharing of medical record on big data is a less cost accessing the data. In big data there is no need to upgrade separate system. Guidelines, backup systems and disaster recovery can be managed in centralized server.

#### 6. CONCLUSION

Big Data is based on large volume of data obtained from the whole process of diagnosis, treatment and extract large quantities of structured and unstructured data. It is necessary to analysis of Saudi citizen's health information as the capability of analysts to recent advancements in analytics and high-performance information technology available and provided by the Saudi government in public health sector to keep the every individual citizen health information. Big Data analytics can perform predictive modeling to determine which patients are most likely to benefit from a care management plan. Proposed method offers a lot of benefits such as disease prevention, reduced medical errors and the right care at the right time and better medical outcomes. This proposed model

is complete collection of data which can improve the Research and Development and translation of new therapies. Normally, Big data facing lot of challenges in medical applications and health information system. Especially, these challenges include consolidating and processing segmented, aggregating and analyzing unstructured data, indexing and processing continuously streaming data, data leakage and unified standards etc., above these problems our proposed methodology solving by NLP and various Text analytical method mentioned in the figure1. Proposed scheme has great potential to improve medicine, guide clinicians in delivering value-based care. At the end, proposed structure addressed several challenges and effective way of normalized data form in the medical environment. In the future proposed plan will see the rapid, widespread implementation and use of big data analytics across the healthcare organization and the healthcare industry using various analytical algorithms.

## 7. REFERENCES

- [1] Jasleen Kaur Bains , “Big Data Analytics in Healthcare- Its Benefits, Phases and Challenges”, International Journal of Advanced Research in Computer Science and Software Engineering Research Paper Available online at: [www.ijarcsse.com](http://www.ijarcsse.com) , Volume 6, Issue 4, April 2016.
- [2] Ahmed E. Youssef, “A Framework for secure Healthcare systems based on Big data analytics in mobile cloud computing environments”, International Journal of Ambient Systems and Applications (IJASA) Vol.2, No.2, June 2014.
- [3] R. Zhang and L. Liu, “Security Models and Requirements for Healthcare Application Clouds”, IEEE3rd International Conference on Cloud Computing, 2010.
- [4] Charles DK J, Patel V, Furukawa M., “Adoption of Electronic Health Record Systems among U.S. Non-federal Acute Care Hospitals: 2008- 2013”, <http://www.healthit.gov/sites/default/files/>. Accessibility verified April 20, 2014.
- [5] Faxvaag A, Johansen TS, Heimly V, Melby L, Grimsmo A., “Healthcare professionals’ experiences with EHR-system access control mechanisms”, Stud Health Technol Inform 2011;169:601–5. [PubMed]
- [6] Wagholikar KB, Sundararajan V, Deshpande AW., “Modeling paradigms for medical diagnostic decision support: a survey and future directions”, J Med Syst 2012;36(5):3029–49. [PubMed]
- [7] Priyanka K, Prof NagarathnaKulennavar, “A Survey On Big Data Analytics In Health Care”, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5865-5868
- [8] Suzhi Bi, Rui Zhang, Zhi Ding, and Shuguang Cui, “Wireless Communications in the Era of Big Data”, arXiv:1508.06369v1 [cs.NI] 26 Aug 2015.
- [9] Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, Guang-Zhong Yang, “Big Data for Health”, IEEE Journal of biomedical and health informatics, Vol.19 No.4, July 2015.
- [10] Lidong Wang, Cheryl Ann Alexander, “Big Data in Medical Applications and Health Care”, American Medical Journal, 6 (1): 1.8, 2015.
- [11] Liyanage H, Liaw ST, de Lusignan S., “Accelerating the development of an information ecosystem in health care, by stimulating the growth of safe intermediate processing of health information (IPHI). Inform Prim Care 2012;20(2):81–6. [PubMed]

## 8. AUTHOR PROFILE

**Dr.G.Rasitha Banu**, Assistant Professor, Department of health Information Management and Technology in Jazan University, KSA. She is having 19 years of teaching experience and 10 years of research experience. She has published more than 20 papers in national and International Research journals. She has presented many Technical papers in national and International conferences. Her research area includes Data Mining, Bio-informatics and Cloud computing etc.

**Dr PrakashKuppuswamy**, Lecturer, Computer Engineering & Networks Department in Jazan University, KSA. Scholar from Dravidian University, India. He has published 25 International Research journals/Technical papers and Participated in many international Conferences in Maldives, Libya and Ethiopia. His research area includes Cryptography, Bio-informatics and E-commerce security, Cloud Security etc.

**Dr.N.sasikala**, Assistant Professor, Department of Computer Science, Mohamed Sathak College of Arts and Science, Chennai, India. She is having 17 years of teaching experience and 10 years of research experience. She has published more than 10 papers in national and International Research journals. She has presented many Technical papers in national and International conferences. Her research area includes Software Engineering, Bio-informatics and Big data etc.