

# Big Data in Healthcare

Mohammad Kamal Hussain, Mohammad Jabed Hussain, Mohammed Bakri Osman, Taher  
Mohammed Abdurraheem, Mahmoud Al-Areefi

**Abstract:** To improve quality in healthcare, it is very much important to store, manage and retrieve as well as use the data & information properly as it has great potential to help leaders in effective decision making. Managing data in healthcare is not easy task as it has many associated risk. As per new trends in healthcare industry, it is observed that the data volume generation in healthcare is growing rapidly. Big Data is substantial, less organized and mixed in nature. In addition to that, big data is considered one of the best tools to reduce the associated and functional cost of healthcare providers worldwide. While income should not be only a main or prime indicator, it is equally important for healthcare providers to gather the most valuable present tools and techniques and setup to force or inculcate big data effectively otherwise it can harm or risk organization to lose money in business as well as profit. **Objective:** This paper is focusing on the special factors of Big Data in healthcare. The main aim of this study is to find the roles of big data in healthcare and know how big data is helping in data transaction in healthcare industry. **Methods:** More than 30(n=30) published papers have been reviewed and suitable papers (n=18) have been included to make the conclusion. Information was condensed utilizing distinct measurable assessment & techniques. **Findings:** According to investigation of published articles, it has established that the role of big data is very much unique and important as well as it is helping healthcare providers to improve the patient safety and quality by providing smooth health information storage and exchange with high privacy and security. **Conclusion:** Big Data in healthcare is a new concept introduced in healthcare data analytics and management which is basically focusing in improving the drug and disease discovery, personal healthcare record, electronic health record, effective decision in diagnosis and treatment by healthcare practitioners and at most helps in getting desired and positive health outcome. The data is one of the crucial factors in healthcare and it is high time for healthcare providers to look into those matters in enormous way.

**Keywords:** Big Data in Healthcare, Cloud Computing, Data Analytics, Health Analytics, Health Meta Data. .

## I. INTRODUCTION

Big Data is well-defined as bulk or huge volume of stored data that is useful for every industry for proper trending and

**Revised Manuscript Received on March 10, 2020.**

\* Correspondence Author

**Dr. Mohammad Kamal Hussain\***, Associate Professor, Department of Health Information Technology & Management, Umm Al-Qura University, Makkah, Saudi Arabia. E-mail: [drkamalhussain@gmail.com](mailto:drkamalhussain@gmail.com)

**Mohammad Jabed Hussain**, Department of Information System, Umm Al Qura University, Makkah, Saudi Arabia.

**Mohammed Bakri**, Assistant Professor, Department of Epidemiology, Umm Al- Qura University Makkah, Saudi Arabia.

**Taher Mohammed Abdurraheem**, Assistant Professor, Department of Health Information Technology & Management, Umm Al- Qura University Makkah, Saudi Arabia

**Mahmoud Al-Areefi**, Assistant Professor, Department of Health Information Technology & Management, Umm Al- Qura University Makkah, Saudi Arabia

use it for healthy decision making. The big factor now days in data analytics is massive high speed in collecting, managing and storing as well as retrieving the data as previous data analytics were depending upon analyzing the earlier data. Big data is having concept of real time current generated data. A healthcare facility is the place where generation of data is from every corner and in huge numbers. Managing or handling, Storage and retrieval of data on fingertip is required by all healthcare providers and that is possible only due to the better amalgamation uses of computer and information technology. It is considered that the data management in healthcare, especially for patient records and other associated data are essential to provide better quality of care in proper way. Data must be useful in a way that may be useful for taking proper and better decision for quality improvement. Big data can be analysed, and it may lead to proper decision making for the betterment of all stake holders of healthcare industry, patients, providers and government as well to prepare plan for disease management.

## II. HISTORICAL BACKGROUND AND CURRENT SCENARIO

The notion of Big Data is relatively fresh in health industry, therefore the essence of generating, organizing, managing and storing the large number of information for ultimate analysis and uses is quite old. As the concept of big data, it is that degree of data that cannot be gathered, stored and processed by a single device. Before one decade, the term Big Data was almost not known to the industry but in recent days it is widely advertised as the new revolution in health data management and analytics. As cloud computing before it, the term has been adopted by all stakeholders from vendors to large scale outsourcing and cloud computing service providers are taken initiatives to encourage their support. The concept gets known to other during 2001 when famous industry analyst Doug Laney published his concept regarding Meta data to big data in three Vs. (Volume, Velocity & Variety, and Figure.1.) [1], V= Volume which basically referring to the point that Big Data includes analysis reasonably big number of data and information collected from the various sources. In the past, storing the large number of information were having problem but new technologies like Hadoop[2] have eased and simplify the burden. V= Velocity which reflects the high speed at which the data is produced and converted in very timely manner, it means data is available to the user with high speed streaming in an efficient way. V=Variety which defines the Big Data can come from various sources from structured to semi-structured to non-structured text documents,

## Big Data in Healthcare

Audio, video, images, geographical information, messages like email and other useful data. Later Gartner and IBM have given another V that is Veracity which refers to vagueness of data like the amount of quality of data being stored. Data like columns on community networking websites are vague [3-4].

Big data has a lot of scope to manage the bigger size of the data that is being created and stored on global level is almost unimaginable, and it keeps on rising.

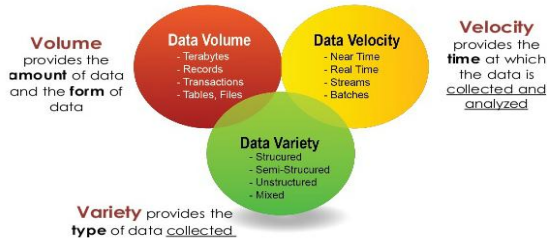


Figure.1. 3Vs

**2.1. Big Data Privacy:** Privacy is always a major issues and concern in health information management, With the growing acceptance of big data in healthcare and the ease of access and retrieve to huge numbers of stored records and repositories, personal data privacy in healthcare is more complex and not assured even healthcare providers put efforts in signifying its available data. To choose the health big data service provider from markets, it is necessary to know what parameter they use to secure the confidential data and its uses to the restricted person. The provider can use the methodology to combined data, information that need action, Intuitive knowledge and Real time wisdom to protect and misuse of health records. Big data analytics with the use of refined machineries has the probability to transform the data warehouses and make informed decisions. In this factors the major part which is privacy and security as well as standards and governance of the same need to be discussed and addressed [11].

**2.2. Big Data in Healthcare:** The uses of healthcare big data for health service providers and policy makers is more appropriate as the health service providers in industry generates tons of data daily which needs to be stored properly and used efficiently to increase the patient satisfaction and avoid the mistakes at all the levels of healthcare facilities including primary, secondary, tertiary care government or private hospitals or providers facilities. The present inclination in Health Data analytics in health systems or information management is toward developing a refined framework, tools and techniques for functions as well as operations of healthcare business as well as clinical intelligence.

Big data in health systems and health informatics is more valuable and it is accepted to as well as predict the outcome of disease, outbreaks, better health care and quality of life (health for all) and prevent morbidity and mortality as well as disease development in the society and help the policy makers and government to decrease the cost of health treatment [5, 7]. Big data is generally give information related to diseases and admonitory indications and sign for proper diagnosis and treatment that to be properly managed and followed at healthcare facilities. [5-6]. Big data is useful in clinical

decision support as well as detection of the epidemic, current diseases pattern, health status of the community and population at large in the country, and also support in research related to health service as it is providing large volumes of health data. Public health organizations or facilities (like PHCs, CHCs) NGOs, Hospitals, Poly-clinic and or medicine company can use this data to articulate their business strategy, operational and governing policies, procedure to be followed for better interventions as well as medical treatment for the patients and drug development for the society at large. Big data has great potential and implications on healthcare providers, user's researchers and professionals. [8].



Figure.2. Big Data Applications

**2.3. Optimizing workflows in Healthcare:** The product industry has many processes which are predictable in nature when it comes to output. Product industry also follow the same input process and output concept but when it is compare with service industry like healthcare which is dynamic and also dependent on many staffs members from medical and supportive staffs department and patients as well. This unpredictable workflow adaptation to improve high productivity is challenging unless healthcare providers will have to the required to integrate multiple data warehouse such as real time data access e health, electronic medical record, health information system, laboratory information system, radiology information system and automated machine enabled to operate and installed. Big data helps in optimizing the workflow in healthcare by providing huge number of useful data together as repositories and provide real time access as and when required and also increases the high quality of work productivity. **Quality Care:** To provide quality care to the patients, proper uses of data in healthcare is essential.

**2.4. Evidence-Based Medicine:** All provided treatment must be evidence based. Big data helps healthcare provider in Merging and analyzing the range of organized and unorganized data which is can be electronic medical records or Health Information,

business and operative data, medical data, genomic data to compare treatments with outcomes and forecast patients at risk for disease or readmission to provide more well-organized care that may lead to more satisfactions of the patients. It may lead to the growing in business of healthcare providers.

**2.5. Medical and Clinical Operations:** Big data helps healthcare provider in provide better medical and clinical care. It also helps to investigate and determine the cost effective scientific way to improve the efficiency and treat accordingly as per established standards.

**2.6. Global & Public Health:** Managing data especially for public and global health issues is now days very challenging. New trends in disease are more evident and to provide proper care and data management, it is very much important to increase the data storage globally. Big data helps healthcare providers worldwide in analyzing disease patterns and tracking disease outbreaks and transmission to improve public health communicable and non-communicable diseases surveillance and speed response. Big data increases faster development of more accurately targeted vaccines like choosing the annual influenza stresses and also turning large amounts of data into actionable information that can be used to identify needs and provide services as well as forecast and prevent crises especially for the benefit of community at large [12].

**2.7. Population Health:** Public health policy is focused on a thorough analysis of the health status of a regionally heterogeneous population and the Social and cultural-Economic with a view to identifying and reflecting on social efforts to improve the health outcomes. Big data research can direct policies through targeting measures that target a certain segment of the population. The policy's success depends crucially on the quality of the underlying research and the consistency (effectiveness) of the interventions. Universally accepted criteria for validating progress are still missing for many interventions (for example in the social / mental health domain). **There are many issues with respect to big data and public health, such as:**

- Data privacy laws make it challenging to integrate data analyzes from various healthcare professionals and programs.
- Unorganized text is a large portion of population health information.
- There are limits on interoperability, data consistency and system integration.
- Current systems are not complex.
- A huge-scale, comprehensive and privacy of measuring and collecting information, together with the diligent testing of quantitative methods for managing missing information, would allow the reinforcement of the scientific basis for decision making and the creation of more detailed and efficient (heterogeneous / personalized) steps.

**2.8. Infectious Diseases:** In recent years, technology makes it feasible not only to obtain data from the health-care system (medical facilities, wellbeing centers, labs, etc.) but also to obtain information from community themselves (sensors, tracking, IoT apps, social networks, etc.). Health

environments can benefit directly from collecting and analyzing the knowledge produced in any kind of environment, like social networking sites, platforms, conversations, social devices, IoT devices, surveillance systems, virtual worlds, to mention some. Such conditions provide an immense and rich quantity of information that could be processed and used for public health benefit.

**2.9. Patient Cantered Area:** Adapt predictive analytics to clinical databases (e.g., optimization and predictive designing) to target people who could improve from preventive treatment or lifestyle changes, like those people at risk of acquiring a particular disease (e.g., diabetes) who could profit from preventive services.

**2.10. Hospital Operations and Business Development:** According to the provided better care with the help of proper use of health information to the patients, it increases the market share and helps to make proper decision for hospital operations [18].

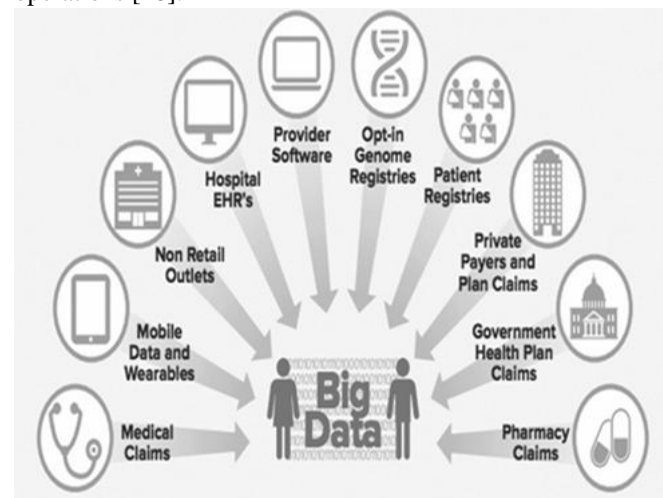


Figure.3. Uses Big Data in Healthcare

### III. ROLE OF BIG DATA TECHNOLOGIES IN HEALTHCARE

Big data innovations have opened up new possibilities and opportunities to enable developments related to health care management that essentially answer different perspectives that are insightful in reacting to much of what occurred, investigating why this occurred, predicting to understand what will occur in the future as well as the perspective about how we can make this work that is rightly explained in Figure.4.

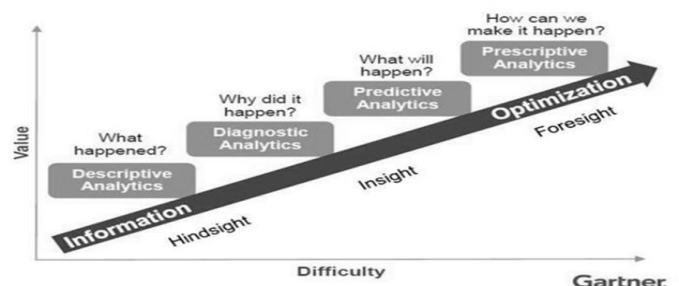


Figure.4. Data patterns including hindsight, insight and foresight



**3.1. Cloud Storage:** You can use cloud storage to upload data or have the entire system built in the cloud. Therefore, the cloud would need adequate storage capacity while at the same time sufficient data uploading rate.

It should also be able to store graphic form such as X ray, CT or MRI, apart from the data containing word supporting documents. The machine should also be in a position to produce graphical displays [15, 18].

**3.2. Hospital Information System:** A hospital information system (HIS) is basically a software program capable of handling all of the information to allow healthcare professionals to do their jobs correctly. The module includes clinical and non-clinical and other support services areas of hospitals including administration, billing etc...Some Common Mistakes [16, 17, 18]

## IV. BIG DATA: PREDICTIVE ANALYTICS AND STATISTICAL TECHNIQUES

Healthcare analytics refers to the comprehensive use of health information and business related knowledge built through application of quantitative models, e.g. mathematical, situational, empirical, analytical, behavioral and other models, to support reality-based decision making in health preparation, leadership, assessment and education [10]. Big data analytics can go further than performance optimization and reducing waste to forecast outbreaks, cure illnesses, improve life and minimize avoidable deaths [5].

## V. BIG DATA ANALYTICS IN HEALTHCARE- ORGANIZATION

Although several specific methods are being built in the big data discipline, here we outline is according to Raghupathi W & Raghupathi V which is realistic and hands-on.

**Table: 1. Source: Adapted from Raghupathi W & Raghupathi V [13].**

<p><b>Step 1 Concept statement</b></p> <ul style="list-style-type: none"> <li>Establish need for big data analytics project in healthcare based on the "4Vs".</li> </ul>
<p><b>Step 2 Proposal</b></p> <ul style="list-style-type: none"> <li>What is the problem being addressed?</li> <li>Why is it important and interesting?</li> <li>Why big data analytics approach?</li> <li>Background material</li> </ul>
<p><b>Step 3 Methodology</b></p> <ul style="list-style-type: none"> <li>Propositions</li> <li>Variable selection</li> <li>Data collection</li> <li>Data transformation</li> <li>Platform/tool selection</li> <li>Conceptual model</li> <li>Analytic techniques -Association, clustering, classification, etc.</li> <li>Results &amp; insight</li> </ul>
<p><b>Step 4 Deployment</b></p> <ul style="list-style-type: none"> <li>Evaluation &amp; validation</li> <li>Testing</li> </ul>

## VI. FINDINGS

The utilization of huge information by Big Data Analytics in healthcare has a great deal of positive and furthermore life-sparing results. After assessing all the papers and literature the findings are really attracting and influence the hole inside organized and unstructured information sources by Big Data. It helps to healthcare providers, users and policymakers in proper decision making. The exponential development of clinical information from different areas has constrained computational specialists to structure creative techniques to dissect and decode such massive measure of information inside a given time allotment. Big Data alludes to the tremendous amounts of data made by the digitization of everything that gets united and broke down by explicit advancements.

## VII. ADVANTAGES OF BIG DATA IN HEALTHCARE & CONCLUSION

Big data may minimize bias in the seasonality impact or prejudice in the selection impact. Selection bias arises when the recent developments weight more strongly than previous events to change the situation, but this may lead to incorrect decisions. Big data can also integrate the data in real time. For example, any organizational mistakes or trouble shooting can be detected instantly and the operational problem can be solved. The services could also be further enhanced, as the actual time offers the latest updates on the topic, For example, it will provide full patient records and will be able to administer it at the same time. Big data can also be used in medical computational modeling to detect and resolve the medical issues until it becomes an uncontrollable issue. Healthcare providers can decrease the risk and fix the problem with the big data information. It has also observed that big data is playing a very big role in smooth data transaction at healthcare facilities by using different management information systems.

## REFERENCES

- Douglas L, "3D Data Management: Controlling Data Volume, Velocity and Variety" February 2001. (Accessed on 01 December 2018).
- Elephantscale. (n.d.). elephantscale/hadoop-book Retrieved from <https://github.com/hadoop-illuminated/hadoop-book>.
- The Four V's of Big Data. (n.d.). Retrieved from <http://www.ibmbigdatahub.com/infographic/four-vs-big-data> (Accessed on 01 December 2018).
- Ward, Stuart, J., Barker, & Adam. (2013, September 20). Undefined By Data: A Survey of Big Data Definitions. Retrieved from <http://arxiv.org/abs/1309.5821v1> (Accessed on 03 December 2018)
- How Big Data Is Changing Healthcare - Forbes. (n.d.). Retrieved from <https://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/> (Accessed on 05 December 2018)
- Wang, L. (2017). Big Data in Healthcare: A New Frontier in Personalized Medicine. Open Access Journal of Translational Medicine & Research, 1(1). doi: 10.15406/oajtmr.2017.01.00005
- Big Data Analytics Use Cases: R/X for Healthcare. (n.d.) Retrieved from <https://mapr.com/blog/big-data-analytics-use-cases-rx-healthcare> (Accessed on 05 December 2018)
- Bigger Data for Better Healthcare. (n.d.). Retrieved from <https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/bigger-data-better-healthcare-ids-insights-white-paper.pdf>.

9. Gartner\_Inc. (n.d.). Predictive Analytics. Retrieved from <http://www.gartner.com/it-glossary/predictive-analytics>
10. Guivarch, C., & Hallegatte, S. (2012). 2C or Not 2C? SSRN Electronic Journal. doi: 10.2139/ssrn.1988201.
11. Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1). doi: 10.1186/2047-2501-2-3
12. Manyika J, Chui M, Brown B, Buhin J, Dobbs R, Roxburgh C, Byers AH: *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. USA:McKinsey Global Institute; 2011.
13. Raghupathi, W. R. V. (2013). An Overview of Health Analytics. *Journal of Health & Medical Informatics*, 04(03). doi: 10.4172/2157-7420.1000132
14. Roesems-Kerremans, G. (2016). Big Data in Healthcare. *Journal of Healthcare Communications*, 01(04). doi: 10.4172/2472-1654.100033
15. Rallapalli, S., Gondkar, R., & Ketavarapu, U. P. K. (2016). Impact of Processing and Analyzing Healthcare Big Data on Cloud Computing Environment by Implementing Hadoop Cluster. *Procedia Computer Science*, 85, 16–22. doi: 10.1016/j.procs.2016.05.171
16. Hussain, Md Kamal & Hussain, Md. (2012). A Brief Study on Role and Functions of Hospital Information System in Tertiary Care Hospitals. *International Journal of Computer Science And Technology*, 3(2), 327–332.
17. Hussain Mdk & AO Babalghith, "Health Information Management & Technology: A New Era of Transforming Healthcare" (2012) *International Researcher Volume 1, Issue 2, Publisher- I researcher ISSN 227-7471*
18. Hussain, Md Kamal. (2018). Healthcare Innovations, Health Needs and Emerging Market Issues Assessment from existing literature. *Research Journal of Social Science & Management*. 7. 160-166.

#### AUTHORS PROFILE



**Mohammad Kamal Hussain**, is Associate Professor at Department of Health Information Technology & Management, Also serving as Head- Quality Unit at Faculty of Public Health & Health Informatics, Umm Al- Qura University Makkah, Saudi Arabia.



**Mohammad Javed Hussain**, is Lecturer at Department of Information System, Faculty of Computer & Information System, Umm Al Qura University, Makkah, Saudi Arabia. He has several years of experience in Computer Science Industry.



**Mohammed Bakri**, is working as Assistant Professor at Department of Epidemiology, Faculty of Public Health & Health Informatics, Umm Al- Qura University Makkah, Saudi Arabia.



**Taher Mohammed Abdurraheem**, is working as Assistant Professor Health Information Technology & Management, Faculty of Public Health & Health Informatics, Umm Al- Qura University Makkah, Saudi Arabia



**Mahmoud Al-Areefi**, is working as Assistant Professor at Department of Health Information Technology & Management, Faculty of Public Health & Health Informatics, Umm Al- Qura University Makkah, Saudi Arabia