International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 02 | Feb -2017 www.irjet.net p-ISSN: 2395-0072

Review paper on Big Data in healthcare informatics

Harshit Kumar[1], Nishant Singh[2]

[1]Student B.E(3rd year),Dept.of Computer Engg., SKNCOE , Pune University
[2]Student B.E(3rd year),Dept.of Computer Engg., SKNCOE, , Pune University

---

Abstract *- The rapid increase in computational power, the number of internet enabled data generating devices and the falling costs of data storage itself which make data available to everybody for virtually no cost have primarily lead to the emergence of big data. Health Care is one of the major areas where the use of big data analytics has become monumental in rendering productive performance as compared to the conventional means. Big data mainly deals with the storage and processing of large scale and complex data sets for which the traditional methods prove to be inept. In this paper a survey on use of big data analytics in health care has been made to provide an insight overview of the technology, methodology and algorithms in big data used for data management and decision making in healthcare .the outcomes of this survey paper will prove to be beneficial to academician, researchers, industries who have interest in big data and healthcare specifically.*

Key Words**: big data, healthcare informatics,volume,velocity.variety,varacity,stakeholders,genomic analytics ,opportunities**

1.INTRODUCTION

The healthcare industry historically has generated huge amount of data, driven by record keeping, compliance & regulatory requirements, and patient care**[1].** Historically, the point of care generated mostly unstructured data: office medical records, handwritten nurse and doctor notes, hospital admission and discharge records, paper prescriptions, MRI, CT and other images.

The increase in digitization of data in healthcare industry has started producing data that fits in the definition of big data by all the attributes and definitions. The analytics of these digital data will provide multidimensional benefits in clinical practices, disease surveillance, population health administration and management in healthcare industry. By definition, big data in healthcare refers to electronic health data sets so large and complex that they are difficult (or impossible) to manage with traditional software and/or hardware; nor can they be easily managed with traditional or common data management tools and methods **[2].** It includes clinical data from CPOE(Computerized physician order entry) and clinical decision support systems (physician's written notes and prescriptions, medical imaging, laboratory, pharmacy, insurance, and other administrative data); patient data in electronic patient records (EPRs); machine generated/sensor data, such as from monitoring vital signs; social media posts, including Twitter feeds (so-called tweets) **[3]**, blogs **[4]**, status updates on Facebook and other platforms, and web pages; and less patient-specific information, including emergency care data, news feeds, and articles in medical journals.

2) V's OF THE HEALTH CARE BIG DATA

2.1)Volume

According to Health Catalyst **[5],** healthcare firms with over 1,000 employees store over 400 terabytes of data per firm (reported in the year of 2009), which qualifies healthcare as a high-data volume industry, despite the real-time streams of web and social media data. Contributing to the huge volume of healthcare data are various sources of data, from traditional personal medical records and clinical trial data to new types of data such as various sensor readings and 3D imaging **[6]**.Recently the proliferation of wearable medical devices has significantly added fuel to the healthcare data. Those devices are able to continuously monitor a series of physiological information, such as bio potential, heart rate, blood pressure, and so forth **[7].**

2.2) Variety

Healthcare data could be characterized by the variety of sources and the complexity of different forms of data. Generally, healthcare data could be classified into unstructured, structured, and semi structured. Historically, most unstructured data usually come from office medical records, handwritten notes, paper prescriptions, MRI, CT, and so on.**[30]** The structured and semi structured data refers to electronic accounting and billings, actuarial data, laboratory instrument readings, and EMR data converted from paper records **[8]**.Nowadays, more and more data streams add variety to healthcare information, both structured and unstructured, including intelligent wearable devices, fitness devices, social media, and so on.

2.3)Velocity

Compared with relatively static data such as paper files, x-ray films, and scripts, it is gradually becoming more important and challenging to process a real-time stream, such as various monitoring data, accurately and in a timely manner, in order to provide the right treatment to the right patient at the right time **[6]**. A concrete example can be found in the prevalence of wearable monitoring devices, which provide continuous and ever-accumulated physiological data. Being able to perform real-time analytics on continuous monitoring data could help predict life-threatening pathological changes and offer appropriate treatment as early as possible.**[30]**

2.4)Veracity

Coming from a variety of sources, the large volume of healthcare data varies in its quality and complexity. It is not uncommon that the healthcare data contains biases ,noise, and abnormalities, which poses a potential threat to proper decision-making processes and treatments to patients. The biggest challenge is determining the proper balance between protecting the patient's information and maintaining the integrity and usability of the data.

As Techcrunch **[9]** points out, "while today we rely on the well-trained eye of the general practitioner and the steady hand of the surgeon, tomorrow's lifesavers will be the number-crunching data scientists, individuals with only a passing understanding of first aid."

3.STAKEHOLDERS

The various stakeholders in healthcare industry have different expected incentives and hopes from Big Data which can be summarized as follows:

3.1)Patients want their everyday use of technology to flow seamlessly into their medical care. Some want to comparison shop for medical treatment as they do for consumer products. Everyone wants customer-friendly service, one-stop shopping, and better coordination of care between themselves, caregivers and various providers, with an ultimate goal of error-free, compassionate and effective care.

3.2)*Providers* want real-time access to patient, clinical and other relevant data to support improved decision-making and facilitate effective, efficient and error-free care. They want technology to be a transparent tool, not an encumbrance.

3.3)Researchers want new tools to improve the quality and quantity of workflow – e.g., predictive modeling, statistical tools and algorithms that improve the design and outcome of experiments and provide a better understanding of how to develop treatments that meet unmet needs while successfully navigating the regulatory approval and marketing process.

3.4)Pharmacy companies want to better understand the causes of diseases, find more targeted drug candidates, and design more successful clinical trials to avoid late failures and market safer and more effective pharmaceuticals. Once in the market, they want accurate formulary and reimbursement information to customize their marketing efforts, as well as less costly post-marketing surveillance.

3.5)Medical device companies, many of which have been collecting data for some time from hospital and home devices for safety monitoring and adverse event prediction, are beginning to wonder what to do with this data, and how to integrate it with old and new forms of personal data.**[10]**

4.OPPORTUNITIES

By digitizing, combining and effectively using big data, healthcare organizations ranging from single-physician offices and multi-provider groups to large hospital networks and accountable care organizations stand to realize significant benefits **[11]**.Implicit benefits of big data analytics in healthcare include earlier detection of diseases and ailments when they are in early stages and can be controlled and treated more easily and efficiently; individual health management by providing patient centric services; improving the treatment methods and detecting healthcare fraud more quickly and efficiently. McKinsey estimates that big data analytics can enable more than $300 billion in savings per year in U.S. healthcare, two thirds of that through reductions of approximately 8% in national healthcare expenditures. Clinical operations and R & D are two of the largest areas for potential savings with $165 billion and $108 billion in waste respectively **[12]**. McKinsey believes big data could help reduce waste and inefficiency in the following areas:

4.1)Clinical Operations: 1) Comparative effectiveness research to determine more clinically relevant and cost-effective ways to diagnose and treat patients.**2)** Clinical decision support systems to enhance the efficiency and quality of operations; i.e., providing real-time information to emergency technicians, nurses and doctors to improve triage, diagnosis, treatment choice, prevent iatrogenic infections and readmissions, prescription and other medical errors.**3)** Other areas include increasing transparency about medical data, remote patient monitoring, and predictive analytics to identify individuals who would benefit from proactive care. **[13]**

4.2) Research & development: 1) predictive modeling to lower attrition and produce a leaner, faster, more targeted R & D pipeline in drugs and devices; 2) statistical tools and algorithms to improve clinical trial design and patient recruitment to better match treatments to individual patients, thus reducing trial failures and speeding new treatments to market; and 3) analyzing clinical trials and patient records to identify follow-on indications and discover adverse effects before products reach the market.

4.3)Public health: 1) analyzing disease patterns and tracking disease outbreaks and transmission to improve public health surveillance and speed response; 2) faster development of more accurately targeted vaccines, e.g., choosing the annual influenza strains; and, 3) turning large amounts of data into actionable information that can be used to identify needs, provide services, and predict and prevent crises, especially for the benefit of populations **[12]**.

4.4)Genomic analytics: Add genomic analysis to the traditional healthcare decision making process by developing efficient and effective gene sequencing technologies.Utilizehighthroughputgeneticsequencerstocap tureorganismDNAsequencesand perform genome-wide association studies (GWASs) for human disease and human microbiome investigations.**[14]**

4.5) Fraud detection: Analyze a large amount of claim requests rapidly by using a distributed processing platform (e.g., MapReduce for Hadoop) to reduce fraud, waste, and abuse, such as a hospital's overutilization of services, or identical prescriptions for the same patient filled in multiple locations**[15]**

4.6) Device/remote monitoring: Capture and analyze continuous healthcare data in huge amounts from wearable medical devices both in the hospital and at home, for monitoring of safety and prediction of adverse events.**[4]**

5.ISSUES AND CHALLENGES

Along with the benefits that healthcare has been leveraging from big data, there is certain issues and challenges that act as barriers in successful implementation of big data for healthcare and prove to be a hindrance in proper and maximum extraction in the terms of advantages that big data can actually offer. Big data analytics not only provides charming opportunities but also faces lot of challenges. The challenge starts from choosing the big data analytics platform. While choosing the platform, some criteria like availability, ease of use, scalability, level of security and continuity should be considered **[16]**. The other challenges of big data analytics are data incompleteness, scalability and security **[17],[18]**.

Some of the challenges are:

5.1) Privacy and Data Security

Internet transactions, cloud storage, social media communications and related data exposes the personal and private data to potential and implicit misuse which makes the privacy of the data a grave issue which need to addressed and tackled earnestly. Privacy of data specific to the field of healthcare is in the terms of 1)The legal traditional doctor-patient confidentiality 2)The concern of patients regarding disclosure of their health status to third parties 3)Conflicting desires of third parties (insurers, employers, etc.) to access data.**[21]** Use of the Internet, cloud computing and pooling of data all raise the data security stakes. "Healthcare data contains the intimate details of a person's life and we must respect and protect it with the highest security possible, ," declares Jason Gilder of Explorys.

5.2) Data Standardization And Data Structure issues:

The data available in the healthcare industry is largely in an unstructured format which is in the format of graphs, prescription notes ,images. Apart from this, the nature of structured data is mostly heterogeneous . Leveraging the patient/data correlations in longitudinal records. And understanding unstructured clinical notes in the right context. Is a grave problem .Although the EHRs share data within the same organization, intra-organizational, EHR platforms are fragmented, at best. Data is stored in formats that are not compatible with all applications and technologies **[21,24]** This lack of data standardization also causes problems in transfer of that data **[25,26]**.

5.3) Data Storage And Transfers:

Data generation is inexpensive as compared to data storage.The real problem is in efficiently storing the data such that different methodologies and technologies can be easily applied to extract the desired information successfully. Once data is generated, the costs associated with securing and storing them remain high **[26]**. Costs are also incurred with transferring data from one place to another as well as

analyzing it **[28,24,29]**.Though many technologies have been devised to store and transfer the structured data, data scientists are still struggling with finding better ways to do the same to unstructured and heterogeneous data. Unstructured data is not as easy as the structured data to be analyzed, processed, stored, or transferred .The need of the hour is finding cheaper and less expensive ways for storage and transmission of secure or insecure data.

5.4)Requirement of Appropriate skills:

The expertises in technical skills make the handling, storage, retrieval and implementation of big data cumbersome. The McKinsey Global Institute estimates that there will be a more than 100,000 person shortage through 2020. It means that mean 50–60% of data scientist positions may go vacant. Data scientists need highly technical skill sets. They must possess soft skills such as communication, collaboration, leadership, creativity and more **[27]**.The scarcity of employees and workers in healthcare industry with adequate skill required in handling the big data is a significant barrier.

6. CONCLUSIONS

This paper presents the overview of Big data in healthcare, the features of big data, the stakeholders of the data and the challenges dealing with the big data in healthcare informatics. Despite many opportunities and approaches for big data analytics in healthcare presented in this work, there are many other directions to be explored, concerning various aspects of healthcare data, such as the quality, privacy, timeliness, and so forth. Computational health informatics in the big data age is an emerging and highly important research field with a potentially significant impact on the conventional healthcare industry .The future of health informatics will benefit from the exponentially increasing digital health data.

REFERENCES

[1] Raghupathi W: Data Mining in Health Care. In Healthcare Informatics: Improving Efficiency and Productivity. Edited by Kudyba S. Taylor & Francis; 2010:211–223.

[2] Frost & Sullivan: Drowning in Big Data? Reducing Information Technology Complexities and Costs for Healthcare Organizations.http://www.emc.com/collateral/analyst-reports/frost-sullivan-reducing-information-technology-complexities-ar.pdf

[3] Bian J, Topaloglu U, Yu F, Yu F: Towards Large-scale Twitter Mining for Drugrelated Adverse Events. Maui, Hawaii: SHB; 2012. 9. Raghupathi W, Raghupathi V: An Overview of Health Analytics. Working paper; 2013

[4] Raghupathi W, Raghupathi V: An Overview of Health Analytics. 2013. Working paper**Google Scholar**K. Elissa, "Title of paper if known," unpublished.

[5] Jared Crapo. 2014. Big data in healthcare: Separating the hype from the reality. https://www.

[6] healthcatalyst.com/healthcare-big-data-realities. (2014).

[7] Bonnie Feldman, Ellen M. Martin, and Tobi Skotnes. 2012. Big data in healthcare hype and hope. *Technical*

[8] Report, Dr. Bonnie 360 (2012).

[9] 4 Kevin Hung, Yuan-Ting Zhang, and B. Tai. 2004.Wearable medical devices for tele-home healthcare.

[10] In *26th* Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS'04).

[11] Vol. 2. IEEE, 5384–5387.

[12] Bonnie Feldman, Ellen M. Martin, and Tobi Skotnes. 2012. Big data in healthcare hype and hope. *Technical*

[13] Report, Dr. Bonnie 360 (2012).

[14] Techcrunch. 2014. Healthcare's Big Data Opportunity. http://techcrunch.com/2014/11/20/healthcares-bigdata-opportunity/. (2014)

[15] Bonnie Feldman, Ellen M. Martin ,Tobi Skotnes (2012). Big Data in Healthcare Hype and Hope. online https://www.ghdonline.org/uploads/**big-data-in-healthcare**_B_Kaplan_2012.pdf

[16] . Burghard C: Big Data and Analytics Key to Accountable Care Success. IDC Health Insights; 2012

[17] . 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.

[18] Manyika J, Chui M, Brown B, Buhin J, Dobbs R, Roxburgh C, Byers AH: Big Data: The Next Frontier for Innovation, Competition, and Productivity. 2011, USA: McKinsey Global InstituteGoogle Scholar

[19] Data in Healthcare Hype and Hope, Bonnie Feldman Ellen M. Martin Tobi Skotnes,October 2012

[20] Daniel A. Reed and Jack Dongarra. 2015. Exascale computing and big data. Communications of the ACM 58, 7 (2015), 56–68.

[21] Wullianallur Raghupathi and Viju Raghupathi. 2014. Big data analytics in healthcare: Promise and potential. Health Information Science and Systems 2, 1 (2014), 3.

[22]] Alexandros Labrinidis and H. V. Jagadish, "Challenges and opportunities with big data," Proc. VLDB Endow. 5, pp. 2032-2033, August 2012.

[23]] I.A.T. Hashem, et al, "The rise of "big data" on cloud computing: Review and open research issues," Information Systems, 2014.

[24]] 22. Hsieh JC, Li AH, Yang CC. Mobile, cloud, and big data computing: contributions, challenges, and new directions in telecardiology. Int J Environ Res Public Health. 2013 Nov 13;10(11):6131–53. doi: 10.3390/ijerph10116131. http://www.mdpi.com/resolver?pii=ijerph10116131. [PMC free article] [PubMed][Cross Ref]

[25] 5. Chawla NV, Davis DA. Bringing big data to personalized healthcare: a patient-centered framework. J Gen Intern Med. 2013 Sep;28(Suppl

3):S660–5. doi: 10.1007/s11606-013-2455-8.http://europepmc.org/abstract/MED/23797912. [PMC free article] [PubMed] [Cross Ref]

[26] 25. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry. 2013;35(4):332–8. doi: 10.1016/j.genhosppsych.2013.03.008.http://linkinghub.elsevier.com/retrieve/pii/S0163-8343(13)00069-8. [PMC free article] [PubMed] [Cross Ref]

[27] Overcoming Healthcare Big Data Challenges, retrieved from http://www.mckesson.com/healthcare-analytics/healthcare-big-data-challenges/#footNote

[28] Song TM, Song J, An JY, Hayman LL, Woo JM. Psychological and social factors affecting Internet searches on suicide in Korea: a big data analysis of Google search trends. Yonsei Med J. 2014 Jan;55(1):254–63. doi: 10.3349/ymj.2014.55.1.254. http://www.eymj.org/DOIx.php?id=10.3349/ymj.2014.55.1.254. [PMC free article] [PubMed] [Cross Ref]

[29] . Naqishbandi T, Imthyaz Sheriff C, Qazi S. Big data, CEP and IoT: redefining holistic healthcare information systems and analytics. Int J Eng Res and Technol. 2015;4(1):1–6.https://www.researchgate.net/profile/Tauseef_Naqishbandi/publication/271515514_Big_Data_CEP_and_IoT_-_Redefining_Holistic_Healthcare_Information_Systems_and_Analytics/links/54ca5ea70cf2c70ce521d024.pdf

[30] ACM Computing Surveys, Vol. 49, No. 1, Article 12, Publication date: June 2016. JIUOGU FANG*, SAMIRA POUYANFAR*, YIMIN YANG, SHU-CHING CHEN, and S. S. IYENGAR]

BIOGRAPHIES

Harshit Kumar
Student B.E (3rd year)
Dept.of Computer Engineering.,
SKNCOE vadgaon bk. ,Sinhgad institutes, Pune University ,Maharashtra , India
sharmaharshit95770@gmail.com

Nishant Singh
Student B.E (3rd year)
Dept.of Computer Engineering.,
SKNCOE vadgaon bk. ,Sinhgad institutes, Pune University ,Maharashtra , India
nishant.singh.210596@gmail.com