The Role of Big Data Analytics in Healthcare

Naveed Mustafa Hussain

ITM 100 - 01 Research Paper 1
09/26/2019

Introduction:

Gartner (circa 2001) defines Big Data as data that contains greater variety arriving in increasing volumes and with ever-higher velocity. To simplify, big data is larger, more complex data sets which are usually so voluminous that traditional data processing software cannot manage them. Such large amounts of data are prevalent in multiple industries, especially healthcare. The healthcare industry generates large amounts of data, driven by record keeping, compliance & regulatory requirements, and patient care. However, despite most of this data being stored in hard-copy form, the trend has arose for rapid digitization of these large amounts of data. Reports say, in 2011, data from the US healthcare system alone reached 150 exabytes and at this rate, the big data will soon reach the zettabyte (10^{21} gigabytes) scale and shortly after, the yottabyte (10^{24} gigabytes). In fact, Kaiser Permanente, the California based health network, supposedly holds between 26.5 to 44 petabytes of potentially rich data from EHRs on its own. (IHTT: *Transforming Health Care through Big Data Strategies*). Big data includes structured data that is easier to process and unstructured data that must be converted and compiled. The table belows differentiates this with examples:

Data category	Item	Description
Structured data	Demographics of the patient	Patient's gender, age, height, weight, etc.
	Living habits	Whether the patient smokes, has a genetic history, etc.
	Examination items and results	Includes 682 items, such as blood, etc.
	Diseases	Patient's disease, such as cerebral infarction, etc.
Unstructured text data	Patient's readme illness	Patient's readme illness and medical history
	Doctor's records	Doctor's interrogation records

(M Chen: *Disease Prediction by Machine Learning Over Big Data from Healthcare Communities,* : Item taxonomy in China hospital data, 2017)

In the healthcare industry, these electronic data sets consist of clinical data from CPOE and clinical decision support systems (written notes, prescriptions, medical imaging, laboratory, pharmacy, insurance and other administrative data), machine-generated/sensor data from patient monitoring, patient data in EPRs (Electronic Patient Records), social media posts, web pages with emergency care data, news feed and medical journal articles. It can be noticed that big data is overwhelming in the healthcare industry both because of its volume but especially because of its diversity of data types and the speed at which it must be managed. However, with proper

analytics, digitizing, combining and effective use of big date, healthcare organizations, as small-scale as a single-physician office to as large-scale as hospital networks, multiple significant benefits can be reaped. The implementation of more advanced technologies such as Machine Learning over Big Data can increase these benefits as well, and overall can also save the healthcare industry from the risk of losing potentially millions of dollars in revenue and profits if available tools, infrastructure and techniques are not acquired to effectively leverage big data (LaValle S, Lesser E: *Big data, analytics and the path from insights to value*).

Advantages to Healthcare:

Financially, McKinsey estimates that big data analytics can enable over $300 billion in savings per year in U.S. healthcare. Clinical operations and R&D are the two largest areas for potential savings with $165 billion and $108 billion in waste respectively (Manyika J,: *Big Data: The Next Frontier for Innovation, Competition, and Productivity)*. Other aspects benefitted include:

- **Clinical Operations:** Research of this data can be used to determine more clinically relevant and cost-effective ways to diagnose and treat patients.
- **Research & Development:** Statistical tools and algorithms improve the clinical trial design and patient recruitment to cater more effectively to individual patients, hence decreasing trial failures and increasing the speed to which new treatments enter the market. Analyzing clinical trials and patient records also allows for identification of follow-on indications and adverse effects before healthcare products enter the market.
- **Public Health:** Analyzing data patterns and tracking disease outbreaks improve public health surveillance and speed up response. Accurately targeted vaccines are developed faster and large amounts of data can be converted to actionable information utilized to identify needs, provide services and predict and prevent crises.
- **Evidence-based Medicine:** Unstructured data-EMRs can be combined and analyzed as well as financial and operational data, clinical data and genomic data to match treatments with outcomes and hence predict patients at risk for disease, providing more efficient care.
- **Genomic Analytics:** Big data analytics can be used to execute gene sequencing both more efficiently and cost effectively, enabling genomic analysis to become a component of the

regular medical care decision process and the growing patient medical record (IBM: *Large Gene interaction Analytics at University at Buffalo, SUNY;* 2012)

- **Patient Profile Analytic**s: Application of advanced analytics to patient profiles is enabled using big data. This identifies individuals who would benefit from proactive care or lifestyle changes (IBM: *IBM big data platform for healthcare. Solutions Brief,* 2012)
- **Device/remote Monitoring:** Large volumes of fast moving data from in-hospital or in-home devices can be captured and analyzed real time, enhancing safety monitoring and adverse event prediction as we know it.

Challenges for Big Data in Healthcare:

There are numerous challenges involved in the utilization of big data for the healthcare industry. One of the biggest issues in handling of big data (which can be applied to any industry) is privacy. While privacy issues have become increasingly urgent in recent times due as Internet transactions, cloud storage, social media and mobile devices expose more and more personal data for misuse, privacy issues in healthcare is a special issue. This is because, doctor-patient confidentiality is traditional and legal, there are concerns by individuals about disclosure of personal health information to third parties (eg. payers and insurers, caregivers, outsiders such as media or criminals) and the fact that majority consider health data privacy a fundamental right, as issued by the HIPAA (Health Information Portability and Accountability Act). In a survey on privacy in 2011, 80% of the public and doctors agreed that privacy safeguards for health information are important (Markle: *The Public and Doctors Agree on Importance of Specific Privacy Protections for Health IT,* 2011). Using big data analytics exposes confidential information, records and statistics in huge numbers to virtually be at risk.

Jason Gilder of Explorys stated, “Healthcare data contains the intimate details of a person’s life and we must respect and protect it with the highest security possible.” (Bonnie Feldman, Ellen M Martin, Tobi Skotnes: *Big Data in Healthcare, Hype and Hope*, 2012) This is a statement agreed upon by many companies in concern of the raising of data security stakes if this big data is digitized on cloud platforms for analytical purposes. Unintentional exposure or loss of the data to unauthorized parties are all major concerns regarding the security, protection and confidentiality of this data.

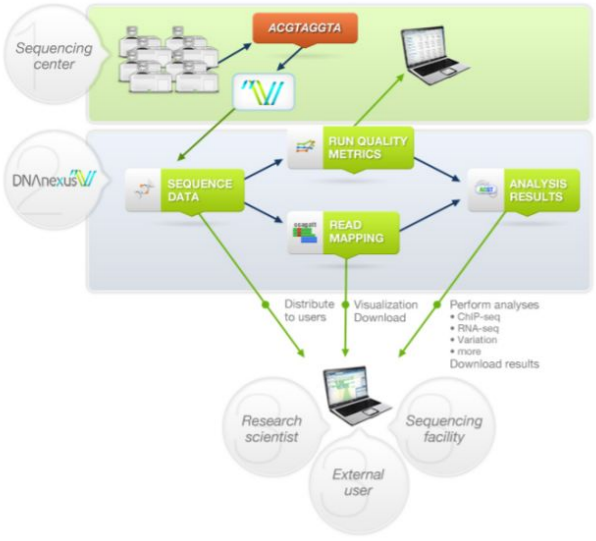
Big data analytics in healthcare raise fiscal and policy issues as well. In a fee-for-service environment, by introducing algorithms and technologies to healthcare data, healthcare practitioners lose their only source of earning- face-to-face encounters with patients. This creates a heavy bias against promoting technologies that streamline face-to-face interactions.

At the same time, technology issues and limitations are equally present in big data analytics. Health data is largely fragmented into institution centered silos. Exchanging individual records between silos, using increasingly standardized vocabularies (code sets) and message formats (ADT, C-CDAs) is where most current effort is being directed, yet the problem of data fragmentation prevails.

Current use of Big Data Analytics in Healthcare:

Although big data analytics have immense applicable potential, digitization and use of such health data is being used even today, by multiple companies, for different aspects. A few of these companies include:

1- **DNAnexus** offers a cloud-based, collaborative and scalable data technology platform that provides next-generation sequencing data management, analysis and visualization. It enables customers to store, manage, analyze and visualize next-generation DNA sequencing data through a web-based cloud service model. The firewalls and encryption support enterprise security and compliance with HIPAA, CLIA and other regulations. DNAnexus' customers include research scientists and clinical research partners in pharma and biotech, as well as physicians who utilise genomics in individual cases for diagnosis and treatment guidance.



(Bonnie Feldman, Ellen M Martin, Tobi Skotnes, *Big Data in Healthcare: Hype and Hope,* 2012)

2- **Predixion Software** uses cloud-based predictive analytic software to explain patterns in hospital datasets to reduce readmissions and help prevent hospital-acquired conditions. It does so but pulling data sets from a variety of sources, using data mining, machine learning and mathematical algorithms to power predictions. Its predictive analytics algorithm risk scores patients upon admission and throughout their hospital stay and identifies those at risk of readmission before they leave the hospital at an 86% accuracy. Currently, the project aims to apply analytics to prevent MRSA infections and deaths in hospital settings and use predictive analytics as a tool for prevention of chronic disease such as diabetes.

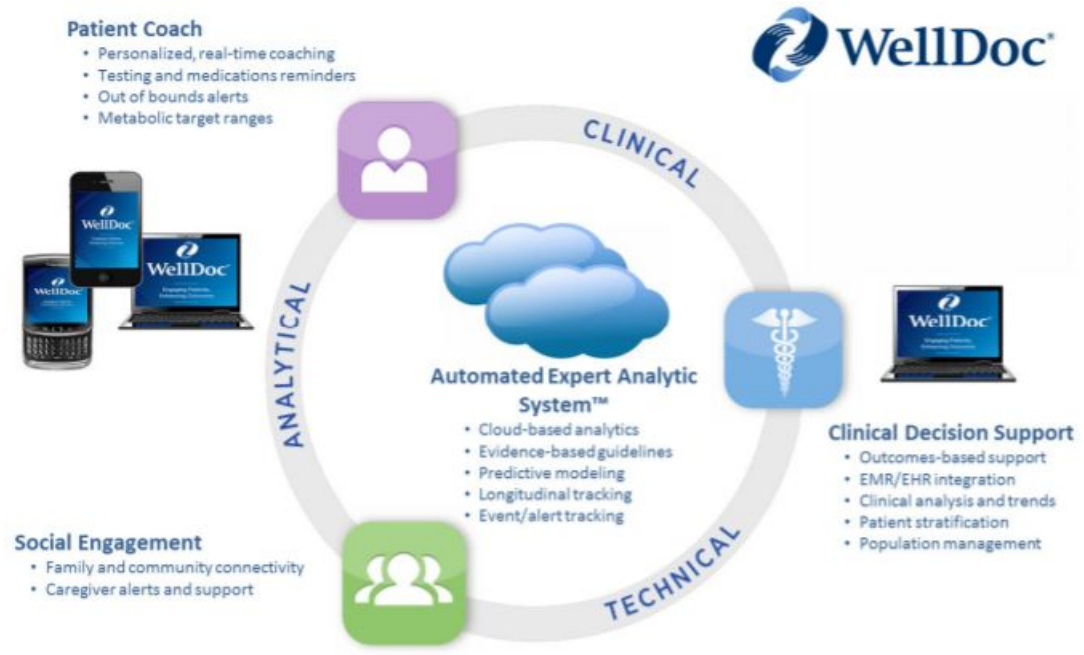
3- **Sickweather LLC** utilises big data from social media (Facebook, Twitter) to track outbreaks of disease, offering users forecasts similar to weather. It offers real time sickness forecasting services to keep individuals aware of outbreaks in their area. Individuals can be members and input information about the situation in their area as well.

(Sick Weather: *sickweather.com/how*)

4- **Sproxil** uses Big Data to identify counterfeit drugs, to protect patient health and enable pharmaceutical companies to track drug distribution and prevent theft. PIN codes are added to each individual drug product package to identify whether the drugs are real or counterfeit. An NLP algorithm is currently being adapted to understand multiple languages to globalize this tool and promote safety in the medicinal drug industry.

5- **Welldoc** uses automated, real-time coaching that integrates behavioral and clinical messaging to help patients manage chronic diseases such as diabetes. It does so by capturing a wide variety of patient-reported structured and unstructured data on clinical and behavioral aspects of patient's health using the patient's mobile phone or web-based application. It offers real-time

feedback in response to the data entered. Utilizing more and more of these real-time data sets of events and responses, big data analytics can be used to improve its messaging and treatment. (Bonnie Feldman, Ellen M Martin, Tobi Skotnes, *Big Data in Healthcare: Hype and Hope,* 2012)



6- **Zeo, Inc.** analyzes over a million nights of data to help consumers improve their sleep. It offers a personal sleep coach device that tracks the quality of users' sleep and then returns personalized advice on how to improve their sleep. They have shared their sleep data with academic institutions to further the collective understanding of sleepy and aim to combine this with data on blood pressure, weight, heart rate and other measures- so that they may aggregate and republish it.

Conclusion:

Big data analytics has the potential to transform the way healthcare providers implement technology to gain insight from clinical and other data repositories and apply it to numerous healthcare functions. In order to see a widespread implementation of big data analytics in health care, challenges such as privacy, safeguarding security, establishment of standards and

governance and streamlining of technology to reduce limitations must all be addressed. Although the applications of big data in healthcare are currently developing, it still has lead to multiple applications that benefit the community in everyday-life, through personal user applications to infrastructures for large establishments. Big data analytics are maturing rapidly and will play an immense role in revolutionizing healthcare to a more adaptive approach.

References:

1 -Oracle, Big Data https://www.oracle.com/big-data/guide/what-is-big-data.html

2- IHTT: *Transforming Health Care through Big Data Strategies for leveraging big data in the health care industry,* 2013. http://www.emc.com/collateral/analyst-reports/frost-sullivan-reducing-information-technology-complexities-ar.pdf

3- LaValle S, Lesser E, Shockley R, Hopkins MS, Kruschwitz N: *Big data, analytics and the path from insights to value*. MIT Sloan Manag Rev 2011, 52: 20-32

4- 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.

5- IBM: *Large Gene interaction Analytics at University at Buffalo, SUNY;* 2012. http://public.dhe.ibm.com/common/ssi/ecm/en/imc14675usen/IMC14675USEN.PDF

6- IBM: *IBM big data platform for healthcare, Solutions Brief,* 2012 .http://public.dhe.ibm.com/common/ssi/ecm/en/ims14398usen/IMS14398USEN.PDF

7- Markle: *The Public and Doctors Agree on Importance of Specific Privacy Protections for Health IT,* 2011 https://www.markle.org/publications/1443-public-and-doctors-agree-importance-specific-privacy-protections-health-it

8- Bonnie Feldman, Ellen M Martin, Tobi Skotnes: *Big Data in Healthcare, Hype and Hope*, October 2012 https://www.ghdonline.org/uploads/big-data-in-healthcare_B_Kaplan_2012.pdf

9- M Chen: *Disease Prediction by Machine Learning Over Big Data from Healthcare Communities*, 2017https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7912315&tag=1