

Big Data for HealthCare

By:

Prashant Badadhe

Tao Sun

Introduction:

In the past decade, Big Data and its techniques are successfully applied in many fields. In the field of weather forecasting, for instance, there are lots of weather sensors and satellites which are deployed all around the world. A huge amount of data, like temperature and humidity, is collected from them. On the other hand, big data is also actively used in e-commerce, especially the online shopping. Countless product information, pictures and videos are uploaded to the internet every day. However, with the support of big data and data mining algorithm, not only the accurate product information become accessible, but also the similar products are recommended automatically by system through learning your search history. There are still many instances about big data application in various fields, including the healthcare what we are going to discuss in this report.

Application Domain: Health Care

Why is Big Data Important for Healthcare?

Big data has become more influential in healthcare due to three major shifts in the healthcare industry: the vast amount of data available, growing healthcare costs, and a focus on consumerism. Big data enables health systems to turn these challenges into opportunities to provide personalized patient journeys and quality care.

- **Increasing Volume of Healthcare Data:** When health records went digital, the amount of virtual data health systems had to handle rose steeply. In addition to EHRs, vast amounts of data are sourced in other ways – through wearable technology, mobile applications, digital marketing efforts, social media, and more. All of this adds up to an incredible amount of information, spurring health systems to adopt big data systems and technologies to effectively collect, analyze, and take advantage of this information.
- **Growing Healthcare Costs:** In the past 20 years, the United States has seen a rapid growth in healthcare costs. Today, healthcare expenses account for around 18 percent of GDP, totaling about \$3.4 trillion. This is partially due to lifestyle factors, as well as government regulations. Through the collection and analysis of large amounts of data, healthcare organizations will find quantifiable ways to improve performance and efficiency. This promotes both increased patient satisfaction and your ability to capture greater market share.

- **Desire for Personalized Care:** Consumers in all industries expect exceptional, convenient, personalized service – a phenomenon that retail industry executives have dubbed “The Amazon Experience.” Healthcare is no different. Customers want convenient, personalized care, a new standard to which health systems must rise. This new model of care focuses on quality, engagement, and retention. Health systems are turning to healthcare big data to provide the insights necessary to drive this level of personalization.

Four V's of Big Data: The characteristics of Big Data are commonly referred to as the four Vs:

1. Volume of Big Data :

The volume of data refers to the size of the data sets that need to be analyzed and processed, which are now frequently larger than terabytes and petabytes. The sheer volume of the data requires distinct and different processing technologies than traditional storage and processing capabilities. In other words, this means that the data sets in Big Data are too large to process with a regular laptop or desktop processor. According to the statistics, healthcare data had explosively growth eight times since 2016, reaching 8.41 petabytes on average in 2018. These data come from many different sources. Pharmaceutical companies have been aggregating years of research and development data into medical databases. In parallel, recent technical advances have made it easier to collect and analyze information from multiple sources.

2. Velocity of Big Data:

Velocity refers to the speed with which data is generated. High velocity data is generated with such a pace that it requires distinct (distributed) processing techniques. An example of a data that is generated with high velocity truly reflect the patient’s status from different angles. The transactional data in the healthcare industry changes rapidly. Claims are paid daily; and the results of diagnostic tests are recorded electronically in real time. All these attributes support the assertion that healthcare data meet the velocity criteria.

3. Variety of Big Data

Variety makes Big Data really big. Big Data comes from a great variety of sources and generally is one out of three types: structured, semi structured and unstructured data. The variety in data types frequently requires distinct processing capabilities and specialist algorithms. An example of high variety data sets would be the audio and video files that are generated in various treatments. On the other hand, there are many different types of medical record which include text report, charts, images, videos and numerical data.

4. Variety of Big Data

Veracity refers to the quality of the data that is being analyzed. High veracity data has many records that are valuable to analyze and that contribute in a meaningful way to the overall results. Low veracity data, on the other hand, contains a high percentage of

meaningless data. The non-valuable in these data sets is referred to as noise. An example of a high veracity data set would be data from a medical experiment or trial.

Data that is high volume, high velocity and high variety must be processed with advanced tools (analytics and algorithms) to reveal meaningful information. Because of these characteristics of the data, the knowledge domain that deals with the storage, processing, and analysis of these data sets has been labeled Big Data.

Big data-related problems:

There are many challenges that make it hard to use healthcare data. First, the data in many healthcare data providers are often isolated or segmented. Administrative data are stored and used by the financial department or operational management teams only. They are responsible to deal with patient's cost information, claims and reimbursement.

- **Billing systems are fragmented and dated** – Data is often very “noisy” – practices, groups, and even service line specialties can be inconsistent. The key is to consider directional data in combination with your local geographic market knowledge; in other words, data should augment interactions and focused outreach to physicians, not replace.
- **Patients do not have unique patient identifiers** – If every patient had a unique identifier, data matching would not be required. Until that happens, data matching mechanisms are required to look for these data anomalies and put the right patient claims together.
- **Diagnosis and procedure codes can be unclear** – Even industry-standard grouper tools can obscure or mis-map physician activity. Perfect data and perfect insights are very hard to achieve, so you have to advocate for, and learn to work with, directional data.
- **Claims data is highly inconsistent** – With claims data, any field data that is not required for payment has a low probability of being completed accurately. In fact, among the few required fields for payment, along with patient, diagnosis, and procedure information, is the “rendering physician” via the NPI1 for that provider.
- **It's difficult to identify the referring physician** – The “referring physician” field on available third-party claims is often inconsistent, incorrect, or not filled at all. In fact, some clearinghouses don't even provide the “referring physician” field because of these inconsistencies.
- **Synchronization across Disparate Data Sources**-The use of big data in healthcare allows for strategic planning thanks to better insights into people's motivations. Care managers can analyze check-up results among people in different demographic groups and identify what factors discourage people from taking up treatment.
- **Getting Meaningful Insights Through The Use Of Big Data Analytics**-It is imperative for business organizations to gain important insights from Big Data analytics, and also it is important that only the relevant department has access to this information. A big challenge faced by the companies in the Big Data analytics is mending this wide gap in an effective manner.

Existing solutions

- To satisfy the need for analysis of different types of data, various initiatives were taken to build specific systems. The most representative of them is electronic health record.
- An electronic health record (EHR) is a digital version of a patient's paper chart. EHRs include all the important clinical data which is related to patients who are taking care by a particular provider, including clinical progress notes, medications, medical history, laboratory data and radiology reports.
- Other examples of big data analytics in healthcare share one crucial functionality – real-time alerting. In hospitals, Clinical Decision Support (CDS) software analyzes medical data on the spot, providing health practitioners with advice as they make prescriptive decisions.
- Many consumers – and hence, potential patients – already have an interest in smart devices that record every step they take, their heart rates, sleeping habits, etc., on a permanent basis. All this vital information can be coupled with other trackable data to identify potential health risks lurking. A chronic insomnia and an elevated heart rate can signal a risk for future heart disease for instance. Patients are directly involved in the monitoring of their own health, and incentives from health insurances can push them to lead a healthy lifestyle (e.g.: giving money back to people using smart watches).
- Also it discovery the treatments to reduce health care costs, reducing the number of hospital readmissions, reducing emergency department visits, reducing the waiting time of patients in the hospital, preventing adverse effects of the medicine, and many more.
- This repository of data has brought new hope and promises. People eagerly expect to discover new knowledge in health care research and breakthrough treatments as well as new challenges in technology and management.
- The use of big data in healthcare allows for strategic planning thanks to better insights into people's motivations. Care managers can analyze check-up results among people in different demographic groups and identify what factors discourage people from taking up treatment.

Solution Proposed:

- Create holistic, 360-degree view of consumers, patients, and physicians.
- Improve care personalization and efficiency with comprehensive patient profiles.
- Identify geographic markets with a high potential for growth.
- Inform physician relationship management efforts by tracking physician preferences, referrals, and clinical appointment data.
- Boost healthcare marketing efforts with information about consumer, patient, and physician needs and preferences.
- Provide straightforward identification of patterns in health outcomes, patient satisfaction, and hospital growth.

- Optimize hospital growth by improving care efficiency, effectiveness, and personalization
- To enhance the use of unstructured medical big data, we must establish the data collection and quality assurance processes.
- Then open platform for integration and utilization of the unstructured clinical data should be developed while reflecting these concepts.
- Additionally, the amount of data available will grow as wearable technology and the Internet of Things (IoT) gains popularity. With this information, healthcare marketers can integrate a large volume of healthcare insights to find and retain patients with the highest propensity for services.
- Develop such systems that can analyze and predict the future market trends for health care industry growth.
- With the help of digitization in health care domain we can predict the root cause of health issues and should develop systems that will help to prevent or identifies diseases in its early stage.
- Medical imaging is vital and each year in the US about 600 million imaging procedures are performed. Analyzing and storing manually these images is expensive both in terms of time and money, as radiologists need to examine each image individually, while hospitals need to store them for several years.
- Telemedicine has been present on the market for over 40 years, but only today, with the arrival of online video conferences, smartphones, wireless devices, and wearables, has it been able to come into full bloom. The term refers to delivery of remote clinical services using technology.

Lee, C., & Yoon, H. (2017). Medical big data: promise and challenges. *Kidney Research And Clinical Practice*, 36(1), 3-11. doi:10.23876/j.krcp.2017.36.1.3

Dash, S., Shakyawar, S., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal Of Big Data*, 6(1). doi:10.1186/s40537-019-0217-0

Liu, P. (2016). Medical Record System Using Blockchain, Big Data and Tokenization. *Information And Communications Security*, 254-261. doi:10.1007/978-3-319-50011-9_20

Mohammad Adibuzzaman, B. (2017). Big data in healthcare – the promises, challenges and opportunities from a research perspective: A case study with a model database. *AMIA Annual Symposium Proceedings*, 2017, 384. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977694/>

Big Data in Healthcare: Challenges & Promise | NEJM Catalyst. (2018). *NEJM Catalyst*. Retrieved 19 September 2019, from <https://catalyst.nejm.org/big-data-healthcare/>