# Chapter 53
# Application of Big Data in Healthcare: Opportunities, Challenges and Techniques

**Md Rakibul Hoque**
*University of Dhaka, Bangladesh*

**Yukun Bao**
*Huazhong University of Science and Technology, China*

## ABSTRACT

*This chapter investigates the application, opportunities, challenges and techniques of Big Data in healthcare. The healthcare industry is one of the most important, largest, and fastest growing industries in the world. It has historically generated large amounts of data, "Big Data", related to patient healthcare and well-being. Big Data can transform the healthcare industry by improving operational efficiencies, improve the quality of clinical trials, and optimize healthcare spending from patients to hospital systems. However, the health care sector lags far behind compared to other industries in leveraging their data assets to improve efficiencies and make more informed decisions. Big Data entails many new challenges regarding security, privacy, legal concerns, authenticity, complexity, accuracy, and consistency. While these challenges are complex, they are also addressable. The predominant 'Big Data' Management technologies such as MapReduce, Hadoop, STORM, and others with similar combinations or extensions should be used for effective data management in healthcare industry.*

## INTRODUCTION

The healthcare industry is one of the most important, largest, and fastest growing industries in the world. It has historically generated large amounts of data, "Big Data", related to patient healthcare and well-being (Nambiar, 2013). These data include clinical data from clinical decision support systems, patient data in electronic patient records, physician's prescriptions, pharmacies, insurance, administrative data, sensor data, social media posts, blogs, web pages, emergency care data, news feeds, and articles in medical journals (Bian et al., 2012; Raghupathi & Raghupathi, 2013). International Data Corporation, a global

DOI: 10.4018/978-1-4666-9840-6.ch053

market research firm, estimates that the amount of digital data will grow from 2.8 trillion gigabytes in 2012 to 40 trillion gigabytes by 2020 (IDC, 2012). A recent study estimates that over 30% of all data stored in the world are medical data and this percentage is expected to increase rapidly. In 2012, the volumes of worldwide healthcare data were 500 petabytes and are projected to reach 25,000 petabytes in 2020 (Feldman et al., 2012).

It is widely accepted that Big Data can transform the healthcare industry by improving operational efficiencies, the quality of clinical trials, and optimizing healthcare spending from patients to hospital systems (Koh & Tan, 2011). The potential for Big Data analytics in healthcare leads to better outcomes by analyzing patient characteristics and outcomes of care. It identifies the most clinically and cost effective treatments and offers analysis and tools. Big Data can assist patients to determine regimens or care protocols by collecting and publishing data on medical procedures. For example, broad scale disease profiling helps to identify predictive events and support prevention initiatives; and, implementing much nearer to real-time by aggregating and synthesizing patient clinical records. Moreover, licensing data can assist pharmaceutical companies to identify patients for inclusion in clinical trials.

The doctors will be able to understand which tests are not necessary and patients will be able to access information on the doctors for specific procedures with Big Data analytics (Chawla & Davis, 2013). The Big Data analytics can avoid errors, diagnostic accuracy and improve coordination of care by using high-quality data. In USA, the Obama Administration has invested USD 200 million for Big Data Research and Development initiative to transform the use of Big Data for biomedical research (STP, 2012). The government proposed "Health 2.0" to manage hospitals, patients, insurance and government efficiently. The U.S. healthcare alliance network, Premier, has more than 2,700 members, hospitals and health systems, 90,000 non-acute facilities and 400,000 physicians. It has assembled a large database of clinical, patient, financial, and supply chain data to generate comprehensive and comparable clinical outcome measures. The Korean government plans to operate the National DNA Management System which will offer customized diagnosis and medical treatment to patients (NICT, 2011).

However, the health care sector lags far behind other industries in leveraging their data assets to improve efficiencies and make more informed decisions. While other industries such as the insurance, banking and retail sectors are far advanced in leveraging Big Data techniques, health care remains poor at handling the flood of data. Researchers have raised concerns about how to ensure that Big Data has central role in a health system's ability to secure improved health for its users (Ohlhorst, 2012). Big Data entails many new challenges regarding security, privacy, legal concerns, authenticity, complexity, accuracy, and consistency.

There is increasing concern that millions pieces of important data are lost every day due to traditional storage technologies. This is problematic, as it does not allow health services to adapt to the needs of patients or diseases, as there are currently no tools being utilized that are capable of storing and managing so much information, although the technology exists (Diaz et al., 2012). While these challenges are complex, they are also addressable. Mass data storage in real-time technologies is needed. The predominant 'Big Data' Management technologies such as MapReduce, Hadoop, STORM, and others with similar combinations or extensions should be used for effective data management in healthcare industry.

## Big Data in Healthcare

About 2.5 quintillion bytes of data are generated every day and almost 90% of the existing global data has been created during the past two years (IBM, 2013). In 2011 alone, 1.8 zettabytes of data were cre-