Chapter 4
Opportunities and Challenges of Big Data in Healthcare

Wafaa Faisal Mukhtar
Sudan University of Science and Technology, Sudan

Eltayeb Salih Abuelyaman
University of Dammam, Saudi Arabia

ABSTRACT

Healthcare big data streams from multiple information sources at an alarming volume, velocity, and variety. The challenge that faces the healthcare industry is extracting meaningful value from such sources. This chapter investigates the diversity and forms of data in the healthcare sector, reviews the methods used to search and analyze these data throughout the past years, and the use of machine learning and data mining techniques to mine useful knowledge from such data. The chapter will also highlight innovations of particular systems and tools which spot the fine approaches for different healthcare data, raise the standard of care and recap the tools and data collection methods. The authors emphasize some of ethical issues regarding processing these records and some data privacy issues.

INTRODUCTION

We live in the age of Big Data, where every single data entry to any computer at home or office is being recorded in data centers and processed later in order to get valuable information or knowledge. IDC (International Data Corporation) defines Big Data technologies as a new generation of technologies and architectures, designed to extract economical *value* from very large *volumes* of a wide *variety* of data produced every day, by enabling high *velocity* capture, discovery, and/or analysis (Gantz & Reinsel, 2011). Other institutes and companies such as SAS, IBM and Oracle have other dimensions for Big Data, but they all address same concepts.

Several organizations create very large databases, such as consumer data and transaction histories of sales records, patient records, images at hospitals, fingerprints and DNA samples at crime scenes, and data acquired via satellites or surveillance cameras. It is apparent that there is a huge gap between the amount of available information and its corresponding knowledge, which gives an opportunity for extracting hidden knowledge.

DOI: 10.4018/978-1-5225-0920-2.ch004

Healthcare can be considered a major sector for producing big data. Healthcare data have many dimensions such as personal information regarding dates of birth, nationality, sex, location and other information about income and education plus other specific data about history of the patient with medical centers and/or disease. It is generated by everything around us at all times in intensive care units and in surgical operation room, inwardfrom every digital process in systems or sensors.

The aim of this chapter is to discuss the notion of big data when applied to the healthcare sector. The next section reveals how and why big data are represented in health care. Big Data's Vs will be emphasized for healthcare data in the third section. The fourth section highlights different techniques that are well suited for intelligent data analysis. Section five illustrates the challenges facing Big Data and the opportunities available for valuing analysis, and some ethical, legal and social concerns about manipulating healthcare data specifically. Some implications are contending in the future directions showing how they can play an important role in the overall improvement of the healthcare industry.

BACKGROUND

Medical data are at once the most rewarding and challenging of all biological data. For decades everyone was infatuated by the liability of keeping every record and collecting any possible information about everything in their life. The healthcare industry has also experienced these practices about generating and keeping large amounts of data driven by record keeping at physicians' clinics, which is referred to as patient records. This includes forms filled by the patient regarding his/her personal information and oral examination recorded by physicians during visits. Other forms of checkups, different laboratory examinations, and CT scan as well as X-ray images are also kept in hospital's emergency room when examining patients. Moreover, data about compliance & regulatory requirements, and patient care is also evolving from national and international organizations that monitor and administer the healthcare industry.

Electronic health records have experienced several studies. Drug safety study (Trifirò et al., 2009) investigated adverse drug reactions with other diseases, in (Jensen, Jensen, & Brunak, 2012), they combined the HER with the genetic data to reveal gene-disease association, (Almodaifer, Hafez, & Mathkour, 2011) discovered the interesting and concise medical rules for prediction purpose to assist the medical decision makers.

Medical diagnosis researches have proven a great success, because the data about the disease and the patient under examination is always available. In fact the medical diagnostic knowledge can be automatically derived from the description of cases solved in the past. (Kumar, Sathyadevi, & Sivanesh, 2011) proposed using an intelligent clinical decision support system to assist physicians in diagnosing. An automatic diagnosis system was presented in (Karabatak & Ince, 2009b). Soni & Ansari, 2011; Kharya, 2012; Huang, Chen, & Lee, 2007; (Ha, 2011) (Kononenko, 2001) summarized several machine learning techniques used for classifying diseases such as naïve Bayesian and neural networks, his work also highlighted the specific requirements for good performing machine learning algorithms in solving medical diagnostic tasks.

Image parameterization observed in medical records is very useful in quality control, identification, image grouping, surveillance, image storage and retrieval, and image querying. Texture classification is closely related to diagnostic process to diagnose coronary artery disease (Kukar, Kononenko, & Grošelj, 2011).