ISSN XXXX XXXX © 2018 IJESC

Research Article Volume 8 Issue No.9

Healthcare Analytics Systems: An Overview

Yash Gupta[1], Akanksha Joshi[2], Dr. Geetanjali Kale[3]
Software Engineer-Eq Technologic, Pune, India[1], Software Engineer-IBM ISL, Pune, India[2]
Assistant Professor, Pune Institute of Computer Technology, Pune, India[3]

Abstract:
In recent years, Healthcare Analytics has been seen as the domain for interdisciplinary research and invention. These inventions and advancements have a promising future. Automation in health care processes is an ongoing practice in the healthcare industry. An increase in patient satisfaction coupled with efficient care and better population health are few of the areas that will benefit from automatic healthcare improvement. In the healthcare industry, data mining plays an important role, as it enables health care systems to systematically use data and analyze it to identify inefficiencies and best practices that improve care and reduce cost. Preventing epidemic outbreaks, reducing cost of treatments and improving quality of life are the potential targets of using analytics on healthcare data. The doctors, clinicians, healthcare researchers and the medical industry as a whole are beneficiary of healthcare analytics. The paper presents a holistic view of the different components, applications, challenges and future direction of data analytics in healthcare.

Keywords-Analytics, Cloud Computing, Data Mining, Healthcare, IBM Watson, IoT, Machine learning, Sensors

I. INTRODUCTION

A revolution in the healthcare domain is now under way. This is an era in which the fast paced life has given rise to tremendous increase in the numbers of diseases and maladies; equivalently however, technology also has developed many folds. This has helped reduce the gap between the patients and the remedies available, leading to the betterment of human race. These days, people are becoming more cautious about their health. The advancements in the sensor technology, coupled with big data analytics and high performance computing are creating a transformation in the healthcare domain.
The continuously increasing reformations in biology and technology have led to the development of wearable sensors, or smart bands, as we call them today. These sensors generate zillions of data everyday that unless analyzed, are of no real value. Apart from this, by virtue of clinical trials and other medical examinations, millions of more other records are generated that don't usually make it beyond a person's personal medical file [1]. Parallelly, the development in the field of big data coupled with various algorithms has been the motivating factor for the success of analytical systems. These systems have thus been able to extract actionable insights from data that may otherwise seem redundant [2].
This study is therefore, aimed at presenting the usefulness of aggregating and correlating health care data from various sources, analyzing and then presenting them to paint a bigger picture. The study also aims to review the different big data frameworks along with the data sources, analytical capabilities and application areas.

II. RELATED LITERATURE

It has been suggested in a report by McKinsey Global Institute that an efficient use of Big Data is capable of creating more than $300 billion in value every year, thus reducing the expenditure on healthcare [3]. In their study of understanding the capabilities and potential benefits of big data analytics in healthcare; Yichuan Wang, LeeAnn Kung and Terry Anthony Byrd identify the different layers of the Big data analytics architecture. According to their research, the architecture consists of data layer, data aggregation layer, analytics layer, information exploration layer and the data governance layer. Technological advancements have made high precision sequencing of DNA possible. Through their research on future of genomics and healthcare, the authors suggest opening of new avenues owing to the analysis of transcriptomes, proteomes and metabolomes now being possible. The study of human genome along with the application of mathematical and computational tools has provided the foundations for deciphering the structure, variation and function of the human genome and relating them to health and disease states. Considering the American healthcare system, the various applications and the barriers to the widespread adoption of healthcare analytical systems have been explored in their study by Michael Ward, Keith Marsolo and Craig Froehle[4].
Apache Hadoop has also proved to be a highly suitable analytical tool for analysing the healthcare data. Healthcare data analytics using Hadoop plays an effective role in real-time analysis. Hadoop has the capacity to analyze huge volumes of data. This is achieved using the Map Reduce framework. The benefits of using Hadoop and similar systems, particularly for Indian healthcare systems has been described by D. Peter Augustine in his research by the name Leveraging Big Data Analytics and Hadoop in Developing India's Healthcare Services[5]. Similar survey has been done by J.Archenaa and E.A. Mary, focusing on healthcare analytics in the government sector[6].

The term Big Data is being frequently used, referred to and is applicable to almost every aspect of human life. A very crude definition of "Big Data" was coined by the Nasa researchers in 1997. Since then, the definition of "Big Data" has been modified owing to its utility and the gargantuan volume. The combination of 3 V's; Velocity(the high frequency/speed at which data is generated), Variety(the diverse nature of data), Volume(boundless amount of data collected) are the core characteristics defining Big Data. These characteristics of Big Data make it highly applicable to the Healthcare domain. If studied and used in a proper way, the insights derived from healthcare data can be very useful.

III. SYSTEM ARCHITECTURE

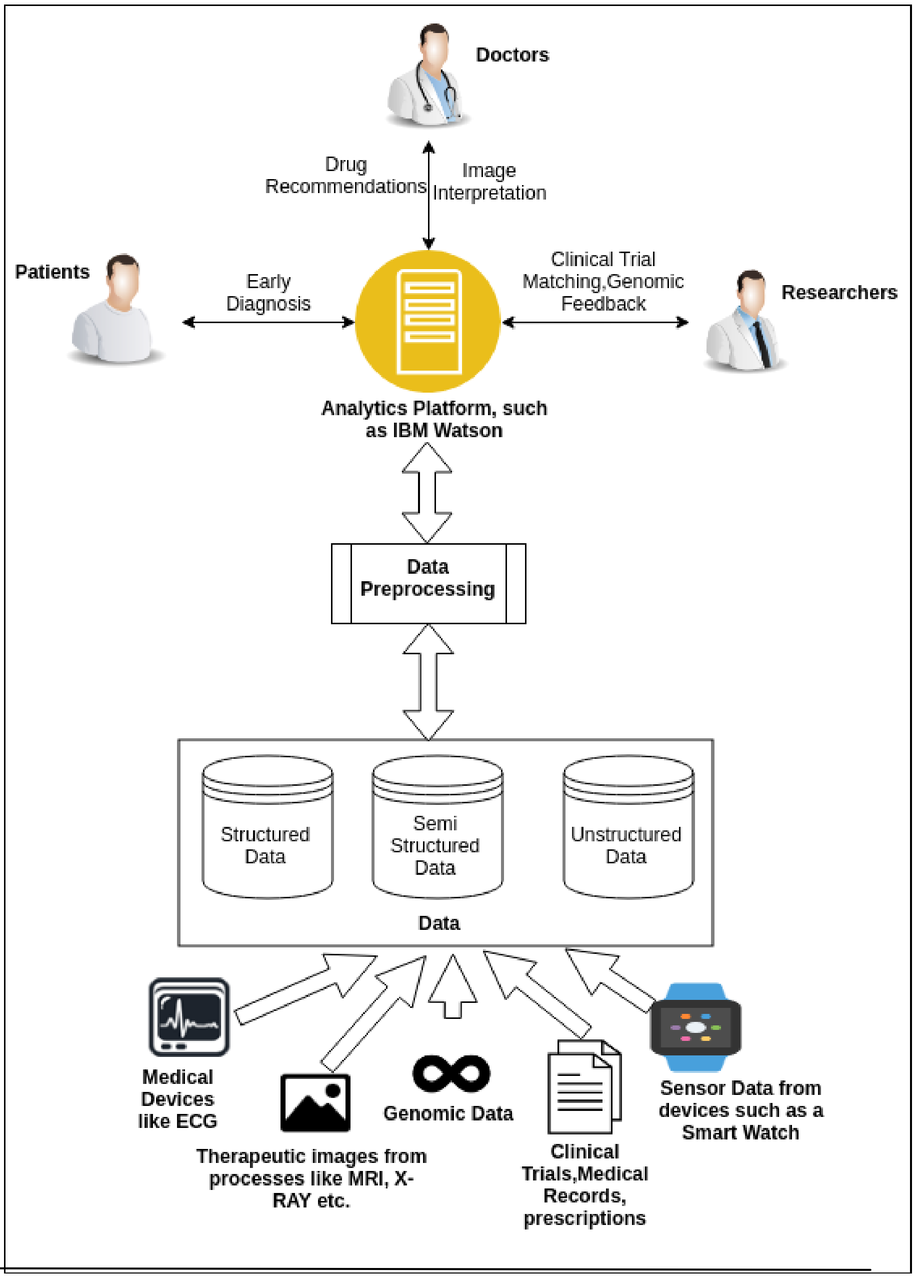

FIGURE 1: PROPOSED SYSTEM ARCHITECTURE

The proposed architecture aims to present a generalized view and functional linking between the various components of a healthcare analytics system. The major stakeholders involved are the patients, doctors and researchers. Information can be collected from a number of sources. Collected data can either be structured, semi-structured or unstructured. Subsequent section describes the data types and sources in more detail.

1. Data Sources

The large volume of data collected about patients comes from a variety of sources[7]:

I. **Clinical Data**: These basically include electronic/medical records, physician system records and pharmaceutical data. Diagnosis of health issues is often done using the images generated from processes like Computed tomography (CT), magnetic resonance imaging (MRI), X-ray, molecular imaging, ultrasound, photoacoustic imaging, fluoroscopy, positron emission tomography-computed tomography (PET-CT), mammography etc. It is necessary to convert this fragmented, unstructured information into a defined structure in order to be used as an input of the analytical systems.

II. **Exogenous data**: Records obtained from personal fitness bands, home monitoring systems etc. Data from these sources can be collected by virtue of either vendor provided SDK'S or RESTful API that allow third parties to gain access to user data controlled by certain permissions for the sake of user privacy. [8].

III. **Genomic Data**: This consists of information related to the genes of an organism. This data requires large storage capacities and is instrumental in detecting, preventing and curing diseases which are genetic or caused by mutations in the genes. Analysis of genomic data helps scientists and medicine practitioners to accurately predict remedies and preventive methodologies for patients belonging to particular lifestyle, having similar genetic make-up or exposed to similar environmental conditions [9]. Over the years, the cost and time to generate and store genomic data has been reduced drastically, catering to development of genomic healthcare analytics [10].

These data sources can also be classified based on the format of data they produce:

I. **Structured data**: Data that follows a well defined format. In the healthcare domain, such data originates from various sources such as laboratory results, patient admission history, billing information etc [11].

II. **Semi-structured data**: Data that is minimally structured but is nevertheless pretty much readable (by virtue of tags, markers etc). In the health care domain such type of data primarily originates from sensors employed to monitor a patient's behavior [11].

III. **Unstructured data:** Data that has no well defined structure, no additional tags or markers. Sources of such data include medical prescriptions written in human languages, clinical letters, biomedical literature, discharge summaries and so on.

Due to the diversity of data sources and the various forms in which they generate data (structured, unstructured, semi-structured), efficient and insightful analytics of healthcare data is a daunting task demanding the use of intelligent analytical platforms [11].

2. Intelligent Platforms

Intelligent platforms are autonomous or semi-autonomous systems that are capable of examining data and deriving suitable conclusions from it. They are a combination of techniques like data mining, pattern mining, machine learning, forecasting, visualization, semantic and sentiment analysis etc. The main focus of intelligent platforms is to extract information from huge volumes of data, thus help in predicting the future trends and making correct decisions.

In the healthcare domain, the already available data can help the patients and the doctors to predict and prevent the diseases in a better way.

Watson:

Watson is IBM's own supercomputer combining the power of Artificial Intelligence and analytics to deliver a platform that serves as an optimal question answering machine. Due to its high processing rate (approximately 80 teraflops) and access to a wide data store of over 200 million pages of information, Watson can perform analysis and text mining on huge volumes of data [9]. Delivered through the cloud, it can understand a variety of complex questions disguised as natural language and give well established, proof based answers. Talking about the more specific application of Watson in healthcare, in April 2015, IBM set up a Watson healthcare cloud platform aimed at driving innovation amongst doctors and researchers by suggesting actionable insights based on its high cognitive computing capability.

Since the late 90s, the IBM R&D teams have been working on Watson technologies for healthcare. In order to help oncologists, Watson for Oncology has been designed. Due to it's all in one unique combination of Natural Language Processing, Hypothesis Evaluation and Dynamic learning, Watson is a real asset in the healthcare domain [12]. Some of its features are summarized as below:

TABLE 1: SUMMARY OF WATSON'S FEATURES

Natural Language Processing	This helps Watson to move beyond the confines of structured data and understand complexities of the remaining unstructured and semi structured data.
Hypothesis Generation and Evaluation	This helps Watson to analyze and fish out responses based on relevant evidence and move towards a more probabilistic system that evolves with its users.
Dynamic Learning	This helps Watson to improve its accuracy based on feedback obtained during user interaction
Visual Recognition	With the help of deep learning algorithms, Watsonis able to identify objects, scenes or faces in the uploaded images on basis of which one can easily train a classified as per requirements with minimal lines of code.
Powerful Visualization	Apart from the processing powerhouse that Watson possesses, it also provides some powerful one click analysis that renders insightful dashboards, bridging the gap between advanced analytics and accessibility.

The next section covers the various applications of applying such powerful analytics on the rich and diverse health care data [13].

3. Applications

Healthcare analytics helps the physicians, doctors and patients in a number of ways. Following are the areas where analytics can improve the quality of treatments and diagnosis.

I. **Early Diagnosis** :Predictive Data Analysis on data obtained from various sources such as sensors, clinical trials etc can help physicians understand an array of different healthcare parameters and help in detecting diseases in their premature stage.[2]

II. **Medical profile matching:**This is especially useful in clinical trials. Current processes of trial matching require clinical coordinators to manuallysift through various patient records in an effort to match a patient's health profile with the requirements of the study in context. Such studies however, have 46 requirements on an average and thus render this manual matching process very tedious. Modern day health care analytics can help out here by analyzing various clinical trial databases with respect to a patients data and suggest his/her eligibility for the particular study [5].

III. **Diagnosis from therapeutic images**: Reports generated from X-RAY, MRI, Ultrasound etc are almost completely based on the doctor's interpretation of the images generated during these processes. Analytics however can aid this interpretation, by running the generated images through its system pre trained on past interpretations by various other doctors, and suggesting evidence based answers [2].

IV. **Drug Recommendation:** Every year, thousands of people die due to medication errors caused by doctors who write prescriptions solely based on their experience. This experience however, is very limited sometimes. Data Mining and recommender systems provide a means to explore knowledge from past diagnosis records and help doctors prescribe medication more correctly [10] [14].

IV. CHALLENGES

Every system comes with its own benefits and challenges. Following are the major challenges involved in the development and widespread use of the healthcare analytical systems.

I. **Incompatible Data Systems**: The healthcare data is collected via different sensors and electronic devices. It is a tedious task to convert the data of different types into a homogenous form.

II. **Patient Confidentiality Issues** - Leakage of health related information of patients is a major issue. There are few governance policies that are necessary to be followed in order to gain the trust of the patients in the healthcare analytical systems.

III. **Large Investment Costs** - Storage and processing of huge amount of data on premises involves heavy initialinvestments. However, the emergence and development of cloud technologies is a promising solution to reduce the investment costs.

V. SUMMARY

Average human lifespan is increasing day-by-day, thus challenging the present methods of diagnosis and treatment delivery. Healthcare industry can be considered to be one of the promising areas where data mining and analytics can provide large benefits. The healthcare domain is ever expanding. The breakthroughs in the field of medical research are opening up new avenues towards improving and understanding this field. Given the complex nature of the human body, there are cases when similar symptoms can lead to different diseases. Having said this, from a technical standpoint, there is never an exhaustive set of parameters on basis of which such technical systems can be modeled. The ratio between the number of patients and the healthcare utility providers is increasing. Given this scenario, specialized treatments along with quick diagnosis can help the patients recover quickly. Real time functioning and enhancements in the way healthcare facilities are provided can lead to physical, mental and emotional well being of the entire population.

VI. REFERENCES

[1] Liang, Ting, and Yong J. Yuan. "Wearable medical monitoring systems based on wireless networks: A review." *IEEE Sensors Journal* 16.23 (2016): 8186-8199.

[2] Thara, D. K., et al. "Impact of big data in healthcare: A survey." *Contemporary Computing and Informatics (IC3I), 2016 2nd International Conference on*. IEEE, 2016.

[3] J.Manyika,M. Chui, B. Brown et al., Big Data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, 2011.

[4] Bao, Youjun, and Xiaohong Jiang. "An intelligent medicine recommender system framework." *Industrial Electronics and Applications (ICIEA), 2016 IEEE 11th Conference on*. IEEE, 2016.

[5] Augustine, D. Peter. "Leveraging big data analytics and hadoop in developing India's healthcare services." *International Journal of Computer Applications* 89.16 (2014): 44-50.

[6] Archenaa, J., and EA Mary Anita. "A survey of big data analytics in healthcare and government." *Procedia Computer Science* 50 (2015): 408-413.

[7] M. N. Ahmed, A. S. Toor, K. O'Neil and D. Friedland, "Cognitive Computing and the Future of Health Care Cognitive Computing and the Future of Healthcare: The Cognitive Power of IBM Watson Has the Potential to Transform Global Personalized Medicine," in *IEEE Pulse*, vol. 8, no. 3, pp. 4-9, May-June 2017. doi: 10.1109/MPUL.2017.2678098

[8] De Arriba-Pérez, Francisco, Manuel Caeiro-Rodríguez, and Juan M. Santos-Gago. "Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios." Ed. Leonhard M. Reindl. *Sensors (Basel, Switzerland)* 16.9 (2016): 1538. *PMC*. Web. 17 Sept. 2018.

[9]Vassy J.L., Korf B.R., Green R.C. How to know when physicians are ready for genomic medicine. Sci. Transl. Med. 2015;7:287fs219. doi: 10.1126/scitranslmed.aaa2401. [PMC free article] [PubMed] [Cross Ref]

[10] Auffray, Charles, Zhu Chen, and Leroy Hood. "Systems medicine: the future of medical genomics and healthcare." *Genome medicine* 1.1 (2009): 2.

[11] Palanisamy, Venketesh, and Ramkumar Thirunavukarasu. "Implications of Big Data Analytics in developing Healthcare Frameworks–A review." *Journal of King Saud University-Computer and Information Sciences* (2017).

[12] High, Rob. The Era of Cognitive Systems: An Inside Look at IBM Watson and How It Works. REDP-4955-00. December 2012. www.redbooks.ibm.com/abstracts/redp4955.html?open. Accessed September 5, 2018.

[13] Ward, Michael J., Keith A. Marsolo, and Craig M. Froehle. "Applications of business analytics in healthcare." *Business horizons* 57.5 (2014): 571-582.