Big Data: A Challenging Opportunity for Biomedical Informatics

Mudassir Khan
Department of Computer Science, College of Science & Arts, King Khalid University, Saudi Arabia
mkmiyob@kku.edu.sa

Abstract:
In the current technological world big data technologies are being used in bio-informatics research and healthcare. The huge amount of clinical data have been generated and collected at an unoccupied speed and scale. For example, the number of sequencing technologies in the new era producing the trillions of DNA sequence data per day, and the different applications of EHRs- Electronic health records are specifying huge amount of patient data. The amount of processing and analyzing healthcare data is about to decrease dramatically with the help of available technologies. The Big data applications provide new opportunities to enhance new knowledge and establish different type methods to improve the quality of existing healthcare system. The objective of the paper is to evaluate the applications of analytics of Big Data in the biomedicine and healthcare field and the associated outcomes.

Keywords: Big data; data-driven application; data analytics; map-reduce; medical care, Tools

1. Introduction:

Big data is the novel hype in biomedicine and healthcare care. The data storage and processing has become cheap, fast, and easy. It is therefore important to understand the reason why big data are assuming a crucial role for the biomedical informatics community. The capability of handling big data is becoming an enabler to carry out unprecedented research studies and to implement new models of healthcare delivery. Therefore, it is first necessary to deeply understand the four elements that constitute big data, namely Volume, Variety, Velocity, and Veracity, and their meaning in practice [1]. Big data refers to datasets whose size, complexity and dynamic nature are beyond the scope of traditional data collection and analysis methods. The potential benefits to critical care are significant, with faster progress in improving health and better value for money. Although not replacing medical trials, big data can improve their design and advance the field of precision medicine. However, there are limitations to analyzing big data using observational methods. In addition, there are ethical concerns regarding maintaining confidentiality of patients who contribute to these datasets [2]. Then, it is mandatory to understand where big data are present, and where they can be beneficially collected. Huge quantity of biological and clinical data have been produced and collected at an unprecedented rate and scale. For example, the new era of sequencing technologies authorizes the processing of trillions of DNA sequence data per day, and the application of electronic health records is documenting large amounts of patient data. The cost of acquiring and analyzing biomedical data is expected to decrease dramatically with the help of technology upgrades, such as the emergence of new sequencing machines, the development of novel hardware and software for parallel computing, and the extensive expansion of electronic health records. Big data applications present new opportunities to discover new knowledge and create novel methods to improve the quality of health care. The application of big data in health care is a fast-growing field, with many new discoveries and methodologies published in the last five years. In this paper, we review and discuss big data application in four major biomedical sub disciplines: (1) bioinformatics, (2) clinical informatics, (3) imaging informatics, and (4) public health informatics [3].There are research fields, such as translational bioinformatics, which need to rely on big data technologies to withstand the shock wave of data that is generated every day. Other areas, ranging from epidemiology to medical care, can benefit from the exploitation of the large amounts of data that are nowadays available, from personal monitoring to primary care. However, building big data-enabled systems carries on relevant implications in terms of reproducibility of research studies and management of privacy and data access; proper actions should be taken to deal with these issues. An interesting consequence of the big data scenario is the availability of new software, methods, and tools, such as map-reduce, cloud computing, and concept drift machine learning algorithms, which will not only contribute to big data research, but may be beneficial in many biomedical informatics applications. The way forward with the big data opportunity will require properly applied engineering principles to design studies and applications, to avoid preconceptions or over-enthusiasms, to fully exploit the available technologies, and to improve data processing and data management regulations.

The aim of this Paper is to deal with the latest technological progress and focus on possible applications in medical care amenable to big-data research that can potentially improve patient care [4]. The first challenge in Big Data research will be the development of large, multicenter, and high-quality databases. These databases could be used to further investigate recent findings from mathematical models, developed in smaller datasets. Big data refers to datasets whose size, complexity and dynamic nature are beyond the scope of traditional data collection and analysis methods. The potential benefits to critical care are significant, with faster progress in improving health and better value for money. Although

not replacing clinical trials, big data can improve their design and advance the field of precision medicine. Such analyses can provide complementary information to medical trials of the standard type. As big data analyses become more popular, various statistical techniques for causality analysis in observational data are becoming more widely available. This is likely to be of benefit to medical science, but specific adaptations will have to be made according to the requirements of the applications. This paper also showed that: (1) integrating different sources of information enables clinicians to depict a new view of patient care processes that consider a patient's holistic health status, from genome to behavior; (2) the availability of novel mobile health technologies facilitates real-time data gathering with more accuracy; (3) the implementation of distributed platforms enables data archiving and analysis, which will further be developed for decision support; and (4) the inclusion of geographical and environmental information may further increase the ability to interpret gathered data and extract new knowledge [5,6].

Recently, the impact of big data analytics in developing novel solution for protein structure prediction problem received significant attention among research community. Proteins are the key functional units in living organism and serve as hormones, receptors, storage, enzymes and as transporters of particles in human body and responsible for biochemical reactions. The huge volume, unstructured format, and ever-changing nature of protein sequence residues found in the Protein Data Bank (PDB) prompt the need of big data intensive computational framework for protein structure prediction problem.

2. The Impact of Big Data in Medical Care and Health Care

In the business sector, the core value of big data has been effectively utilized for the identification of behavioral patterns of the consumers to develop innovative business services and solutions. In the healthcare sector, the implication of big data serves predictive analytical techniques and machine learning platforms [7] for the provision of sustainable solutions such as the implementation of treatment plans and personalized medical care. Some scientists [8] compared the healthcare big data with the big data generated from the business sector under different attributes and their values. They redefined the characteristics of the healthcare big data into three features namely Silo, Security, and Variety instead of Volume, Velocity and Variety. Silo represents the legacy database that contains public healthcare information maintained in stakeholders' premises such as hospitals. The security feature implies the extra care needed in maintaining healthcare data. The variety feature indicates the existence of healthcare data in many forms such as structured, unstructured and semi-structured. With the advent of big data analytics and its associated technologies, the healthcare domain witnessed pragmatic transformations at various stages from the perspective of involved stakeholders [9]. The impact of big data in healthcare results in identifying new data sources such as social media platforms, telemetric, wearable devices etc. in addition to the analysis of legacy sources that includes patient medical history, diagnostic and clinical trials data, drug effectiveness index etc. When the mixture of these data sources and analytics are coupled together, it provides a valuable source of information for healthcare researchers towards attaining novel healthcare solutions [10]. A typical patient centric healthcare ecosystem with its significant stakeholders and their diversified data sources (structured/semi-structured/unstructured) is perceived in Figure 1.

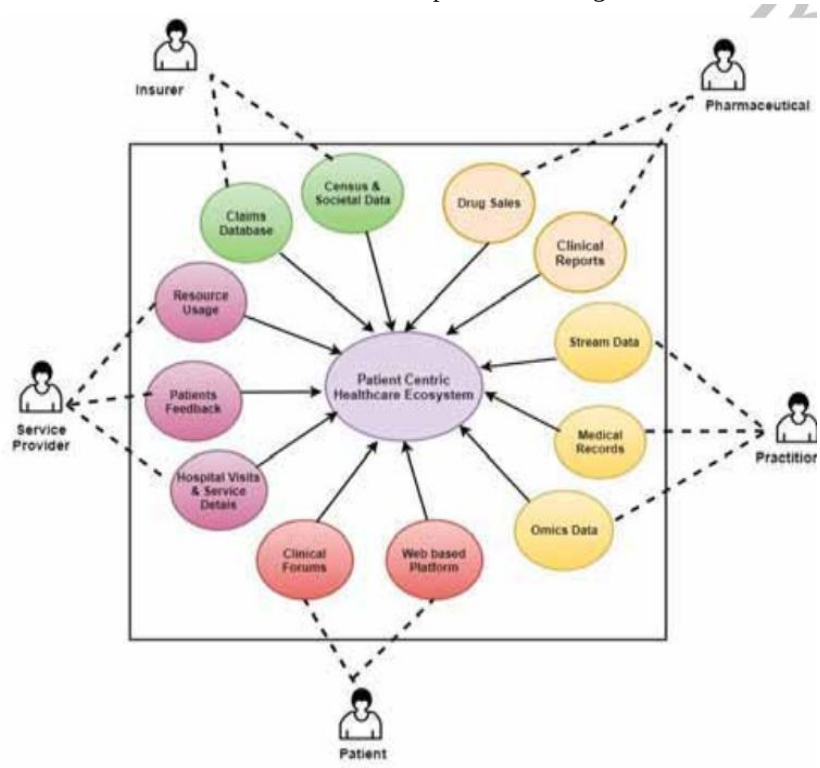

Fig. 1. A patient centric healthcare environment – from the big data perspective

When these stakeholders work collaboratively and share their data insights effectively, healthcare solutions would be offered in a cost-effective manner with improved personalized care for patients. By considering the importance of these stakeholders in building big data healthcare ecosystem, the next section throws insights on their perspective over the effective utilization of big data sources.

Table 1: Big Data Tools for Healthcare Ecosystem

Task	Tools	Merits and Applications
Data Unification	Pentaho	Tool for processing knowledge discovery from a scalable environment in a flexible manner
	Plantir	To serve decision makers for highlighting the process insights by overwhelming the treatment options and improving the quality of patient care
	Ayata	Used to perform prescriptive analytics from huge amount of data towards helping organizations for making smarter decisions
	Attunity	Advance data integration platform for performing automated data pipeline at a very fats rate
	Informatica	To provide enterprise data management solutions
	Jitterbit	Facilitates integration of data from saas based cloud services and on-premise applications in an intelligent manner
Searching and Processing	Appache Lucene	High performance, full-featured text search engine for analysing full text across variety of platforms
	Google Dremel	Map/Reduce based computations supported by hadoop for processing nested data with high accuracy
	Cloudera Impala	Executes low latency and high concurrency analytical queries on top of hadoop
	Dryad	Provides high performance distributed execution engine with good programming constructs
Machine Learning	Mahout	Offers distributed machine learning library for processing scalable mining algorithms.
	Skytree	Tool to performing machine learning and advanced analytics of massive data sets at high speed.
	BigML	Environment for offering solutions to big data use cases through predictive analysis.
Stream Data Processing	Appache Storm	Scalable and flexible real time computation system for processing massive amount of data.
	SQL Stream Blaze	Support creation of distributed streaming applications that gives data integration and analytics in real time.
	SP Lunk	Highly capable tool to collect and harness machine data.
	SAP Hana	Tool for in-memory computing of stream data for real time analytics.
Visual Data Analytics	Qlik	To discover clinical and operational data through visual analytics for discovering insights.
	Tableau	Provides faster, highly interactive dashboards to project extracted pattern.
	Jaspersoft	Supports interactive visual analytics at big rate by extracting data from different data sources.

The above-discussed tools are helpful to the researchers for deploying effective healthcare frameworks that facilitate end-to-end healthcare solutions by improving patient outcomes with the advent of big data.

3. Conclusion

In conclusion, employing efficient and streamlined analytics to big data will contribute to quick and accurate diagnosis, appropriate treatment, reduced costs and improved overall healthcare quality. Framework based solutions always cater to the comprehensive requirement of various stakeholders involved in the healthcare domain. With the impact of big data, healthcare domain was revamped and offer intensive solutions for handling diversified big data sources that range from patient health records to medical images. This project shows various research attempts in establishing medicine and healthcare frameworks and summarizes their significant outcomes. The summary of contributions by various researchers highlights the data source utilized, adopted analytical techniques and other features. At the end, the implication of various big data tools in developing healthcare framework is also extensively studied.

The extensive research efforts discussed so far in developing various healthcare frameworks shows the importance of adopting big data-based platforms and analytical techniques for reaping quality knowledge and disseminating it to the diversified healthcare stakeholders.

REFERENCES

[1] R. Bellazzi, Yearbook of Medical Informatics, May 22; 9:8-13 (2014)
[2] AB Docherty, NI Lone, Current Opinion to Critical Care, Oct;21(5):467-72 (2015).
[3] Jake Luo, Min Wu, Deepika Gopukumar, and Yiqing Zhao, Biomedical Informatics Insights, 8: 1–10 (2016).
[4] Molly E. McCue, and Annette M. McCoy, Frontier and Veterinary Science, 4: 194 (2017).
[5] Jake Luo, Min Wu, Deepika Gopukumar, Yiqing Zhao, Biomedical Informatics Insights (2016):8, 1-10.
[6] Isaac Cano, AkosTenyi, Emili Vela, Felip Miralles, Josep Roca, Current Opinion in Systems Biology, 3: 36-42, (2017).
[7] Al-Jarrah, O.Y., Yoo, P.D., Muhaidat, S., Karagiannidis, G.K., Taha, K., Big Data Res. 2, 87–93, (2015). doi: 10.1016/j.bdr.2015.04.001
[8] Jee, K., Kim, G.H., Healthc. Inform. Res. 19, 79–85 (2013) doi:10.4258/hir.2013.19.2.79
[9] Wang, L., Alexander, C.A., Am. Med. J. 6, 1–8 (2015). doi:10.3844/amjsp.2015.1.8
[10] Zhou, L., Pan, S., Wang, J., Vasilakos, A. V., Neurocomputing 237, 350–361 (2017). DOI: 10.1016/j.neucom.2017.01.026
[11] Bharti Karla, an Inspection on Big Data Computing, IJESR/ICRIT 2018/ Special Issue/Article No-52/326-329.
[12] Mudassir Khan, Big Data Analytics Evaluation, (IJERCSE), vol-1. Issue-2, Feb. 2018,
[13] Mudassir Khan, Mohd Dilshad Ansari "Security and Privacy Issue of Big Data over the Cloud Computing: A Comprehensive Analysis", (IJRTE-Scopus Indexed), Vol. 7, Issue 6s, March 2019, 413-417