International Journal of Chemical and Biochemical Sciences
(ISSN 2226-9614)
Journal Home page: www.iscientific.org/Journal.html

Application of Big Data in health care

Haroon Ur Rashid[1], Fatma Hussain*[2] and Khalid Masood[1]

[1]Department of Computer Science, Lahore Garrison University, Lahore, Pakistan and [2]Department of Biochemistry, Faculty of Sciences, University of Agriculture, Faisalabad

Abstract

In this article an overview of big data applications in health care is presented. It summarizes sources and characteristics of big data. In addition, an update of big data uses in milieu of disease diagnosis, progression, prevention, management, treatment strategies are also highlighted. Now a days, generation of data is surprising high and it will keep on expanding beyond all limits. Academic and information technology sectors have focused greatly on big data. Predictive analytical techniques like regression, computational statistics, simulation modeling and machine learning algorithms are used with a few modifications for big data. Next-generation sequencing (Omics), data models and clinical records are integrated for precision medicine, drafting policies and drug development. Harnessing the big data using machine learning techniques enhances existing knowledge, envisages poly-pharmacology, supports pharmacovigilance and predicts future directions, thus benefiting all stakeholders in health care. Open-innovation paradigm has created new alliances between industry and academia to promote research and development by data sharing.

Key words: Electronic Health Records, Medical Imaging, Omics, Precision Medicine and Pharmaceutical Industry

Full length article **Corresponding Author*, e-mail: fatmauaf@yahoo.com; Tel: +92 41 9200161-170, Ext. 3309

1. Introduction

Nowadays health care system not only provides guidance, support, management and treatment but also combat numerous ailments. Being composed of public and private segments, it is challenged increasingly by the outbreak of multifactorial pandemics and growing number of patients. To encounter such disquiets, health industry is striving to boost its amenities. Improved service output mainly depends on comprehensive information assembled from all potential sources. Now a days, generation of data is surprising high and it will keep on expanding beyond all precincts [1]. Data is generated from numerous sources: clinical trials, population data, patient's history, self-reported and physician-diagnosed symptoms, self-medication, life style, physician's prescriptions, dietary habits, laboratory test reports, scheduled and emergency appointments with doctor, treatment/recovery patterns and information about doctors, nurses and family caregivers. Collection, transfer, storage, calculation and examination of large size data is an immensely laborious work. Highly diversified, fragmented and unorganized health system databases are problematic during transfer, processing and examination that can lead to wrong or inconsistent conclusions [2].

Amazingly, the capacity to handle data is growing exponentially owing to a new concept referred to as "Big Data". That is, massive data sets that can yield surprising insights when analyzed [1]. Big data signifies large size, extremely complicated and vibrant collection of information that cannot be studied by conventional methods [3]. Three main characteristics of big data are volume (data quantity), variety (data type) and velocity (data speed). Simple elucidation can be "big data can collect, save and scrutinize large size of diverse data (organized, raw and dis-organized) at high speed" [4]. There are certain other factors that should also be deliberated as depicted in table 1.

Table 1: Characteristics of big data

Volume	Multi-domain data, user device data, geo-location data
Velocity	Stream computing, batch and real-time algorithm
Value	Modeling, prediction
Variety	Crowd sensing, heterogeneous sensors
Viability	Selection, relevance, relation
Veracity	Quality, uncertainty, security

Review of literature revealed that major contributions in health related big data research came from United States, China, United Kingdom and Germany [5].

Academic and information technology sectors have concentrated greatly on big data. Rapid production and assembly of data excels beyond the permissible limits in current digitalized domain. More than 2 billion people have access to internet and about 5 billion are mobile phone users. According to an estimate, internet will link 5 billion devices by the year 2020, with 44 times more data creation than that in 2009 [6].

Application of big data can lay a foundation for a comprehensive picture of past, present and future health issues. Predictive analytical techniques like regression, computational statistics, simulation modeling and machine learning algorithms are used along-with big data. However, in recent era of digitalization, collection of massive data from social networking sites such as facebook, twitter, instagram, linkedin, flicker, pharmaceutical companies webpages, blogs, medical/clinical websites, and online journals has further broaden the horizons of computational data science [7]. In this review, role of big data in health care is discussed. Big data applies to five major aspects of health care.

2. Administration and Delivery

Analytics tools suitable for administrative (consumable and non-consumable facilities) and delivery (services rendered to patients) purposes are preferred by health care systems. Administrative and delivery (provision) data can be either combined or analyzed separately. Various data models can be used to solve problems arising from inaccurate data collection that can affect analysis and conclusions. As healthcare data is the most diversified type, many organizations have to develop their own customized software [8]. Although implementation of analytic practices faces legal and ethical issues, text and electronic health records are (figure 1) used to extract information in using big data approaches [7]. Computerized medical records are consistently maintained and are reported as a main source for data collection [9].

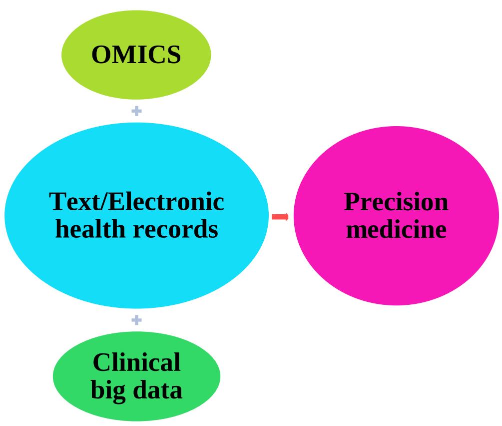

Figure 1: Integration of data for precision medicine

Huge data obtained from case-control studies and population trials is quite informative for hospital administrators to meet the requirements of patients and physicians. It can be used to forecast pathogenesis and thereby, to arrange alternative therapeutic options at the site. Similarly, big data techniques can be used to estimate infrastructure up-gradation to meet flourishing healthcare business [10]. Usage of big data has coined term "business intelligence" among health care providers owing to enormous benefits expected in analyzing big data. From financial viewpoint, big data inferences cut healthcare costs by reducing undesirable purchases, treatments, re-admissions, thereby reducing patient's burden and improving quality of life [11].

3. Diagnosis and Progression

Health data highlights timely identification of disease, progression, treatment options, prevention choices and thus provides a complete cure. Health oriented computational research focuses on three aspects: disease aspect (type, onset, progress), technical aspect (data mining, machine learning) and health service provision (prevention, treatment, management) [5]. The vast amount of information in the form of electronic medical records helps to develop novel models of disease progression to ultimately predict health outcomes. Therapeutic interventions are revised regularly depending upon disease progression patterns resulting from data models [9]. Variable genetic makeup and environmental exposure responsible for individual differences has led to a new concept of precision medicine. Merger of high profile OMICS tools (table 2) and big data deliver unique pathological conceptions that are transformed into personalized deterrence and medications [12].

Table 2: "OMICS" approach for health care

Phenomics	Structural, functional traits (phenotype) by medical imaging, sensors, monitors
Metabolomics	Metabolic profiling by mass spectrometry, nuclear magnetic resonance spectrometry
Proteomics	Protein sequencing by chromatography, mass spectrometry
Transcriptomics	mRNA sequencing/expression by microarray
Epigenomics/genomics	DNA sequencing

4. Prevention, Management and Treatment Support

Healthcare records are accessed to appraise preventive strategies, management skills and treatment efficiencies [13]. Although big data applications cannot replace clinical trials, however, it can promote precision medicine process. As conventional computational techniques are replaced by big data, improved medical care facilities with minimal costs are expected [3]. Next-generation sequencing (NGS) is increasing employed in molecular biochemistry. NGS mostly deal with limited-volume sample size that demands awareness regarding downstream assay and computation tools. Computational mechanisms of the NGS (figure 2) transform raw sequence data into clinically actionable genomic information [14]. Genome-wide association studies reflect the importance of personalized diagnosis and treatment. Many unseen patterns, unidentified correlations and undetected links in genomics are exposed by big data analytics [15].

Mining the "Big Data" using innovative machine-learning methods are used to predict polypharmacology and eliminate hindrance in phenotype screening. Polypharmacy deals with medicines that attacks numerous delicate sites interlinked in a physiological network. These medicines are more potent with limited disadvantages as compared to those that target one site at a time. Recently, open-innovation paradigm has created new alliance between industry and academia to promote research and development by data sharing [16]. Quality and efficiency of health care delivery are re-defined by big data applications (table 3). Big data applications refine the attribute and efficacy of health care delivery. In order to take advantage of the promising opportunities of big data technologies, complete awareness and demands of all stakeholders should be recognized [17].

5. Drafting Healthcare Policies

Various data base sources provide pseudo real-time access to patient across laboratories and manufacturers. The data is used for the development, implementation and revision of health care policies [18]. However, variable data access policies existing across the globe either endorse or restrain research opportunities. Policy makers, data owners, data suppliers and data users should draft such policies that support data access and utilization from all sources [19]. Massive quantities of health care data can become knowledge generating engine to formulate health guidelines [10]. Successful policies address strong aims and debate obstacles that are possible by means of innovative data sharing [20].

6. Pharmaceutical Sector

Pharmaco-epidemiological research is warranted for the introduction of new drugs. Enormous growth of data resources has made such investigations practically achievable [21]. Big data mining is implemented in pharmacovigilance of drugs for many diseases owing to increasing patient size. Big data tools detect adverse side effects that will be minimized in future trials. Unfortunately, pharmacovigilance of drugs for orphan diseases face obstacles such as small patient population size [22]. Big data is used to validate claims made in the pharmaceutical approval process to foresee the functioning of medicines and devices (figure 3) [13]. Increasing raw material prices have declined productivity of pharmaceutical industries that are already under pressure from doctors to improve product quality. Pharmaceutical companies have entered an economical era. Industry has gained its revival due to continual increase in big data initiatives that ensure cost-effective production with excellent quality and quantity [23].

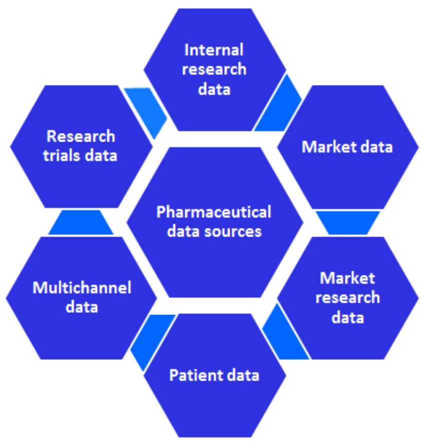

Figure 3: Big data sources in pharmaceutical industry

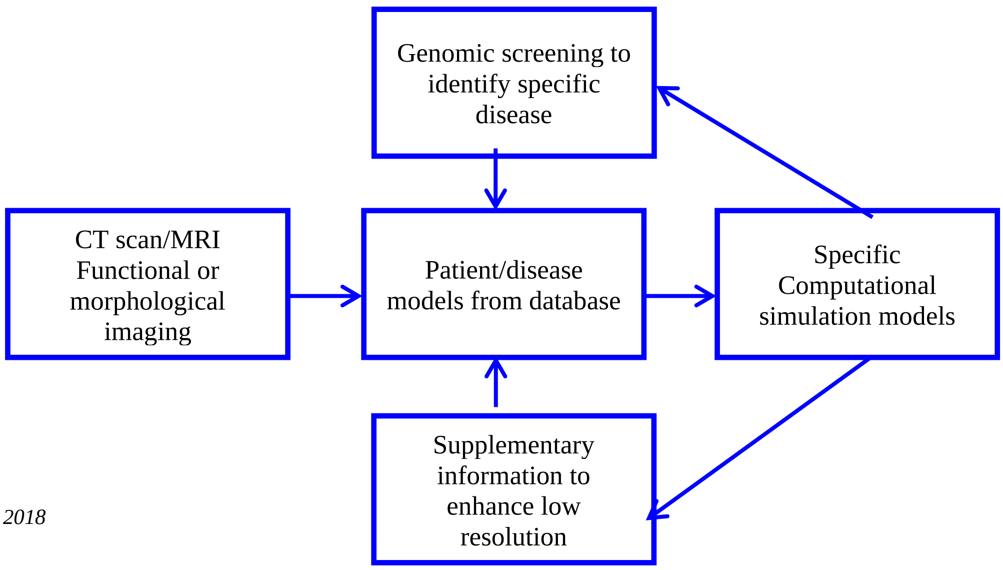

Figure 2: Combination of imaging, modeling and real-time sensing for disease diagnosis, progression management, and therapeutic planning.

Table 3: Big data in health care - Selected tools and platforms

Omic data processing	GMAP, STAR, GATK, RSEM, Trans-ABySS, MACS, SISSRs
Omic biomarker identification	SNPassoc, VAT, PLINK, edgeR, MATS, DBChIP, QDMR
Omic data modeling	WGCNA, CODENSE, MEMo, CellDesigner, NetLogo, BoolNet, Snoopy
Big data Analytics	Apache Hadoop, IBM InfoSphere Platform, Apache Hive, Apache Sqoop, Apache Pig, Apache Ambari, Apache Spark Streaming, Zookeeper

7. Conclusions

Large amount of data flowing in health care system can be pooled together and investigated by big data analytics to support all stakeholders. Big data tools once implemented persistently may uplift the whole scenario, reviving the health care system. Big data has resulted in business intelligence that is of fundamental importance to health care professionals. Optimum analysis of big data can cut health care costs drastically and can improve the quality and efficiency in dealing with health related problems.

References

[1] M.B. Hoy. (2014). Big data: An introduction for librarians. Medical Reference Services Quarterly. 33(3): 320-326.

[2] J.A. Patel and P. Sharma. (2014). Big data for better health planning In IEEE International Conference on Advances in Engineering and Technology Research, 1-5.

[3] A.B. Docherty and N.I. Lone. (2015). Exploiting big data for critical care research. Current Opinions in Critical Care. 21(5): 467-472.

[4] S. Melani. (2015). Philosophy of big data expanding the human-data relation with big data science services. IEEE First International Conference on Big Data Computing Service and Applications, 468-477.

[5] D. Gu, J. Li, X. Li and C. Liang. (2017). Visualizing the knowledge structure and evolution of big data research in healthcare informatics. International Journal of Medical Informatics. 98: 22-32.

[6] N. Khan, I. Yaqoob, I.A. Hashem, Z. Inayat, W.K. Ali, M. Alam, M. Shiraz and A. Gani. (2014). Big data: Survey, technologies, opportunities, and challenges. The Scientific World Journal. 2014.

[7] S.Rüping. (2015). Big data in medicine and healthcare. Bundesgesundheitsblatt-Gesundheitsforschung Gesundheitsschutz. 58(8): 794-798.

[8] S.R. Sukumar, R. Natarajan and R.K. Ferrell. (2015). Quality of big data in health care. International Journal of Health Care Quality Assurance. 28(6): 621-634.

[9] O. Arandjelović. (2015). Discovering hospital admission patterns using models learnt from electronic hospital records. Bioinformatics. 31(24): 3970-3976.

[10] H.M. Krumholz. (2014). Big data and new knowledge in medicine: The thinking, training and tools needed for a learning health system. Health Affairs. 33(7):1163-1170.

[11] C. Schaeffer, L. Booton, J. Halleck, J. Studeny and A. Coustasse. (2017). Big data management in US hospitals: Benefits and barriers. The Health Care Manager. 36(1): 87-95.

[12] K.H. Yu, S.N. Hart, R. Goldfeder, Q.C. Zhang, S.C. Parker and M. Snyder. (2017). Harnessing big data for precision medicine: Infrastructure and applications. In Pacific Symposium on Biocomputing. 635-639.

[13] T. Wasser, K. Haynes, J. Barron and M. Cziraky. (2015). Using big data to validate claims made in the pharmaceutical approval process. Journal of Medical Economics. 18(12): 1013-1019.

[14] S. Roy-Chowdhuri, S. Roy, S.E. Monaco, M.J. Routbort and L. Pantanowitz. (2017). Big data from small samples: Informatics of next-generation sequencing in cytopathology. Cancer Cytopathology. 125(4): 236-244.

[15] K.Y. He, D. Ge and M.M. He. (2017). Big data analytics for genomic medicine. International Journal of Molecular Sciences. 18(2): 412.

[16] I.V. Tetko, O. Engkvist, U. Koch, J.L. Reymond and H. Chen. (2016). BIGCHEM: challenges and opportunities for big data analysis in Chemistry. Molecular Informatics. 35(11-12): 615-621.

[17] S. Zillner, N. Lasierra, W. Faix and S. Neururer. (2014). User needs and requirements analysis for big data healthcare applications. In MIE. 657-661.

[18] K. Goossens, K. Van Uytfanghe, P.J. Twomey and L.M. Thienpont. (2015). Monitoring laboratory data across manufacturers and laboratories-A prerequisite to make big data work. Clinica Chimica Acta. 445: 12-18.

[19] J.A. Doshi, F.B. Hendrick, J.S. Graff and B.C. Stuart. (2016). Data, data everywhere, but access remains a big issue for researchers: A review of access policies for publicly funded patient-level health care data in the United States. eGEMS (Wash DC). 4(2):1204.

[20] A. Heitmueller, S. Henderson, W. Warburton, A. Elmagarmid, A.S. Pentland and A. Darzi. (2014). Developing public policy to advance the use of big data in health care. Health Affairs (Millwood). 33(9): 1523-1530.

[21] D. Macías Saint-Gerons, C. de la Fuente Honrubia, F. de Andrés Trelles and F. Catalá-López. (2017). Future perspective of pharmacoepidemiology in the era of "big data" and the expansion of information sources. Spanish Journal of Public Health. 90: e20010.

[22] J. Price. (2016). What can big data offer the pharmacovigilance of orphan drugs? Clinical Therapeutics. 38(12): 2533-2545.

[23] P. Tormay. (2015). Big Data in pharmaceutical R&D: Creating a sustainable R&D engine. Pharmaceutical Medicine. 29(2): 87-92.