Nidhi Raj Int. Journal of Engineering Research and Application *www.ijera.com*
ISSN : 2248-9622, Vol. 8, Issue5 (Part -I) April 2018, pp45-52

RESEARCH ARTICLE OPEN ACCESS

Big Data Science and Its Applications in Biomedical Research and Healthcare: A Review

Nidhi Raj*, Aabriti Karki**, Madhu BR***, Manjunath CR****
B.Tech. (VIII-Sem), Department of Computer Science & Engineering, School of Engineering & Technology, Jain University, Bangalore, India.
****B.Tech. (VIII-Sem), Department of Computer Science & Engineering, School of Engineering & Technology, Jain University, Bangalore, India.***
****- Associate Professor, Department of Computer Science& Engineering, School of Engineering & Technology, Jain University, Bangalore, India.*
*****- Associate Professor, Department of Computer Science& Engineering, School of Engineering & Technology, Jain University, Bangalore, India.*
Corresponding author:Nidhi Raj

ABSTRACT

There is a consensus among scientists that the analysis of Big Data in health care (such as electronic health records, patient reported outcomes or in motion data) can improve clinical research and the quality of care provided to patients. Big Data mainly deals with the storage and processing of large scale and complex structure data sets for which the traditional methods prove to be incapable. They show a slow responsiveness and lack of scalability, performance and accuracy. The paper provides a systematic review of recent progress and advances in Big Data science, healthcare and the human genome research one of the most promising medical and the healthcare domains. Further the paper includes some of the major challenges and opportunities with a focus on the upcoming and promising areas of medical research: image, signal, and genomics-based analytics to improve big data applications in healthcare in the summarized form.

Keywords: Big Data, Challenges and Opportunities. Electronic Health Records, ePROs, Heterogeneous Datasets.

Date of Submission: 18-04-2018 Date of acceptance: 03-05-2018

I. INTRODUCTION:

In the last decade a Big Data revolution is under way in the health care sector: as a result, the new challenge for world healthcare is to take advantage of these huge amounts of data destructured among them [1,2]. Various attempts at defining Big Data essentially characterize it as a collection of data elements whose size, speed, type, and/or complexity require one to seek, adopt, and invent new hardware and software mechanisms in order to successfully store, analyze, and visualize the data [3,4]. The current trend toward digitizing healthcare workflows and moving to electronic patient records has seen a paradigm shift in the healthcare industry. The quantity of clinical data that are available electronically will be then dramatically increased in terms of complexity, diversity and timeliness, resulting what is known as Big Data [5].

In this context, scientists have been focused on the improvement of public health policies, clinical research and the care provided to patients through the analysis of health-related big datasets [6]. There are several large sources for Big Data in health care: genomics, electronic health record (HER), medical monitoring devices, patients, hospitals, providers and organizations and health-related mobile phone apps. More medical discoveries and new technologies such capturing devices, novel sensors, and wear-able technology have contributed to additional data sources [7].

It is a multidisciplinary field involving various sciences such as biomedical, medical, nursing, information technology, computer science, and statistics. Using Information and Communication Technologies (ICTs), health informatics collects and analyzes the information from all healthcare domains to predict patients' health status. Recently the proliferation of wear-able medical devices has significantly added fuel to the healthcare data. Those devices are able to continuously monitor a series of physiological information, such as biopotential, heart rate, blood pressure, and so forth [8]. The enormous scale and diversity of temporal-spatial healthcare data have created unprecedented opportunities for data assimilation, correlation, and statistical analysis [9]. However, the advent of modern ubiquitous and social networking technologies has given rise to new forms of patient generated data, such as electronic patient reported outcomes (ePROs); physiological and psychometric data (especially real-time data collecteddirectly through sensor devices); and data generated online (for example, patients' comments and posts in online social networking tools).

The analytical analysis of big volumes of unstructured, semi-structured and structured heterogeneous data sets is an important predicate of modern science and one of the key drivers of innovation and economic growth in developed countries. Despite the fact that data has been used for a long time in the sciences to understand natural phenomena (especially through simulations and algorithms) in different domains [10]. The modern term Big Data refers to data whose heterogeneity and complexity require new architectures, techniques, algorithms, and analytics to manage it and extract value and hidden meaning from it [11].

Big Data also arises with many challenges, such as difficulties in data capture, data storage, and data analysis and data visualization the characteristics of Big Data require powerful and novel technologies to extract useful information and enable more broad-based health-care solutions. In most of the cases reported, I found multiple technologies that were used together, such as artificial intelligence (AI), along with Hadoop and data mining tools. Parallel computing is one of the fundamental infrastructures is capable of executing algorithm tasks simultaneously on a cluster of machines or supercomputers. In recent years, novel parallel computing models, such as MapReduce have been proposed for a new Big Data infrastructure. More recently, an open-source MapReduce package called Hadoop was released by Apache for distributed data management. The Hadoop Distributed File System (HDFS) supports concurrent data access to clustered machines. Hadoop-based services can also be viewed as cloud-computing platforms, which allow for centralized data storage as well as re-mote access across the Internet. [12]. Scientists define Big Data by the following four main characteristics (called the 4 Vs): volume (amount or scale of data), velocity (data in motion and more specifically to the speed at which data is created, processed and analyzed). Variety (managing the complexity and heterogeneity of multiple data sets), veracity (refers to data uncertainty and to the level of reliability/quality associated with certain types of data) [13] and some experts added more Vs :Vision (a purpose), Verification (processed data conformed to some specifications), Validation (the purpose is fulfilled) and finally Value- Refers to techniques of deriving value from data [14,15]. There is an intrinsic value that the data may possess and must discover for analysis. [16]. The purpose is to perform a systematic review of the literature in order to deter-mine the extent to which Big Data applications in health care systems have managed to improve patient experiences and clinicians' behavior as well as the quality of care provided to patients.

II. METHODOLOGY:

While several different methodologies are being developed in this rapidly emerging discipline, here I outline one that is practical and hands-on. Which shows the main stages of the methodology [17].The first step involves identification of problem and then adopts appropriate approach and strategy to select relevant papers for solving the problem .The second step involves search for four bibliographic databases to find related re-search articles: PubMed, Science Direct, Springer, and Scopus. In searching these databases I used the main keywords"Big Data,""Healthcare," "Genomics research" and "Electronic Health Records" Then selected relevant papers based on following criteria. At the third step the paper was published within the last five years i.e.2014–2018 discussing the design and the use of a Big Data application in healthcare domains. At the fourth one the paper reported a method for processing Big Data and discussed the performance of the method. In the last step the paper evaluated the performance of a new and existing Big Data application [18]. All potentially related papers were collected by reviewing the title and abstract. This initial search resulted in 115 papers from 2014 to 2018. Finally, I screened the papers based on the abovementioned criteria and subsequently selected 41 candidate papers.

III. HEALTHCARE STAKEHOLDERS:

The different interested independent person / organization in healthcare industry have dissimilar expected incentives and hopes from Big Data which can be summarized as follows:

3.1 Patients community: Patients want to avail a broad range of healthcare services at affordable cost with personalized recommendations [19]. In addition to physician's clinical diagnosis, they have an opportunity to gain more medical knowledge through digital platforms: social media networks, clinical forums etc. These big data sources enable the patients to connect with similar people for gaining information such as dis-ease symptoms, side-effects, hospitalization, and drug information [20]. Everyone wants customer-friendly service, one-stop shopping, and better coordination of care between themselves, caregivers and various providers, with an ultimate goal of error-free effective care. [21]

3.2 Healthcare Providers: Providers may recognize high risk population and act appropriately the massive amount of data generated from various phases of diagnosis and treatment plans of patients helps healthcare providers to identify the real insight about the progress of the treatments that are offered by them [22]. They want technology to be a transparent tool, not an encumbrance.

3.3 Pharma and Clinical Researchers: Pharmaceutical and clinical research Big Data helps to build predictive models for understanding the biological and drug processes that attribute to the high success rate in attaining effective drug designs[23]. Effective analysis of health data from diversified big data sources helps pharma companies to measure the outcome of designed drugs with smaller and shorter trials [24]. They want new tools to improve the quality and quantity of workflow – e.g., predictive modeling, statistical tools and algorithms that improve the de-sign and outcome of experiments and provide a better understanding of how to develop treatments.

3.4 Medical device: The strategic operators of the hospitals can also utilize the location aware-ness data to decide the co-location of various departments to optimize the use of expensive healthcare equipment. Companies, many of which have been collecting data for some time from hospital and home devices for safety monitoring and adverse event prediction, are beginning to wonder what to do with this data, and how to integrate it with old and new forms of personal data.[25] .

IV. IMAGING ANALYSIS:

Medical imaging data is a type of Big Data in medical research and genomics is a rapidly growing field derived from recent advancement in multimodal imaging data [26]. Imaging informatics is the study of methods for generating, managing, and representing imaging information in various biomedical applications. It is concerned with how medical images are ex-changed and analyzed throughout complex health-care systems. The goal of medical image analytics is to improve the interpretability of depicted contents. Many methods and frame-works have been developed for medical image processing. However, these methods are not necessarily applicable for big data applications. One of the frameworks developed for analyzing and transformation of very large datasets is Hadoop that employs MapReduce. MapReduce is a programming paradigm that provides scalability across many servers in a Hadoop cluster with a broad variety of real-world applications. However, it does not perform well with input-output intensive tasks [27]. Imaging informatics developed almost simultaneously with the advent of EHRs (Electronic Health Records) and the emergence of clinical informatics. Imaging can be classified in to three types they are: data storage and retrieval, data sharing, and data analysis.

4.1. Data storage and retrieval : Imaging informatics is predominantly used for improving the efficiency of image processing workflows, such as storage, retrieval, and interoperation. PACS (Picture Archiving and Communication Systems) are popular for delivering images to local display workstations, The Big Data technologies based on the implementation of cloud platforms with PACS, developed a massive Hadoop-based medical image retrieval system that extracted the characteristics of medical images using a Brushlet transform[28] and the MapReduce computation model and Hadoop Distributed File System (HDFS) storage model used to address the challenges of content-based image retrieval systems.

4.2. Data and workflow sharing. PACS primarily provide image data archiving and analysis workflow at single sites. Radiology groups operating under a disparate delivery model face significant challenges in a data-sharing infrastructure. Super PACS provides two approaches: (a) the federated approach and (b) the consolidated approach,

4.3. Data analysis. Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains. Requirements, Collection, Migration, Integration, Data Management, Communication & Visualization, Profiling, Data Auditetc, are common types of data analysis. Data analysis is closely linked to data visualization and data dis-semination. Seeking to overcome the challenges brought by large-scale data derived from pathological images, proposed Hadoop-GIS, an efficient and cost-effective parallel system [29].

V. SIGNAL GENERATION ANALYSIS:

Families of methods commonly referred to as DPA (disproportionality analysis) are currently the most widely used approach for automated ADR (Adverse Drug Reactions) signal detection in pharmaco vigilance [30]. A well-accepted database of all currently known ADRs does not exist at this time. For reference standard of drugs associated with each of the targeted ADRs is divided into the following two inclusion criteria: (i). Established- drugs confirmed to be causally related to the ADR: drug label warning, a Micromedex listing literature reviews of well-established ADRs, and other published reports having conclusive evidence such as laboratory data, clinical data, or a rechallenge/dechallenge study. and (ii). Plausible- drugs that have a high likelihood of being causative: analysis of one or more case reports mentioning the drug as a potential cause for a certain adverse event.

VI. GENOMIC ANALYSIS:

Genetic analysis is the overall process of studying and researching in fields of science that involve genetics and molecular biology. Ge-nomics is an interdisciplinary field of science focusing on the structure, function, evolution, map-ping, and editing of genomes. Genomics also involves the sequencing and analysis of genomes through uses of high through put DNA sequencing and bioinformatics to assemble and analyze the function and structure of entire genomes. Ge-nomic research areas can be broadly categorized into following way.

6.1. Molecular Analyses. Genetic analyses include molecular technologies such as PCR(The polymerase chain reaction is a biochemical technology in molecular biology to amplify a single or a few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence), DNA sequencing (DNA sequencing process is used to determine the order of nucleotide bases) and DNA microarrays (A DNA microarray is a collection of microscopic DNA spots attached to a solid surface. Scientists use it to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome), and cytogenetic (cytogenetic is a branch of genetics that is concerned which the study of the structure and function of the cell) methods

6.2. Functional genomics: is a field of molecular biology that attempts to make use of the vast wealth of data produced by genomic projects to describe gene (and protein) functions and interactions. Functional genomics attempts to answer questions about the function of DNA (Deoxyribose Nucleic Acid) at the levels of genes, RNA (Ribonucleic Acid) transcripts, and protein products. A key characteristic of functional genomics studies is their genome-wide approach to these questions, generally involving high-throughput methods rather than a more traditional “gene-by-gene” approach.

6.3. Structural genomics: seeks to describe the 3-dimensional structure of every protein encoded by a given genome. This genome-based approach allows for a high-throughput method of structure determination by a combination of experimental and modeling approaches.

6.4. Epigenomics: is the study of the complete set of epigenetic modifications on the genetic material of a cell, known as the epigenome. Epigenetic modifications are reversible modifications on a cell’s DNA that affect gene expression without altering the DNA sequence Two of the most characterized epigenetic modifications are DNA methylation and histone modification.

6.5. Metagenomics: is the study of metagenomes, geneticmaterial recovered directly from environmental samples. The broad field may also be referred to as environmental genomics, economics or community genomics.

Genomics research is the dream use case for Big Data technologies which, if unified, are likely to have a profoundly positive impact on mankind. The significance of genomic sequence data in the medical discoveries of the future will largely depend on our ability to process and analyze large genomic data sets, which continue to expand as the cost of sequencing decreases. Next-generation sequencing (NGS) technology enables genomic data acquisition in a short period of time. The role of Big Data techniques in bioinformatics applications is to provide data repositories, computing infrastructure, and efficient data manipulation tools for investigators to gather and analyze biological information. Genomic data generated by NGS technologies are a vital component in supporting genomic medicine, but the volume and complexity of the data raise challenges for its use in clinical practice. NGS involves breaking DNA into large amounts of segments. Each segment is called a ‘read’. Due to biases in sample processing, library preparation, sequencing-platform chemistry, and bioinformatics methods for genomic alignment and assembly of the reads, the distribution and/or length of reads across the genome can be uneven [31]. Therefore, some genomic regions are covered with more reads and others with fewer reads. For RNAseq, read depth is more often designated as number of millions of reads. Read alignment involves lining up the sequence reads to a reference sequence to allow comparison of sequence data from a sample sequenced with the reference genome [32]. A number of genomic alignment tools including Acharjya [33], Yang [34], and Oussous [35] have been developed on Big Data infrastructures.

VII. CHALLENGES OF HANDLING BIG DATA:

Millions of data points regarding tens of thousands of clinical elements within the electronic health records are available for EHR-based phenotyping. Like sequence data, it will also become a significant challenge to manage and manipulate the complete data of millions of individuals. The top challenges were issues of data structure, data standardization, data storage and transfers, security, regulatory compliance, inaccuracies in data, real-time analytics and managerial skills such as data governance [36]. At minimum, a Big Data analytics platform in healthcare must support the key functions necessary for processing the data. The criteria for platform evaluation may include availability, ability to manipulate at different levels of granularity, continuity, scalability, privacy and security enablement, ease of use, and quality assurance [37]. The major challenges in adopting Big Data analytics from enterprises are more managerial and cultural than associated with data and technology [38].Realtime big data analytics is a key requirement in healthcare. *Real-time analytics*is typically performed on data collected from sensors. In this situation, data change constantly,and rapid data analytics techniques are required to obtain an analytical result within a short period. The lag between high-quality data collection and processing system has to be addressed.The important managerial issues of ownership, governance and standards have to be considered. Health care data is rarely standardized, often fragmented, or generated in legacy IT systems with incompatible formats. It is difficult to determine the proper balance between protecting the patient's information and maintaining the integrity and usability of the data. Open access, integration, standardization of readable and useabledata is a challengeA vital challenge of incorporating genomic data into clinical practice is the lack of standards for generating NGS data, bioinformatics processing, data storage, and clinical decision support [39]

VIII. OPPORTUNITIES:

Big Data offers big opportunities for Retail, Financial, Web Companies. The top opportunities revealed were quality improvement, quality control and performance measurement, population management, early disease monitoring, data quality, structure, and accessibility, improved decision making, drug and medical device surveillance, routine clinical practice and research, and cost reduction. Despite the challenges that Big Data needs to overcome, the advanced analytics that are promised through big data offer tremendous opportunities for most stakeholders in the health care industry. If even some of the opportunities of Big Data are realized, they can radically change patient outcomes and the way decisions are made by providersand help solve some macro-level issues related to health care.[40]

IX. DISCUSSION:

In healthcare/medical field, large amount of information about patients' medical histories, symptomatology, diagnoses and responses to treatments and therapies is collected. To conduct pilot studies on incorporating genomic data into clinical care, a number of healthcare systems have developed bioinformatics infrastructures to process NGS data through group of databases supplementary to the EHRs [41]. Most of the infrastructures are locally developed and proprietary, but this is because these centers are among the first healthcare providers. However, there are still challenges with integrating genomic data into EHRs in clinical practice, including reliable bioinformatics systems/pipelines that translate raw genomic data into meaningful and actionable variants, the role of human curation in the interpretation of genetic variants, and the requirement for consistent standards to genomic and clinical data. A vital challenge of incorporating genomic data into clinical practice is the lack of standards for generating NGS data, bioinformatics processing, data storage, and clinical decision sup-port. Standards could promote interoperability in data quality. Obedience to standards would enable the routine use of genomic data in clinical care. However, it is challenging to build standards when NGS technology and bioinformatics tools are frequently evolving. Furthermore, approaches to clinical decision support differ among healthcare institutions. Appropriately integrating genomic data with EHRs for the discovery of clinically actionable variants can generate novel insights into disease mechanisms and provide better treatments. To improve our understanding on the nature of the disease from comprehensive EHRs, new methods such as ML (machine learning), NLP (natural-language-processing), and other artificial intelligence approaches are needed. However, not all patients are likely to benefit from the use of Big Data in healthcare due to our current knowledge gaps on how to extract useful information from large-scale genomic and clinical data and how to interpret discovered variants properly. In the meantime, targeted therapies are not yet available for many important genes, and regulatory issues need to be solved before some useful bioinformatics tools can be applied to clinical setting.

Apache Hadoop is a wellknown Big Data technology that has an important supporting community. It has been designed to avoid the low performance and the complexity encountered when processing and analyzing Big Data using traditional technologies. One main advantage of Hadoop is its capacity to rapidly process large data sets, thanks to its parallel clusters and distributed file system. In fact, unlike traditional technologies, Hadoop do not copy in memory the whole distant data to execute computations. In-stead, Hadoop executes tasks where data are stored. Thus, Hadoop relieves network and servers from a considerable communication load. Furthermore, there is a current trend towards further developing analytics and visualization technologies on top of the Hadoop platform, to enable better standardization of reporting and summarization of results. This is a problem which is not adequately addressed in the technology sector and is vital if the technology is to be widely embraced by diverse Industry sectors.

For Big Data healthcare systems, the Hadoop-Map Reduce framework is uniquely capable of storing a wide range of healthcare da-ta types including electronic medical records, genomic data, financial and claims data etc. and offers high scalability, reliability and availability than traditional Database Management Systems (DBMS).

X. CONCLUSION

We are currently in the era of emerging field of “Big Data science”, in which Big Data technology is being rapidly applied to biomedical research and healthcare fields. This systematic review provides some examples in which Big Data technology has played an important role in modern day healthcare revolution, as it has completely changed people’s view of health care activity. The increasing development of medical imaging and genomic data will certainly play an important role in clinical activities. Further, the review revealed both challenges and opportunities that Big Data science offers to the health fields. The main objective is to provide the healthcare services to all and to improve the accessibility of healthcare amenities to all the stratum of the society. The outcomes of this review will surely be helpful in the knowledge of the Big Data in healthcare in area of research, such as processing heterogeneous datasets, patient reported outcomes and in motion data.

Most of the above challenges can be future research topics. These future research topics can be: Aggregating and analyzing unstructured health care data, indexing and processing of continuously stream data, medical data confidentiality and interoperability and e-Infrastructures as persistent platforms for health care Big Data, etc.

REFERENCES

[1] J. Li ,F.Tao, Y. Cheng And L. Zhao ,”Big Data In Product Lifecycle Management”,Int J Adv Manuf Technol, 81, 667–684 D, 2015.

[2] L. Wang, G. Wang, And C. AnnAlexander,“Big Data And Visualization:Methods, Challenges And Technology Progress”, Digital Technologies,Vol.1, No.1, Pp 33-38, 2015 , Doi: 10.12691/Dt-1-1-7 [Pubmed]

[3] H. Wang, Z. Xu, And W. Pedrycz, “ An Overview On The Roles Of Fuzzy SetTechniquesIn Big Data Processing: Trends, Challenges And Opportunities”. Knowledge-Based Syst. 118, 15–30. 2017.

[4] M. Giacalone And .Vito Santarcangelo, “Big Data Compliance For InnovativeClinical Models, Big Data Research, Available Online 5 February 2018 (In Press].

[5] K. Abouelmehdi, A. Beni-Hssane, H.Khalouflai, And M. Saadi’ “Big Data SecurityAndPrivacyInHealthcare: A Review”, Procedia Computer Science, Vol. 113 , 73–80, 2017

[6] A.F.Simpao, L.M.Ahumada, J.A Galnez, And M.A.Rehman, “ A Review Of Analytics An Clinical Informatics In Health Care.” Journal Of Medical Systems, Vol.38, No 4, P 45, 2014.

[7] US Department Of Health And Human ServicesGenomicsDataSharing.2014Https://Gds.Nih.Gov/17summary_DBgap_StatiStics.Html Webcite

[8] R. Da MotaUeti, D. F Espinosa, D. F. Espinosa, L. Rafferty And PatricC.K.Hung , “Big Data Applications And Use Cases” Editors Patrick C. K. (Ed.), Sprig-er, 2016DOI 10.1007/978-3-319-30146-4_9

[9] Daniel A Read AndT.Donnngarra, “ExascaleComputting And Big Data”, Magazine Archive, Vol 58, No.7, 2015.

[10] G. Bello-Orgaz, J. J.Jung, And David Camacho, “Social Big Data: Recent

AchievementsAndNew Challenges", Information Fusion, Vol.28, 45-49. 2016.

[11] D.Apiletti,Baralis,T.Cerquitelli,P.Garza,F.Pulvirenti, And P.Michiardi, "A Parallel Mapreduce Algorithm To Efficiently Support Itemset Mining On High Dimensional Data", Big Data Research, Vol. 10, Pp 53-69, 2017.

[12] C.L.Philip And Chenchun-Yang Zhang, " Data-Intensive Applications, Challenges, Techniques And Technologies: A Survey On Big Data" Information Science,Vol.275,314-347, 2014

[13] Q. Huang, S. Jing, J.Yi, W. Zhen, "Innovative Testing And Measurement Solutions For Smart Rid" John Wiley &Sons: Singapore, 2015

[14] C.K. Emani, N.Cullot, And C. Nicolle,"Understandable Big Data: A Survey". Comput. Sci.Rev. Vol.17, 70–81.2015.

[15] A. Gandomi, And M. Haider, " Beyond The Hype: Big Data Concepts, Methods, And Analytics." Int. J. Inf. Manage, Vol.35, 137–144. 2015.

[16] B.Furht, And F.Villanustre, "Introduction To Big Data.In: Big Data Technol." App..Springer International Publishing, Cham, Pp. 3–11. 2016.

[17] M.Dayal AndNanhay Singh, "Indian Health Care Analysis Using Big Data ProgrammingTool," ,Procedia Computer Science , Vol.89 , 521 – 52, 2016.

[18] J. Luo, M .Wu, D.Gopukumar, And Yiqing Zhao, "Big Data Application In Biomedical Research And Health Care: A Literature Review" Published By Libertas Academica, University Of Wisconsin–Milwaukee, Milwaukee,WI, USA. Biomedical Informatics Insights , 8, 2016.

[19] M. Mancini, "Exploiting Big Data For Improving Healthcare Services J", E-Learning Knowl. Soc., Vol.10 , Pp. 23-33, 2014.

[20] M. Bouhriz, And H. Chaoui, "Big Data PrivacyInHealthcareMoroccanCo-text"ProcediaComput. Sci., Vol. 63 , Pp. 575- 580, 2015

[21] H.K. Patil, And R.Seshadri , " Big Data Security And Privacy Issues In Healthcare. " IEEEInt. Congr. Big Data Pp. 762–765, 2014.

[22] H.Stoklosa, E.Showalter, A.Melnick, And E.F.Rothman, "Health Care Providers' Experience With A Protocol For The Identification, Treatment, And Referral Of Humantrafficking Victims." Journal Of Human Trafficking, Vol. 3 (3) , 182–192. 2016, Doi:10.1080/23322705.2016.1194 668.

[23] PY. Wu, C.W. Cheng, C.D. Kaddi, J. Venugopalan,R. Hoffman,And.D. Wang,"Omic And Electronic Health Record Big Data Analytics For Precision Nmedicine" Ieee Trans. Biomed. Eng Vol. 64 , Pp. 263- 273, 2017.

[24] I. Merelli, H. Pérez-Sánchez, S. Gesing, And D. D'Agostino, "Managing, Analysing, And Integrating Big Data In Medical Bioinformatics: Open Problems And Future Perspectives" Biomed Res. Int., Vol. 2014 (2014), Http://Dx.Doi.Org/10.1155/2014/134023

[25] H. Kumar And N.Singh, "Review Paper On Big Data In Healthcare Informatics" International Research Journal Of Engineering And Technology, Volume: 04 Issue: 02 ,2017

[26] R.R. Kitchen, J.S. Rozowsky, M.B. Gerstein And A.C. Nairn, "Decoding Neuroproteomics: Integrating The Genome,Translatome And Functional Anatomy". Nat. Neurosci. Vol. 17, 1491–1499. 2014. [Crossref] [Pubmed]

[27] A. Belle, R. Thiagarajan, S. M. Reza Soroushmehr, F. Navidi, Daniel A. Beard, And K.Najarian "Big Data Analytics In Healthcare" Biomed Research International, Volume 2015 (2015), [Article ID 370194

[28] Q-A .Yao, H. Zheng, Z-Y. Xu Et Al. "Massive Medical Images Retrieval System Based On Hadoop". J Multimed.; Vol.9(2), 216, 2014.

[29] H. Jiang, Y.Chen , Z. Qiao ,T.H. Weng, And K.C. Li ."ScalingUpMapreduceBasedBigDataProcessing On Multi-GPU Systems" Hadoop Based E-Book Conversion System, ClusterComputing, 18 (1),Pp. 369-383, 2015.

[30] A.Sarker,R.Ginn.A.Nikfarjam,K.Connor,K.Smith,S.Jayaraman,T.Upadhaya,AndG.Gonzalez, "Utilizing Social Media Data For PharmacoVigilance: A Review"Vol. 54, PP 202-212, 2015.

[31] D. Sims, I. Sudbery, N.E.Ilott, A. Heger, And C.P. Ponting, " Sequencing Depth And Coverage: Key Considerations In Genomic Analyses" Int. J. Mol. Sci. , Vol.18, 412 -15, 2017 [Pubmed]

[32] D.M.Church,V.A.Schneider, K.M.Steinberg, And M.C.Schatz, A.R.Quinlan, C.S Chin, P.A.Kitts, B.Aken, G.T. Marth, And M.M.Hoffman, Et Al. "Extending Reference Assembly Models". Genome Biol. Vol.16, 13. 2015, [Pubmed]

[33] D. Acharjya, And K.P. Ahmed," A Survey On Big Data Analytics: Challenges, Open Research Issues And Tools". Int. J. Adv.

Comput. Sci. App., Vol. 7, 511–518, 2016a.. [Pubmed]

[34] A. Yang, M. Troup, And Joshua W.K. Ho, " Scalability And Validation Of Big Data Bioinformatics Software", Comput Struct Biotechnol J.; Vol. 15:, 379–386. 2017.

[35] A. Oussous, Fatima-Zahra Benjelloun, A. A. Lahcen And, S. Belfkih, "Big Data Technologies: A Survey "Journal Of King Saud University – Computer And Information Sciences, Vol. Xxx Xxx ,2017(Available At Science Direct , In Press)

[36] H.Wang, Z. Xu, H. Fujita, And S. Liu, "Towards Felicitous Decision Making: An Overview On Challenges And Trends Of Big Data". Inf. Sci. Vol.367, 747–765, 2016.

[37] W. Raghupathi And V. Raghupathi, "Big Data Analytics In Healthcare: Promise And Potential", Health Information Science And Systems , 2, 3, 2014.

[38] G. Skourletopoulos, C.X.Mavromoustakis, G. Mastorakis, J.M. Batalla, C. Dobre, S.Panagiotakis, And E.Pallis, "Big Data And Cloud Computing: A Survey Of The State-Of-The-Art And Research Challenges. In: Variousadvances In Mobile Cloud Computing And Big Data In The 5G Era". Springer, Pp. 23–41. 2017.

[39] Z. Yin, H. Lan, G. Tan, Mian Lu, Athanasios V. Vasilakos, And W. Liu, "Computing Platforms For Big Biological Data Analytics: Perspectives And Challenges" Comput Struct Biotechnol J. Vol.15: 403–411.2017.

[40] L. Zhang, H. Wang, Researcher, Q. Li, Ming-Hui Zhao And Qi-Min Zhan, "Big Data And Medical Research In China" BMJ 2018; 360. Doi: Https://Doi.Org/10.11 36/Bmj.J5910 ,

[41] P.L.Peissig,A.Nikolai,AndM.Brilliant,"Personalized Medicine. In Drug Discovery And Evaluation": Pharmacological Assays; Hock, F.J.,Ed.;Springer:Berlin/Heidelberg,Germany, Pp. 1–16 ,2015.

Nidhi Raj "Big Data Science and Its Applications in Biomedical Research and Healthcare: A Review "International Journal of Engineering Research and Applications (IJERA) , vol. 8, no.5, 2018, pp. 45-52