

Big Data Technology In Health and Biomedical Research: A Literature Review

Revati Raman Dewangan, Deepali Thombre and Chitranjan Patel

Asst. Prof., CSE Department, Asst. Prof., IT Department, Senior Consultant
CCET Bhilai (CG), SSEC, Bhilai (CG), Capgemini India
revati2004@gmail.com, deepthombre@gmail.com, chitranjan.patel@gmail.com

Abstract

Big data and hadoop technologies are used for biomedical and health-care informatics research purpose. Huge information advances are progressively utilized for biomedical and medicinal services informatics research. A lot of natural and clinical information have been produced and gathered at a remarkable speed and scale. For instance, the new era of sequencing innovations empowers the expert censing of billions of DNA grouping information every day, and the use of electronic health records (EHRs) is reporting a lot of patient information. The expense of securing and examining biomedical information is relied upon to diminish drastically with the assistance of innovation overhauls, for example, the rise of new sequencing machines, the improvement of novel equipment and programming for parallel processing, and the broad extension of EHRs. Huge information applications show new chances to find new learning and make novel techniques to enhance the nature of human services. The utilization of enormous information in social insurance is a quickly developing field, with numerous new disclosures and procedures distributed in the most recent five years. In this paper we present different areas of biomedical fields like bioinformatics, clinical informatics, imaging informatics, and general health care informatics. In particular, in bioinformatics, high-throughput tests encourage the exploration of new expansive affiliation investigations of ailments, and with clinical informatics, the clinical field profits by the boundless measure of gathered patient information for settling on savvy choices. Imaging informatics is presently all the more quickly incorporated with cloud stages to share therapeutic picture information and work processes, and general wellbeing informatics influences huge information methods for foreseeing and observing irresistible illness flare-ups, for example, Ebola. In this paper, we survey the late advance and leaps forward of enormous information applications in these human services areas and compress the difficulties, holes, and chances to enhance and progress huge information applications in social insurance.

Keywords: *Big data, electronic health records (EHRs), DNA, bioinformatics, clinical informatics, imaging informatics, public healthcare informatics and analysis*

Big Data Technologies

Big Data Analytics is the process of examination in such a way toward looking at enormous information to reveal shrouded designs, obscure connections and other helpful data that can be utilized to settle on better choices. With enormous information investigation, information researchers and others can break down gigantic volumes of information that traditional examination and business insight arrangements can't touch. Consider that your association could amass billions of columns of information with a huge number of information blends in different information stores and plenteous arrangements. Elite investigation is important to process that much information so as to make sense of what's critical and what isn't [1].

Presently Big Data is a reality: The volume, assortment and speed of information coming into your association keep on reaching exceptional levels. The Big Data wonder has developed as an aftereffect of immense measures of information that are getting to be accessible over an extensive variety of use areas crosswise over science, business, and government.

Data may originate from many different sources, including:

- digitally media, including text, images, audio, and emails;
- scientific instruments;
- Streaming information from weblogs, recordings, money related/business exchanges;
- social interactional information from interpersonal organizations, twitter bolsters and snap streams;
- scientific data from huge scale reviews, and cerebrum research.
- from omnipresent detecting and control applications in built and normal frameworks;



Figure 1. Big Data and Health Care Application

Medicinal services information, and related metadata, gathered by machine-to-machine (M2M) advancements. For instance, social insurance suppliers may utilize remote sensors to accumulate information, for example, pulse, heart rate or blood levels, and exchange that information through the Internet to a controlling server where it can be examined and followed up on.

There are n number of application of Big Data Technology in HealthCare field few of them are presented here.

1. Big Data Hadoop Technology in Cancer Treatments and Genomics

Deepak Singh, the essential item supervisor at Amazon Web Services, said, "We've certainly seen an uptake in receiving Hadoop in the life sciences group, generally focusing cutting edge sequencing, and straightforward read mapping since what designers found was that various bioinformatics issues exchanged extremely well to Hadoop, particularly at scale."

Industry reports show that, there are around 3 billion base combines that constitute the human DNA and it is fundamental for such a lot of information to be sorted out in a powerful way in the event that we need to battle malignancy. The main motivation why growth has not been cured yet is a direct result of the way that disease changes in various examples and responds in various routes in light of the hereditary cosmetics of a person. Subsequently, oncology analysts have thought of an answer that so as to cure tumor,

patients should be given customized treatment in view of the kind of disease the individual patient's hereditary qualities make up. Utilizing Hadoop innovation will offer awesome backing for parallelization and help in mapping the 3 billion DNA base sets utilizing MapReduce programs [15].

2. Hadoop Technology in the Hospital Network

A Cleveland Clinic spinoff organization known as Explorys is making utilization of Big Data in human services to give the best clinical backing, lessen the expense of consideration estimation and deal with the number of inhabitants in at-danger patients. Explorys has purportedly manufactured the biggest database in the medicinal services industry with over a hundred billion information directs all thanks toward Hadoop.

Explores utilizes Hadoop innovation to help their restorative specialists break down information bombardments progressively from various sources, for example, money related information, finance information, and electronic wellbeing records.

3. Hadoop Technology in Healthcare Intelligence

Medicinal services Insurance Business works by examining the related costs (the danger) and similarly isolating it by the quantity of individuals in the danger bunch. In such circumstances, the information and the results are constantly rapid and changing. Using Hadoop innovation in Healthcare Intelligence applications helps doctor's facilities, payers and medicinal services organizations expand their upper hands by concocting brilliant business arrangements [16].

For example, we should expect that, a medicinal services insurance agency is occupied with finding the age in a specific area where people beneath that age are not casualties of specific illnesses. This information will help the safety net provider process the expense of protection strategy. To accumulate craved age, insurance agencies will need to handle enormous information sets to separate significant data such as medicines, illnesses, side effects, sentiments, geographic area subtle element and so forth. In this situation, utilizing Hadoop's Pig, Hive and MapReduce is the best answer for procedure such vast datasets.

4. Hadoop technology in Monitoring Patient Vitals

There are a few healing facilities over the world that utilization Hadoop to help the doctor's facility staff work productively with Big Data. Without Hadoop, most patient consideration frameworks couldn't envision working with unstructured information for investigation.

Kids' Healthcare of Atlanta treats more than 6,200 youngsters in their ICU units. All things considered, the length of stay in Pediatric ICU differs from a month to a year. Youngsters' Healthcare of Atlanta utilized a sensor adjacent to the bed that helps them consistently track understanding signs, for example, circulatory strain, pulse and the respiratory rate. These sensors deliver vast pieces of information, which utilizing legacy frameworks can't be put away for over 3 days for analysis. The primary intention of Children's Healthcare of Atlanta was to store and investigate the fundamental signs. In the event that there is any adjustment in example, then the clinic needed a caution to be produced to a group of specialists and aides [18]. This was effectively accomplished utilizing Hadoop environment parts - Hive, Flume, Sqoop, Spark, and Impala.

Applications of Big data in Health Care fields

1) Neuroinformatics

Neuroinformatics examination is a youthful subfield, as every information occasion, (for example, MRIs) is very substantial prompting datasets with Big Volume. No one but as of late can computational force stay aware of the requests of such research.

Neuroinformatics focuses its exploration on examination of mind picture information (tissue level) keeping in mind the end goal to: figure out how the cerebrum works, discover connections between's data assembled from mind pictures to restorative occasions, and so on., all with the objective of promoting therapeutic learning at different levels. We picked the field of Neuroinformatics to speak to the more extensive area of Medical Image Informatics in light of the fact that by restricting the degree to cerebrum pictures, more top to bottom examination might be performed while as yet assembling enough data to constitute Big Data [4]. Starting here on Neuro informatics research utilizing tissue level information will be referenced by information level as opposed to the subfield [17].

2) Public health information

General Health Informatics applies information mining and examination to populace information, keeping in mind the end goal to increase restorative understanding. Information in Public Health Informatics is from the populace, accumulated either from "conventional" means (specialists or clinics) or assembled from the populace (online networking) [3]. In either occasion, populace information has Big Volume, alongside Big Velocity and Big Variety. Information accumulated from the populace through online networking could have low Veracity prompting low Value, yet methods for separating the helpful data from social networking, (for example, Twitter posts), this line of information can likewise have Big Value cdot.

3) Clinical informatics

Clinical Informatics research includes making expectations that can help doctors improve, quicker, more precise choices about their patients through examination of patient information. Clinical inquiries are the most vital inquiry level in Health Informatics as it works straightforwardly with the patient. This is the place a disarray can emerge with the expression "clinical" when found in exploration, as all Health Informatics examination is performed with the consequent objective of foreseeing "clinical" occasions (straightforwardly or by implication). This perplexity is the explanation behind characterizing Clinical Informatics as just research which straightforwardly utilizes persistent information. With this, information utilized by Clinical Informatics research has Big Value [5].

Indeed, even with all exploration in the long run noting clinical domain occasions, as per Bennett *et al.* there is around a 15 ± 2 year hole between clinical examination and the genuine clinical consideration utilized as a part of practice. Choices nowadays are made for the most part on general data that has worked some time recently, or in light of what specialists have found to work before [19]. Through all the exploration introduced here and also with all the examination being done in Health Informatics, the human services framework can grasp new ways that can be more precise, solid, and effective.

4) Bioinformatics

Research in Bioinformatics may not be considered as a feature of customary Health Informatics, however the examination done in Bioinformatics is an imperative wellspring of wellbeing data at different levels [2]. Bioinformatics concentrates on expository exploration keeping in mind the end goal to figure out how the human body functions utilizing sub-atomic level information as a part of expansion to creating strategies for viably taking care of said information. The expanding measure of information here has enormously expanded the significance of creating information mining and investigation systems which are effective, touchy, and better ready to handle Big Data.

Information in Bioinformatics, for example, quality expression information, is consistently developing (because of innovation having the capacity to create more sub-atomic information per individual), and is unquestionably classifiable as Big Volume. The issue of Big Volume inside sub-atomic information prompts research, for example, McDonald *et al.* who made a Bioinformatics suite of programming instruments they call khmer. This suite looks to tackle equipment computational issues through

programming. The instruments in this suite pre-prepare Big Volume genomic arrangement information by separating long successions into generally short strings which can be put away in a Bloom channel based hash table, helping both the capacity and proficiency of investigation of Bioinformatics information.

5) Imaging informatics applications

Imaging informatics is the investigation of strategies for producing, overseeing, and speaking to imaging data in different biomedical applications. It is worried with how medicinal pictures are traded and broke down all through complex human services frameworks. With the developing requirement for more customized consideration, the need to join imaging information into EHRs is quickly expanding.

In this segment, we arranged huge information advances/instruments into three classifications: (1) information stockpiling and recovery, (2) information sharing, and (3) information examination. Imaging informatics grew all the while with the approach of EHRs and the rise of clinical informatics; nonetheless, it is altogether different from clinical informatics because of the heterogeneous information sorts created from various modalities of medicinal pictures. Information security remains an imperative thought here, but since ebb and flow frameworks principally depend on business cloud stages and existing conventions, for example, computerized picture correspondence in medication (DICOM), there is no exploration concentrating on enhancing information security in imaging informatics.

Information stockpiling and recovery. Imaging informatics is dominantly utilized for enhancing the productivity of picture preparing work processes, for example, stockpiling, recovery, and interoperation. PACS are mainstream for conveying pictures to nearby show workstations [20], which is expert basically through DICOM conventions in radiology divisions. Numerous electronic restorative applications have been produced to get to PACS, and more prominent utilization of huge information innovation has been enhancing their execution. Silva *et al.*⁸² proposed a way to deal with coordinate the information in PACS, given the present pattern among social insurance foundations to outsource the two vital segments of PACS (DICOM object vault and database framework) to the cloud. Silva *et al* proposed to furnish a unique layer with a Cloud IO (info/yield) stream instrument to bolster more than one cloud supplier notwithstanding their disparities in information access benchmarks [6].

Literature Review

For this research we have studied more than 20 paper related to our research few of them we have written here.

1) Big data in global health: improving health in low- and middle-income countries

Throughout the most recent decade, a gigantic increment in information gathering and investigation has happened in numerous fields. In the wellbeing segment, be that as it may, there has been moderately little advance in information examination and application in spite of a quick ascent in information generation. Given satisfactory administration, changes in the quality, amount, stockpiling and examination of wellbeing information could prompt generous upgrades in numerous wellbeing results.

In low-and center pay nations specifically, the production of a data criticism system can move medicinal services conveyance towards results-based practice and enhance the viable utilization of rare assets. We survey the advancing meaning of huge information and the conceivable points of interest of – and issues in – utilizing such information to enhance medicinal services conveyance in low-and center salary nations. The gathering of enormous information as cellular telephone based administrations enhance may imply that improvement stages required somewhere else can be skipped. In any case, poor framework may avert interoperability and the protected utilization of patient information. A proper administration system must be produced and upheld to secure people and

guarantee that human services conveyance is custom fitted to the qualities and estimations of the objective groups [7].

The conveyance of medicinal services is a perplexing try at both individual and populace levels. At the clinical level, the custom fitted procurement of consideration to people is guided, partially, by medicinal history, examination, fundamental signs and proof. In the 21st century these conventional precepts have been supplemented by an emphasis on learning, measurements and quality change. The gathering and investigation of information of good quality are basic to upgrades in the adequacy and proficiency of human services conveyance. A significant portion of the waste in human services consumption results from not realizing what works for specific patients specifically clinical settings. Intercessions that seem compelling in populace based studies are frequently broadly executed with no checking or any endeavor to recognize the patients more prone to profit by – or to be hurt by – the mediations.

The difficulties of producing, breaking down and applying clinical information are especially intense in low-and center salary nations. Given the sheer size of the human populace and the mind blowing many-sided quality of social insurance conveyance – with a large number of infections and a huge number of solutions and mediations – the compromise of information driven changes in clinical drug with great populace wellbeing is unpredictable. Be that as it may, the late advancement of new strategies to gather, investigate and apply information on a phenomenal scale – the alleged enormous information approach – may permit the hole between social insurance conveyance and populace wellbeing to be spanned and numerous wellbeing results to be made strides. These new techniques for gathering, curating and conceptualizing numbers offer comparable points of interest to populaces as those offered to patients by therapeutic imaging – *i.e.* they offer the chance to see with more prominent accuracy [21].

In numerous fields, over late decades, the era and utilization of information have quickly extended and new information applications have been produced. Be that as it may, despite the fact that there has been a simultaneous increment in the generation of information in the wellbeing division, there has been small relating change in the utilization of such information to enhance wellbeing care. Some confirmation of-idea applications have been made for – and energetically got by – wellbeing experts be that as it may, in this way, they have brought about not very many substantial, framework wide information activities. In this article we investigate some present and potential utilizations of enormous information to general wellbeing and human services conveyance in low-and center wage nations.

In the field of medicinal services conveyance, the enormous information methodology may speak to a noteworthy turning point – encouraging the advancement of learning frameworks of consideration and empowering more exact administration of people to enhance the strength of whole populaces. Sheer size increments both the potential dangers and potential advantages of the methodology. In spite of the fact that the methodology may have most esteem in low-asset settings, it is additionally most helpless against fracture and abuse in such settings. Shared administration, cautious investigation and specialized associations are expected to minimize the dangers. The complexities ought not be thought little of. In low-and center pay nations, the shepherding of the move from paper records to petabytes of advanced stockpiling gives another chance to worldwide wellbeing foundations to offer helpful administration.

2) Big Data Analytics in Bioinformatics: A Machine Learning Perspective

Bioinformatics examination is described by voluminous and incremental datasets and complex information investigation strategies. The machine learning techniques utilized as a part of bioinformatics are iterative and parallel. These techniques can be scaled to handle huge information utilizing the disseminated and parallel figuring advances.

Generally huge information devices perform calculation in bunch mode and are not streamlined for iterative handling and high information reliance among operations. In the

late years, parallel, incremental, and multi-view machine learning calculations have been proposed. So also, diagram based designs and in-memory huge information instruments have been produced to minimize I/O cost and upgrade iterative handling [22].

Be that as it may, there need standard enormous information designs and apparatuses for some vital bioinformatics issues, for example, quick development of co-expression and administrative systems and remarkable module distinguishing proof, location of edifices over developing protein-protein connection information, quick examination of huge DNA, RNA, and protein grouping information, and quick questioning on incremental and heterogeneous malady systems. This paper addresses the issues and difficulties postured by a few major information issues in bioinformatics, and gives an review of the cutting edge and the future exploration opportunities.

Aside from digitization of administrations and endeavors, another pattern has developed as of late to network all the man-made things around us, for example, autos, home apparatuses, weapons, movement lights, and power meters. These things speak with each other to share information caught through different sensors, so as to take canny operational choices without anyone else. This system has been termed as the Internet of Things (IoT) [8]. The initially organized apparatus, a coke candy machine, was sent at the Computer Science bureau of Carnegie Mellon College in the year 19901. The IoT is developing quick and machine-to-machine associations will achieve 1.2 billion in 2017, up from just 200 million in 2012 [9,23].

Notwithstanding, it ought to be noticed that not all information, that we produce, are helpful for distinct or prescient examination. Just a part of the information in the computerized universe is helpful, whenever labeled, termed as target-rich information. Metadata are more target-rich than the information itself. As indicated by Turner *et al.* [10], around the majority of the objective rich information were general IT information in the year 2014; in any case, by the year 2020, IoT information will possess more than 20% of the objective rich information lake.

This paper talks about the late surge in bioinformatics information stores regarding volume and in addition measurement. With the appearance of new high throughput and shoddy information catching devices, this fast development in information will proceed in the coming a long time. Bioinformatics information are voluminous, heterogeneous, incremental, and disseminated geologically everywhere throughout the world. Therefore, the enormous information examination methods are required to take care of the issues in bioinformatics. The issues, information sources and information sorts in bioinformatics are differing in nature. The current huge information designs try not to give a thorough answer for huge information investigation, which is quick, blame tolerant, expansive scale, incremental, dispersed, and improved for iterative and complex calculations. The surely understood MapReduce design for disseminated figuring executes in a clump mode furthermore, has high circle read/compose overhead. Then again, the chart based models for gushing applications come up short to give adaptation to internal failure. An incorporated huge information investigation design that satisfies the prerequisites of the issues in bioinformatics is a dire need. Machine learning has been the most used instrument for information investigation. Extensive scale information existed well before the huge information period, especially in bioinformatics.

Machine learning devices have been effectively used to break down both little scale and also extensive scale information utilizing different systems for example, testing, highlight determination, and conveyed calculations. In any case, huge information postures more difficulties on the conventional learning strategies as far as speed, assortment, what's more, incremental information. Conventional learning techniques as a rule insert iterative handling and complex information reliance among operations. Thus, the conventional machine learning strategies can't be utilized to perform quick handling on enormous information utilizing huge information stages, for example, MapReduce. This paper talks about the conventional machine learning techniques, their confinements,

and the endeavors made in the late years to develop them for huge information, for example, the incremental, parallel, and multi-view grouping techniques to handle complex bioinformatics issues.

3) Big Data for Health

This paper gives a review of later improvements in enormous information with regards to biomedical and wellbeing informatics. It plots the key attributes of huge information and how medicinal and wellbeing informatics, translational bioinformatics, sensor informatics, and imaging informatics will profit by a coordinated methodology of sorting out various parts of customized data from an assorted scope of information sources, both organized and unstructured, covering genomics, proteomics, metabolomics, and also imaging, clinical determination, and long haul nonstop physiological detecting of a person. It is normal that late advances in enormous information will grow our insight for testing new speculations about illness administration from analysis to aversion to customized treatment. The ascent of huge information, in any case, likewise brings challenges up as far as protection, security, information possession, information stewardship, and administration [24].

This paper examines some of the current exercises and future open doors identified with huge information for wellbeing, sketching out a portion of the key basic issues that need to be handled. In this paper, a few methods for characterizing enormous information exist as a wide term to embody the difficulties identified with the handling of a huge measure of organized and unstructured information. Plainly, the size (or volume) of information is a critical variable of enormous information.

Without a doubt, the US social insurance framework alone as of now achieved 150 exabytes (10¹⁸) five years prior [11]. After a short time, we will bargain with zettabyte (10²¹) and yottabyte (10²⁴) information for nations with extensive populaces including developing economies, for example, China and India. This pattern is because of the way that multiscale information produced from people is ceaselessly expanding, especially with the new high-throughput sequencing stages, continuous imaging, and purpose of consideration gadgets, and additionally wearable processing and versatile wellbeing innovations. They give genomics, proteomics and metabolomics, and also long haul consistent physiological components of a person. In parallel, ecological elements exhibit yet another arrangement of variables that can be caught by consistent detecting that are vital to populace wellbeing. In human services, information heterogeneity and assortment emerge subsequently of connecting the differing scope of biomedical information sources accessible. Sources can be either quantitative (*e.g.*, sensor information, pictures, quality clusters, research center tests) or subjective (*e.g.*, free content, demographics). The goals hidden this information challenge are to bolster the premise for observational proof to reply clinical inquiries, which would not generally been comprehended by means of thinks about in light of randomized trials alone. What's more, the issue of summing up results in view of a slender range of members might be comprehended by exploiting the capability of enormous information for sending longitudinal studies[25].

Volume, Velocity, and Variety are the three Vs in the first meaning of the key attributes of enormous information in the exploration report distributed by META Group, Inc. (presently Gartner, Inc.) [12]. From that point forward, different variables have likewise been considered, including Variability (consistency of information after some time), Veracity (dependability of the information got), and Value. These attributes are elements that each catches.

Veracity is critical for huge information as, for instance, individual wellbeing records may contain typographical mistakes, condensings, what's more, mysterious notes. Walking estimations are once in a while taken inside less dependable, uncontrolled situations analyzed to clinical information, which are gathered via prepared experts. The utilization of unconstrained unmanaged information, for example, those from online networking, can prompt wrong forecasts as the information setting is not generally

known. Moreover, sources are regularly one-sided toward those youthful, web clever, and expressive on the web.

Huge information can serve to support the relevance of clinical exploration examines into genuine situations, where populace heterogeneity is an obstruction. It similarly gives the open door to empower powerful and accuracy solution by performing tolerant stratification. This is for sure a key assignment toward customized human services. A superior utilization of restorative assets by method for personalization can prompt all around oversight wellbeing administrations that can beat the difficulties of a quickly expanding and maturing populace. In this manner, progresses in huge information preparing for wellbeing informatics, bioinformatics, detecting, and imaging will have an incredible effect on future clinical examination. Another vital element to consider is quick and consistent wellbeing information securing, which will add to the achievement of huge information in pharmaceutical [13].

In particular, detecting gives a strong arrangement of answers for fill this hole. Frequencies of wellbeing information procurement still include a moderate what's more, complex procedure requiring the inclusion of unique wellbeing individual and research centers. In this setting, speedier and subtle wellbeing information can be given by method for pervasive detecting. The utilization of sensors means the limit of covering vast times of nonstop observing without the requirement for performing sporadic screening, which may just speak to a tight picture of the improvement of an ailment. Be that as it may, the reality of conveying nonstop detecting over a vast populace will bring about an extensive measure of data that requires both on-hub information reflection what's more, appropriated surmising. From a populace level, one's heartbreaking past can give critical knowledge into estimating furthermore, keeping the same episode from happening in others. Last yet, not the minimum, the administrative strategy and direction are required to guarantee security amid information transmission and capacity, and additionally amid consequent information examination undertakings.

5. Conclusion

This paper gives an understanding of how we can reveal extra esteem from the information produced by human services. Expansive measure of heterogeneous information is produced by these organizations. Be that as it may, without legitimate information examination techniques these information got to be futile. Huge Data Analytics utilizing Hadoop assumes a successful part in performing significant constant investigation on the immense volume of information and ready to foresee the crisis circumstances before it happens. It portrays about the huge information use cases in medicinal services. The issue is not the absence of information but rather the absence of data that can be utilized to bolster basic leadership, arranging and system. The whole government framework can understand profits by using huge information innovations. To effectively distinguish and execute huge information arrangements and advantage from the worth that enormous information can bring, government need to dedicate time, assign spending plan and assets to visioning and arranging. With the assistance of Hadoop the objective of powerful national consideration administration can be accomplished by giving a compelling information driven administrations to natives by foreseeing their needs taking into account the examination of review led among various classes of subjects.

References

- [1] <http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care>
- [2] D. E. Genome and nation, "Iceland's health sector database and its legacy", *Innovations*, doi: <http://dx.doi.org/10.1162/itgg.2006.1.2.80>, vol. 1, no. 2, pp. 80–105.
- [3] M. L. Miranda, J. Ferranti, B. Strauss, B. Neelon and R. M. Califf, "Geographic health information systems: a platform to support the 'triple aim'", *Health Aff (Millwood)*, doi: <http://dx.doi.org/10.1377/hlthaff.2012.1199> PMID: 24019366, vol. 32, no. 9, (2013), pp. 1608-15.
- [4] G. Robins, P. Pattison, Y. Kalish and D. Lusher, "An Introduction to Exponential Random Graph (p*) Models for Social Networks," *Social Networks*, no. 29:2, (2007), pp. 173-191.
- [5] D. Watts, "Six Degrees: The Science of a Connected Age", New York: W. W. Norton, (2003).
- [6] I. H. Witten, E. Frank and M. Hall, "Data Mining: Practical Machine Learning Tools and Techniques (3rd ed.)", San Francisco: Morgan Kaufmann, (2011).
- [7] "Big data in global health: improving health in low- and middle-income countries Rosemary", Wyber, a Samuel Vaillancourt, b William Perry, c Priya Mannava, c Temitope Folaranmic & Leo Anthony Celid, *Bull World Health Organ*, doi: <http://dx.doi.org/10.2471/BLT.14.139022>, vol. 93, (2015), pp. 203-208.
- [8] "Big Data Analytics in Bioinformatics: A Machine Learning Perspective", HIRAK KASHYAP, HASIN AFZAL AHMED, NAZRUL HOQUE, SWARUP ROY, and DHRUBA KUMAR BHATTACHARYYA, *JOURNAL OF LATEX CLASS FILES*, vol. 13, no. 9, (2014).
- [9] V. Turner, J. Gantz, D. Reinsel, and S. Minton, "The digital universe of opportunities: Rich data and the increasing value of the internet of things", International Data Corporation, White Paper, IDC 1672, (2014).
- [10] N. Gershenfeld, R. Krikorian and D. Cohen, "The internet of things", *Scientific American*, vol. 291, no. 4, (2004), pp. 76.
- [11] "Big Data for Health", Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrifield, Stephen T. C. Wong, and Guang-Zhong Yang, Fellow, *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 4, (2015).
- [12] H. Xu, M. C. Aldrich, Q. Chen, H. Liu, N. B. Peterson, Q. Dai, M. Levy, A. Shah, X. Han, and X. Ruan, "Validating drug repurposing signals using electronic health records: A case study of metformin associated with reduced cancer mortality", *J. Amer. Med. Informat. Assoc.*, (2014), pp. 1-10.
- [13] Y. Hagar, D. Albers, R. Pivovarov, H. Chase, V. Dukic, and N. Elhadad, "Survival analysis with electronic health record data: Experiments with chronic kidney disease", *Statist. Anal. Data Mining, ASA Data Sci. J.*, vol. 7, (2014), pp. 385-403.
- [14] "Impact of Biological Big Data in Bioinformatics", Divya Kumari Liqid e Learning Services Private Limited, Noida, Ravi Kumar Persistent System Ltd International Journal of Computer Applications, vol. 101, no. 11, (2014), pp. 0975-8887.
- [15] V. Marx, "The Big Challenges of Big Data", *Nature*, vol. 498, (2013), pp. 255-260.
- [16] J. Clarke, "Continuous base identification for single-molecule nanopore DNA sequencing", *Nat Nanotechnol.* vol. 4, (2009), pp.265-70.
- [17] A. D. Baxevanis, B. F. Francis and B. F. F. Quellet, "Bioinformatics: A practical guide to the analysis of genes and proteins", (2005).
- [18] F. F. Costa, "Big Data in Biomedicine," *Drug Discovery Today*, in Press, (2013).
- [19] F. S. Collins, E. D. Green, A. E. Guttacher and M. S. Guyer, "A vision for the future of genomics research", *Nature*, vol. 422, (2003), pp. 835-847.
- [20] "A review of data mining using big data in health informatics", Matthew Herland, Taghi M Khoshgoftaar and Randall Wald* Herland et al. *Journal of Big Data*, <http://www.journalofbigdata.com/content/1/1/2>, vol. 1, no. 2, (2014).
- [21] A. Thommandram, J. E. Pugh, J. M. Eklund, C. McGregor and A. G. James, "Classifying neonatal spells using real-time temporal analysis of physiological data streams: Algorithm development In: IEEE Point-of-Care Healthcare Technologies (PHT 2013)", IEEE, based in New York, USA, Bangalore, India, (2013), pp. 240-243.
- [22] C. Ballard, K. Foster, A. Frenkiel, B. Gedik, M. P. Koranda, S. Nathan, D. Rajan, R. Rea, M. Spicer, B. Williams and V. N. Zoubov, "IBM Infosphere Streams: Assembling Continuous Insight in the Information Revolution", [<http://www.redbooks.ibm.com/abstracts/sg.pages=247970.html>], (2011).
- [23] J. Zhang, T. D. Y. Chung and K. R. Oldenburg, "Validation of high throughput screening assays", *Journal Biomolecular Screening*, vol. 4, no. 2, (1999), pp. 67-73.
- [24] J. L. Gall, S. Lemeshow and F. Saulnier, "A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study", *JAMA*, [<http://dx.doi.org/10.1001/jama.1993.03510240069035>], vol. 270, no. 24, (1993), pp. 2957–2963.
- [25] H. Akaike, "A new look at the statistical model identification", *IEEE Trans Automatic Control*, vol. 19, no. 6, (1974), pp. 716-723.