International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Big Data in Medical Field

Tigulla Hruthika Goud[1], Tigulla Hrushitha Goud[2], Kandhala Ananth Pal Reddy[3], Kandukuri Jagan Mohan[4]

[1]Student, Department of Information Technology, Mahatma Gandhi Institute of Technology
[2,4]Student, Department of Computer Science Engineering (Artificial Intelligence & Machine Learning), CVR College of Engineering
[3]Student, Department of Computer Science Engineering, Sreyas Institute of Engineering and Technology

Abstract

Large volumes of structured, unstructured, and semi-structured data have been produced in recent years by numerous global institutions; this heterogeneous data is collectively known as big data. Because several sources are known to provide large volumes of heterogeneous data, the health industry sector has to deal with the challenge of managing big data. In the healthcare industry, numerous big-data analytics technologies and approaches have been created to manage these enormous volumes of data. The influence of big data on healthcare, as well as the range of applications and technologies available within the Hadoop ecosystem, will all be covered in this paper. The fragmentation of healthcare data, ethical concerns, usability challenges, security and privacy concerns, and other issues are some of the barriers preventing the successful application of big data in healthcare, according to this research.

Keywords: Big data, Healthcare

1. Introduction

The global healthcare sector is among the largest and most expansive emerging industries. Global healthcare management has been shifting in recent years from a disease-centered to a patient-centered approach, and from a volume-based to a value-based healthcare delivery model. Healthcare companies are gradually becoming more and more in need of big data. Managing and analysing massive amounts of health data is necessary to deliver patient-centered treatment that works. Big data tools and technologies that are cutting edge and both equal to or better than current healthcare data management capabilities are desperately needed. The majority of big data studies in healthcare today only focus on big data's technological understanding. The big data analytics in healthcare IT had been discussed on very few topics. Examining the big data applications, processes, opportunities and challenges in healthcare management is the main idea behind this.

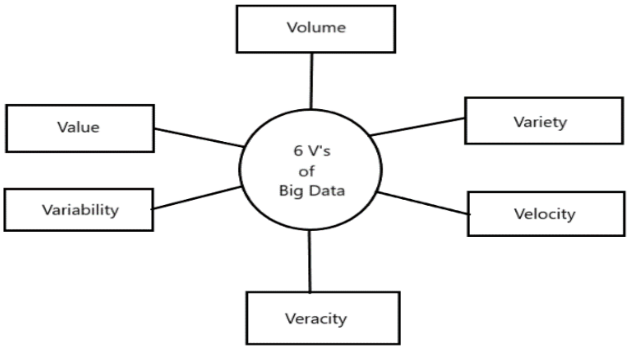
2. A short review of big data in healthcare

Big data refers to large volumes of complex, variable, and high-velocity data that necessitate the use of cutting-edge technologies and techniques for information collection, archiving, distribution, management, and analysis. Big data, as used in the healthcare industry, is a collection of large, complicated electronic health data that is challenging to distribute, process, and analyse using conventional methods. In the context of healthcare, big data can also be defined as an assortment of instruments, technologies, processes, and guidelines that are employed to efficiently generate, store, process, analyse, and retrieve sizable amounts of electronic health data. Big data, however, refers to more than just massive data sets; it also refers to the capacity to extract important and valuable information from a sizable collection of data sets by utilising cutting-edge tools and technologies. Big data in healthcare comes from the following sources: Machine Generated Data, Biometric Data, Human Generated Data, Transactional Data, Behavioural Data, Epidemological Data and Publication Data.

3. The meaning of big data

Various authors have described big data in healthcare. Usually, these qualities consist of volume, variety, and velocity. However, this paper categorises big data in healthcare into 6V's due to the shift in the nature of healthcare data. These include: volume, variety, velocity, veracity, variability and value.

Figure 1. Big Data's 6 V's in Health Care



- Volume: Volume is a term used to describe the amount of data that an organisation produces. Terabytes (1012 bytes), petabytes (1015 bytes), or exabytes (1018 bytes) are the units of measurement used today for healthcare data. Clinical data records will eventually amount to either a zettabyte (1021 bytes) or a yotta byte (1024 bytes). Such a large amount of data raises storage and analysis challenges.
- Variety: Broadly disparate data bases or mash-ups of data derived from independent sources in format are the sources of variety. Data in the healthcare industry can be categorised as organised, semi-organized, or unorganised based on its various formats. Data that is stored in Extensible Markup Language (XML) format, laboratory data, clinical data, sensor data, and data from relational databases are examples of semi-organized data. On the other hand, unorganised data are free-text data that typically lacks a precise design, such as manual written notes, data from X-ray, radiological, and other medical imaging, graphics, patient discharge summaries, physiological measures (signals), and healthcare data from social media and mobile phones. Eighty percent of big data are unstructured data sets.
- Velocity: The massive frequency at which current data is created, supplied, and managed is referred to as velocity. As a result, velocity encompasses both the production of data quickly and the processing of data quickly enough to satisfy demand. The third feature of big data is the rapid growth of data. There are two types of generated data: batch and real-time. The concentration of matching data assemblies, the summary of earlier data or inheritance gatherings, and the various forms of streamed data from various sources frequently cause fluctuations in the data's contents. One example of velocity

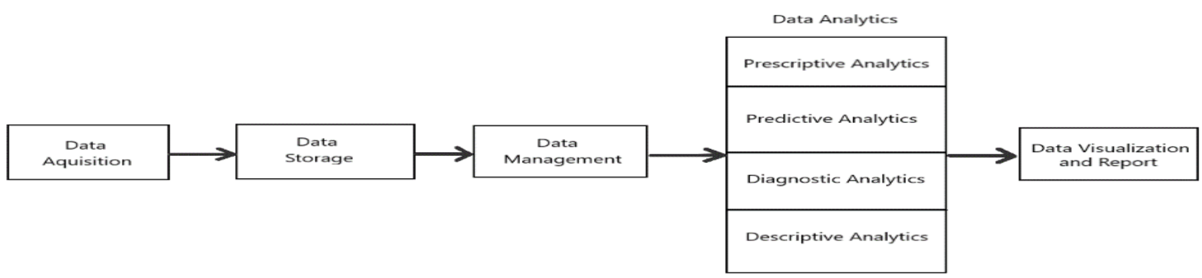
is the population's ongoing ageing, which leads to an increase in patients and a yearly growth rate of data of between 55 and 60 percent.

- Veracity: Information's accuracy and correctness are referred to as its veracity. Big data is hard to verify, has poor veracity, and is never 100% accurate. Since the majority of the data originates from unidentified and unverified sources, it is crucial to establish a standard to guarantee the data's features even before it is used.
- Variability: Variability describes changes in data that occur during handling and lifecycle. As range and variability increase, so does data's appeal and its potential to reveal important information that is hidden and unexpected.
- Value: Value, also known as big data analytics, is the process of gleaning important information from massive data sets. Value of data is helpful in making wise decisions.

4. Big Data Analysis Process in the Healthcare Sector

The potential for big data analysis to alter how healthcare suppliers use cultured equipment to raise awareness from their clinical and other data repositories and reach a conclusion is significant. Data acquisition, data storage, data management, data analytics, and data visualisation & report are the five steps in big data healthcare analytics. The big data analysis procedure in healthcare management is shown in the figure.

Figure 2. Health Care analysis process in Big Data



- Data acquisition: Big data for the healthcare industry can come from primary or secondary sources and can be organised, semi-structured, or unstructured.

  Primary sources:
  - Electronic health records, or EHRs, are databases that hold a patient's test results, demographic data, prescriptions, shots, and progress notes.
  - MRIs in medicine: X-rays, CT scans, and MRIs are examples of medical imaging methods that can be used to identify and track illnesses.

  Secondary sources:
  - Social media: Data on the frequency of diseases, their symptoms, and the effectiveness of treatments can be gathered using social media.

- Smartphones: Data on patient activity, mood, and blood glucose levels can be gathered using smartphone apps. Websites: Data on DNA and the microbiome can be gathered using websites like uBiome and 23andMe.

  Big data in healthcare can be applied to large-scale cohort studies, disease transmission analysis, better self-management of chronic conditions, and medical research support. It can also be used to customise medicine.
- Data storage: The most promising technology for big data analysis and storage in the healthcare industry is cloud computing. It can manage the massive volumes of data produced by healthcare organisations, both internal and external, and it is elastic, efficient, and economical. Additionally, cloud computing removes the need to purchase and maintain pricey computer software, hardware, and dedicated space. Healthcare organisations can save money on infrastructure costs, enhance clinician and researcher access to data, expedite the development of new treatments and therapies, and enhance patient care by examining big data through cloud computing.
- Data management: Healthcare data management encompasses data governance, data mining, organising, cleaning, and retrieval. It also covers how to verify if any values are missing or if there is some scraped data. Such information must be deleted. It supports patient risk assessment and tailored discharge planning. HCatalog and Apache Ambari are important tools for data management. The process of obtaining a file or important information from sizable healthcare databases is called data retrieval. Healthcare big data analysis usually involves data mining and information recovery.
- Data Analytics:

  Four categories of big data analytics are available in the healthcare industry:
- Descriptive analytics: Condenses past data to reveal patterns and insights.
- Diagnostic analytics: Makes use of past data to determine the underlying causes of issues.
- Predictive analytics makes use of both historical and current data to forecast future trends and outcomes.
- Prescriptive analytics: Makes recommendations for actions based on insights from predictive analytics.
- Data visualization: The process of translating healthcare data analysis findings into a visual or graphical format is known as data visualisation, and it helps people comprehend complex information and make better decisions. It can be applied to comprehend data patterns and correlations.

5. Big Data's Impact on the Healthcare System

Five ways in which big data analytics could transform healthcare are as follows:

- Right Living: By utilising data analytics, patients can make more informed decisions regarding their health and wellbeing.
- Appropriate Care: Healthcare providers can prevent effort duplication while patients can obtain the most suitable and efficient treatment.
- Appropriate Provider: By using data analytics, providers can gain a comprehensive understanding of their patients and offer more effective treatment options.
- Proper Innovation: It is possible to create novel medical technologies, therapies, and disease states more swiftly and effectively.
- Appropriate Value: Medical professionals can concentrate on the unique requirements of each patient while offering the best calibre and value in care.

- Big data analytics has the ability to enhance patient outcomes, lower costs, and boost system efficiency in the healthcare industry overall.

6. Opportunities

- Superior quality of care: Big data's ability to paint a more comprehensive picture of the patient is transforming the healthcare industry. Big data can assist physicians in identifying and resolving medical issues before they worsen by analysing data from a range of sources, such as social media, medical records, and DNA records. For instance, a physician can use big data to identify patients—even those without any outward signs of illness—who are susceptible to having a heart attack. After that, the physician can create a customised treatment plan to assist the patient in lowering their risk. Another example is the use of big data by physicians to determine which patients are most likely to experience adverse effects from a specific treatment or to respond well to it. The doctor can use this information to inform their decision-making about patient care. Big data has the ability to enhance patient outcomes, lower costs, and boost system efficiency in the healthcare industry overall.
- Symptoms of Fraud: Big data can be used to spot patterns and irregularities in patient claims data, which can be used to uncover healthcare fraud. A big data analytics platform, for instance, can be used to find patients who are filing an abnormally high number of claims or to compare patient claims data with medical records to find discrepancies. The Trend wise Analytics article "Healthcare Fraud Management using Big Data" goes into detail about this. The authors of this paper explain how patterns of fraudulent activity in Medicare claims data were found using a big data analytics platform. Medicare was able to save almost $1 billion thanks to the platform's ability to recognise and flag questionable claims. All things considered, big data is an effective instrument for identifying and stopping healthcare fraud.
- Enhanced Patient Care: Electronically recorded health information is very helpful in compiling demographic and medical details like a boy's examination, clinical details, health issues, and circumstances that would further assist. Quality procedures are provided by healthcare providers. Benefits to patient care derived from using Big. A research study refers to data in the healthcare industry.
- Boost your operational expertise: Healthcare organisations employ Big Data as a business strategy to examine past patient hospitalisations as well as staff competency. Predictive statistics allow healthcare organisations to reduce costs without sacrificing quality of care. Big Data is also helpful in the healthcare industry because it lowers re-entry and decreases medication errors by enhancing economic and managerial competence.
- Finding a treatment for diseases: The healthcare sector can benefit from the application of big data analytics (BDA) to enhance patient care, lower expenses, and boost productivity. BDA can help patients create personalised treatment plans, make decisions based on health concerns, act early to prevent problems, continuously monitor their health, and enhance their quality of life. BDA can be used by hospitals to predict patient outcomes, identify patients who are likely to stay longer, and make better decisions. BDA can be used by insurance units for fraud analysis, detection, and prevention. BDA can be used by pharmaceutical companies to produce drugs and medicines more rapidly and effectively. All things considered, BDA has the power to completely transform healthcare by increasing its effectiveness, efficiency, and personalization.

- Cut down on expenses and wait times: Big data has the potential to lower healthcare costs and wait times. Predictive analytics, for instance, is used by one French hospital to control staffing levels. By doing this, the hospital can make sure that it has enough personnel on hand to handle patient demand, which can reduce costs and raise patient satisfaction. Although there are numerous other ways to use big data to cut costs, very few hospitals are doing so. This is due in part to the complexity of hospital finances and the reluctance of certain hospitals to make big data investments. The possible savings are substantial, though. For instance, big data has been used by four hospitals in Paris to predict admission rates. They have been able to effectively allocate doctors and resources as a result.
- Availability of treatments, like chemotherapy: Big data can be utilised in the healthcare industry to forecast workload and schedule appointments, resulting in cost savings and increased efficiency. Without information science, this is hard to accomplish. For instance, it is challenging to book appointments for 70 patients in a centre with only 35 chairs. In order to guarantee that there are adequate hospital beds for admitted patients, big data can also be utilised to manage hospital beds. Every day, this difficult task is completed in the hopes of creating a plan.
- Reminder Attention: In order to avoid hospital overcrowding, big data can also be used to identify patients who are likely to adhere to their doctor's recommendations. Recently, there has been a lot of interest in wearable technology that tracks users' heart rates and physical activity. The physical well-being and involvement of patients in their care can both be enhanced by this data. Additionally, it can be used to track health and lower the chance of developing diseases like high blood pressure and asthma. Numerous applications are available that monitor a patient's vital signs, medication compliance, amount of movement, and other activities. Countries such as Japan are merging robotics and healthcare due to an ageing population. Robots are being used in many different contexts, such as assisting elderly patients living alone, assisting physicians in providing medical care to patients in remote areas, and assisting animals suffering from Alzheimer's disease.

7. Challenges

- Privacy: Perhaps the biggest drawback is the absence of intimacy and seclusion. Big Data needs access to almost everything in order to function effectively, including social media accounts and private audio files. However, there is a cost associated with disclosing personal health information in order to address health issues. However, a patient is not granted freedom. Regulations protecting the privacy of medical records do exist, but they are not given much weight because it is generally agreed that information about an individual's health should not be restricted.
- Substitute Medical Personnel: Big data has the benefit of allowing for the early detection of potential health issues, but it also carries a significant risk: the replacement of doctors. Big Data is not so powerful as to be used in isolation from human intervention, but as its application grows, there is concern that patients may become less likely to visit physicians and instead turn to technology, undermining their authority. Big Data in the healthcare industry is unavoidable since an increasing number of institutions and businesses are investing in this expanding sector. However, one should take into account its drawbacks and recognise that the process is safe for both physicians and patients. Discussions about whether Big Data will result in the replacement of medical personnel have surfaced on prominent websites such as Forbes and Fortune.

8. Limitations

- Opposition to change: When it comes to technology adoption, the healthcare system lags behind other industries like the banking and oil and gas sectors. This is brought on by a lack of trust, legal concerns, insufficient administrative support for information technology (IT) and related practise changes, and a shortage of suitable and necessary skills for using ICT tools. For this reason, the healthcare system continues to use the paper-based system. However, the healthcare system's integration of data from various sources is not supported by the paper-based system.
- Data on healthcare is fragmented: The fragmentation and dispersion of data, which are stored in proprietary heterogeneous systems across healthcare organisations, exacerbates the challenges posed by big data in the industry. Healthcare data are also kept in antiquated systems with limited interoperability, most notably electronic medical record systems. However, because the underlying data is composed of various schemas, formats, metadata, and standards, integrating health data to create big data is challenging.
- Ethical difficulties: Effective information exchange between patients and healthcare providers is hampered by ethical issues like data privacy, confidentiality, control over access to patient information, commercialization of de-identified patient information, ownership and governance of patient information, and so on. As a result, integrating healthcare data from various sources becomes difficult. As a result, it becomes difficult to quickly obtain a thorough and comprehensive picture of a patient while they are receiving care.
- Expansion of healthcare guidelines: Standards are established guidelines that enable disparate platforms, tools, systems, and technologies to function together. Healthcare facilities, however, don't follow one set of rules. For example, different hospitals have different titles and codes for drugs, diseases, examinations, and case reports. Therefore, interoperability between heterogeneous systems is hampered by the absence of a common standard in the healthcare system.
- Privacy and security concerns: Security and privacy concerns are one of the main obstacles to the integration of various healthcare information sources into big data. Security risks to patient data include improper disclosure, unauthorised use, and unauthorised destruction of patient information. These threats can affect healthcare data.

9. Resolutions

- Policies for Security and Privacy: Healthcare data should be sufficiently secured and protected, according to big data analytics developers and providers. When analysing big data in healthcare, tools that guarantee the privacy, availability, and integrity of protected health information should be employed. Furthermore, user authentication, data encryption, application security, and physical security should all be used to sufficiently protect healthcare data. Additionally, audit trail system use ought to be promoted.
- Big Data Analytics Usability: User interfaces for big data analytics should be well-designed. Additionally, big data analytics should be created with ease of learning and use in mind.
- Creation and Acceptance of a Common Standard: For big data to be successfully implemented in healthcare, a standard vocabulary and terminology—that is, a shared language for describing medical terminology—must be established and adopted. This is done to guarantee the sharing, reuse, and consistency of medical data.

10. Tools and techniques for analysing Big data in Healthcare

- Healthcare data has increased dramatically as a result of the development of information and communication technology, particularly the Internet of Things (IoT). Modern data architectures are unable to effectively handle the volume and complexity of this data. Big data analytics is the analysis of large amounts of data in the healthcare industry using a variety of complex tools, platforms, technologies, and applications. Some common instances of big data analytics are as follows:
- Google Big Query: Google Big Query stores and queries enormous datasets in a matter of seconds by utilising Google's cloud infrastructure. Google Big Query employs multiple security layers to safeguard its data.
- Map Reduce: A software tool called Map Reduce consists of the two fundamental operations Map and Reduce. When it comes to distributing subtasks, Map offers an interface, and Reduce gathers work and resolves results into a single value.
- Jaql: Declarative and functional, Jaql is a query language built to handle massive amounts of data. To enable parallel processing, Jaql leverages Map Reduce tasks to translate high-level queries into low-level queries.
- Hadoop: Massively parallel server clusters using Hadoop, an open-source software framework, are used to process massive volumes of data. It is a database management system that is not relational. The Hadoop Distributed File System (HDFS), which distributes data among several servers, is the main part of Hadoop. HDFS is used to store a variety of data types and structures.
- Non-Relational Databases: Relational databases use Structured Query Language (SQL) to access data that are organised into rows and columns. Non-relational databases, on the other hand, do not use SQL to retrieve data. Rather, key-value pairs that serve as links to the locations of files on discs are used to store and retrieve data. Strong security-ensuring mechanisms are lacking in non-relational databases. As a result, they cannot be extensively used in healthcare settings.
- Cloud Based Services: Virtualization of computer resources via the Internet is made possible by the cloud. There is also sufficient access to computer resources thanks to the cloud. Software-as-a-service (SaaS), platform-as-a-service (PaaS), and infrastructure-as-a-service (IaaS) are three examples of the various types of cloud-based services. Hardware, operating system, application software, and storage can all be provided by a vendor thanks to SaaS. Similar to Google App Engine, Microsoft Windows Azure, and Oracle Cloud Computing, PaaS offers a fundamental platform. The vendor is able to offer the raw processing and storage capacity thanks to IaaS.

11. Applications

- Applications of Big Data in 'Omics': Large molecular data sets, including those from proteomics, metabolomics, and genomics, are referred to as "mics data." Utilising big data analytics, this data can be analysed to find patterns and trends that can be leveraged to create novel approaches to disease prevention and treatment, as well as to provide patients with individualised care. The study of genes and their functions is known as genomics. Gene variants linked to diseases can be found using big data analytics, and new treatments and diagnostic procedures can be created using this knowledge. The study of proteins and their roles is known as proteomics. Proteins implicated in disease processes can be found using big data analytics, and new medications that target these proteins can be created. The study of metabolites, or the tiny molecules produced by cellular metabolism, is known as metabolomics. Big data analytics can be utilised to find metabolites linked to specific diseases as well as to create new biomarkers for early disease detection and diagnosis. All things considered, big data analytics holds the potential to completely transform healthcare by facilitating the creation of more efficient and customised disease treatments.

- Insurance Sector / Recipient: Medical Care Big data is being used by insurance companies and payers for claim management, fraud deduction, and underwriting. Insurance companies are keeping an eye not just on algorithmic fraud but also on person-centric practises that expose claim-centric behaviour. For instance, the number of related claims filed by the same person or stating the same treatment in several insurance companies.
- Medical Device Development and Production: A larger range of device materials, delivery techniques, tissue interactions, and anatomical configurations can be assessed thanks to the use of big data. Big Data and calculation techniques can be very important to the manufacturing and strategy of medical systems.
- Pharmaceuticals: Big data is employed in every stage of the development of pharmaceuticals, but it is especially useful for drug discovery. Pfizer recently launched the Precision Medicine Analytics Environment programme, which connects clinical trial, genomic, and electronic health record data to find opportunities to quickly deliver novel medicines for specific patient populations.
- Healthcare with Personalised Patient Care: The best and most updated patient care will be made possible by big data. In the near future, new influences from big data will drive appropriate updates to clinical guidelines, patient triage, and diagnostic aids to enable more individualised and tailored treatment to improve patient outcomes.

12. Conclusion

Large volumes of complex, variable, high velocity data are referred to as "big data," and managing, storing, distributing, and analysing them calls for sophisticated methods and tools. To economically extract values from very large volumes of a wide variety of data is the fundamental aim of big data. Big data is needed in the healthcare system to enable smooth communication between patients and healthcare providers, boost patient involvement in the treatment process, produce evidence-based care, and aid in the early detection of fraud and security threats. Nevertheless, despite the many advantages of big data in healthcare, obstacles to its effective adoption include security concerns, the longitudinal nature of health information, reluctance to shift from traditional modes of care to ICT-based care, and a lack of standards in the field. Thus, this paper proposes that enhancing the use of big data in healthcare will require the development of efficient tools with well-designed interfaces, the adoption of security-based policies, and the application of a standard vocabulary.

13. References

1. J. W. Cortada, D. Gordon, B. Lenihan, The value of analytics in healthcare: From insights to outcomes, IBM Global Business Services, Executive Report, 2012.
2. T. Huang, L. Lan, X. Fang, P. An, J. Min, F. Wang, Promises and Challenges of Big Data Computing in Health Sciences, Big Data Res. 2 (2015) 2–11. doi:10.1016/j. bdr.2015.02.002.
3. G. O. Matheson, M. Klugl, L. Engebretsen, Prevention and management of non-communicable disease: the IOC consensus statement, Clinical Journal of Sport Medicine. 23 (2013) 419–429.
4. K. D. Moore, K. Eyestone, D. C. Coddington, The big deal about big data, Healthcare Financial Management. 67 (2013) 60–68.

5. T. Schultz, Turning healthcare challenges into big data opportunities: A use-case review across the pharmaceutical development lifecycle, Bull. Association Inform. Sci. Technol. 39 (2013) 34-40. doi: 10.1002/bult.2013.1720390508.
6. W. Raghupathi, and V. Raghupathi, “Big data analytics in healthcare: promise and potential,” Health Information Science and Systems, vol. 2, pp. 1-10, 2014.
7. A.E. Youssef,” A framework for secure healthcare systems based on big data analytics in mobile cloud computing environments,” International Journal of Ambient Systems and Applications, vol.2, pp. 1-11, 2014.
8. K. Priyanka, and N. Kulennavar, “a survey on big data analytics in health care,” International Journal of Computer Science and Information Technologies,” vol. 5, pp. 5865-5868, 2014.
9. C.J. McCall, “Big data in healthcare: Hype or hope,” A White Paper by Carol J McCall, 2015.
10. O. Iroju, A. Soriyan, I. Gambo, and J. Olaleke, “Interoperability in healthcare: benefits, challenges and resolutions,” International Journal of Innovation and Applied Studies, vol. 3,pp. 262-270, 2013.
11. A. Gandomi and M. Haider, Beyond the hype: Big data concepts, methods and analytics, International Journal of Information Management, vol. 35, no. 2, pp. 137–144, 2015.
12. A. O’Driscoll, J. Daugelaite, and R. D. Sleator, “Big Data”, Hadoop and cloud computing in genomics, Journal of Biomedical Informatics, vol. 46, no. 5, pp. 774–781, 2013.
13. X. Wu, X. Zhu, G. Q. Wu, and W. Ding, Data mining with big data, IEEE transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97–107, 2014.
14. J. A. Patel and P. Sharma, Big data for better health planning, in Proc. International Conference on Advances in Engineering and Technology Research, 2014, pp. 1–5.
15. A. E. Youssef, A framework for secure healthcare systems based on big data analytics in mobile cloud computing environments, International Journal of Ambient Systems and Applications, vol. 2, no. 2, pp. 1–11, 2014.
16. D. Blazquez and J. Domenech, "Big Data Sources and Methods for Social and Economic Analyses.," Technological Forecasting & Social Change, pp. 100- 113, 2017.
17. R. Montoliu, J. Blom and D. Gatica-Perez, "Discovering places of interest in everyday life from smartphone data.," Multimedia Tools Appl. 62, pp. 179-207, 2017.
18. R. Herschel and V. Miori, "Ethics & Big Data," Technology in Society, pp. 31-36, 2017.
19. J. Lee and M. Kang, "Geospatial Big Data: Challenges and Opportunities.," Big Data Research, pp. 74-81, 2015.
20. H. Liu and G. Guo, "Opportunities and Challenges of Big Data for the Social Sciences: The Case of Genomic Data.," Social Science Research, pp. 13-22, 2016.