Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2019;5(3): 194-203

Review Article

Big Data in Healthcare: A to Z

Elham Nazari[1], Elias Ameli[2], Hamed Tabesh[1]*

[1]Department of Medical Informatics, faculty of medicine, Mashhad University of Medical Sciences, Mashhad, Iran.
[2]Department of Computer Engineering, Azad University, Mashhad, Iran.

ARTICLE INFO | ABSTRACT

Received 05.02.2019
Revised 07.05.2019
Accepted 19.05.2019
Published 01.08.2019

Key words:
Big data;
Healthcare;
Datamining

Background & Aim: Today, with the advent of technology, due to the growing data in the field of health care, it is difficult to manage and analyze this type of data known as the Big Data. This analysis has many capabilities to improve the quality of care, reduce errors and reduce costs in care services.

Methods: This study is based on search of databases (PubMed, Google Scholar, Science Direct, and Scopus). This investigation has done with the websites and the specialized books with standard key words. After a careful study, 50 sources were in the final article.

Results: Since the Big Data Analysis in the field of health has been growing and also considered in recent years, this survey identified the necessity of these analyses, the definition of the Big Data, the benefits, resources, architecture, applications, analysis, platforms, Examples and challenges in the field of health care.

Conclusions: Familiarity with the big data concepts in the field of healthcare can help researchers in conducting applied research and thus improve the quality of health care services and reduce costs.

Introduction

Today, with the advent of various technologies in the healthcare industry and the use of information systems such as EHR (Electronic Health Record) for storing information, the production of digital data in this area is increasing [1]. The California Health Network believes its EHR system has 26.5 to 44 petabytes (10^{15} terabytes) [2]. This type of data, in addition to its large volume, has other features and known as Big Data. Big Data in Healthcare refers to an electronic dataset that is large, sophisticated, varied and has a fast producing speed and cannot be manage with old software/hardware [2-4]. These analyzes are very important in the field of health care and have become a hot topic, as population aging, increasing the prevalence of chronic diseases and increasing the cost of medical technologies have

* Corresponding Author: Tabeshh@mums.ac.ir

Please cite this article in press as Nazari E, Ameli E, Tabesh H. Big Data in Healthcare: A to Z. J Biostat Epidemiol. *2019; 5(3): 194-203*

increased the magnitude of the cost of public health care and long-term care. As a result, these analyzes reduce costs [5], improving the quality of health care processes, including prevention, diagnosis, sharing of information, discovery of knowledge and the use of evidence-based medicine, will improve decision making [3,6-8].

Therefore, given the importance of Big Data Analysis in health care, which has been very much considered in recent years, this survey introduced the services of health care, the need for Big Data Analysis, how Big Data is formed, Big Data Definition, benefits, resources, architecture, applications, analysis, platforms, examples and challenges in the field of health care.

Health care field

The health care field has always been serving various levels of care, including prevention, diagnosis and treatment. Today, provision of optimal services by healthcare providers, including physicians, nurses, insurers and others, has come with introduction of new technologies. For this reason, health care costs have risen more than before and are rising with age of population and chronic diseases [9, 10]. On the other hand, it is always emphasized on diagnosis and treatment using updated clinical data and Evidence-base [11]. Also, given the fact that individual characteristics of each patient, such as family factors, lifestyle and factors affecting his life in the process of diagnosis and treatment should not be overlooked by importance of Personalize medicine [12, 13].

Need for Appropriate Medical Decision Making and Big Data Analysis

Decision making in some areas is more decisive; For example, in a field like medicine, which is related to human life, decision-making process should be done with sufficient accuracy and as far as possible without making mistakes so that the results of a correct decision can be taken to improve the quality of health care [14]. In the meantime, the importance of group decision making in medical profession should not be ignored, because in many cases, this kind of decision will help to provide useful treatment to patients [15-17]. Because this type of system provides a shared system to health care providers and the quality of health care services is facilitated and improved, and thus savings in healthcare costs for caregivers and patients [14]. This will not be possible except by performing a Big Data Analysis and creating a shared database. The question now is why the Big Data Analysis is essential for medical applications? Big Data extracts useful patterns from a large amount of data and can use these analytical results to make decisions. Since any decision in health care is related to human life, analyzing it can have a great impact on improving quality of care and reducing costs. It is also possible to share information to make a decision about the appropriate diagnosis between different specialists [18-21]. Consequently, new evidence for diagnosis and treatment should be used. This new attitude will require and demand much to make Big Data analysis [6, 2].

How Do Big Data Appear in Health Care?

Today, healthcare industry, like other industries, uses information and communication technologies such as computers, Internet and other modern equipment to improve care services. For example, today's patient information is not same as the past, and hospital information systems and electronic health record system are used to store and retrieve their digital and lifelong information [22, 23]. Computers, robots and other electronic devices have come for diagnosis and treatment to help the health care providers. For example, medical imaging with MRI (Magnetic Resonance Imaging), CT (Computed Tomography) is used and CAD (Computer Aided Diagnosis) is used to analyze these images [24]. Apps and games on cell phones and tablets can help patients track their illnesses, especially in chronic diseases and rehabilitation [25, 26]. Internet provides a rich source of information in form of social networks and recommender systems to patients and physicians.

Therefore, considering the expansion influence of technology and the need for care providers to provide higher quality services with the help of these technologies, they generate massive amounts of data in this area and are constantly increasing [27-30].

Big Data Definitions in the Field of Health Care

- Big data in the field of health care includes a set of digital data related to the field of care, which is a large and complex set of data, and difficult to manage with old software and hardware [2].
- Big data in the field of health has features such as high volume, a diverse clinical environment, and a health connection at one or more points [31].
- In the area of health, Big Data refers to complex data of a large size that is difficult to extract knowledge.
- Big Data in the field of health refers to data in different formats and come from different sources [8].

Big Data Benefits in Health Care

The benefits of Big Data in healthcare include specifically improve operational efficiencies, help predict and plan response to disease epidemics, improve the quality of clinical trials and optimize healthcare spending at all levels from patients to hospital systems to governments [1]. The major benefit of managing Big Data in health care is to improve the quality of care and efficiency, including:

- Big Data has the capacity to generate new knowledge. Managing undocumented medical data using computational techniques such as natural language processing will provide valuable information.
- Big Data can transfer the basic medical information of each individual to clinical sites. For example, a biology system will link to EHR data.
- Big Data can help transfer knowledge. Due to the difficulty in accessing resources, physicians prefer to stay in their current state of information. By digitizing the medical texts, access to the manuals and related material is facilitated, but sort of information becomes very complex for finding patients with a few chronic illnesses to find the right treatment method. This problem can help physicians make a correct decision by conducting analyzes on the EHR and designing a dashboard.
- Big Data, by transmitting health care information to patients, enables them to play a greater role in the care of their illness. It also improves the data in medical records, linking the patient's previous data, such as the drug list and family history, and personal data

such as eating habits, exercise schedules, education and income [6].

Applications of Big Data Analysis

Big Data on providing patient-centered services, early diagnosis, hospital quality monitoring, improved therapeutic approaches, as well as in the field of information technology (IT) infrastructure, operational, organizational, managerial and strategic area Health is applicable [33]. The Big Data Analysis is useful in supporting information, admin and delivery, behavior / consumer, clinical decision support, clinical information [34]. Also, in the context of individual analytics, it has been applied for a patient profile, personalized medicine, performance-based pricing for staff, and analysis of disease patterns, improving public health, PHR (Personal Health Record) and EHR [3]. The EHR has a huge amount of data, which is very valuable, including quantitative data (laboratory data), qualitative data (textual and demographic data), and transaction data (drug delivery records). Managing these data is done using computer science techniques, especially machine learning. Most data are in non-structured formats such as textual data that make up the basis for better decision-making in medicine [6]. By using Big Data Analysis, multiple expert decisions can be made at the same time in digital health analysis. With real-time analysis, large volumes of rapidly growing data are being collected from in-hospital and in-home devices for monitoring safety and predicting adverse events. EBM (Evidence Base Medicine) is used to provide effective care and to provide models for anticipating R & D in medications and discovering side effects before entering the market and treating [2].

Big Data sources in health care

Big Data analysis in the field of health care includes management data, test results, Medical images (x-ray, MRI image etc.), history, diet, medications and surgical instruments, biomedical signals (EEG, ECG, EMG, etc.), physician note, genetic information and physician information. These data are available from internal sources such as EMR (Electronic Medical Record), CDSS (Clinical Decision Support System), CPOE (Computerized Physician Order Entry) and EHR. It also comes from external sources such as government resources, insurance companies, laboratories, HMO (Health Maintenance Organization), Web and social media (from Facebook, Twitter, LinkedIn) or data from apps and smartphones, sensors, and biometrics (finger, genetic, handwriting). These data have different formats (flat files, .csv, relational tables, ASCII / text, etc.) and from different places (geographic locations, different health care providers, different applications, databases, and various transactions) [8,35,32,2].

Architecture of the Big Data Functional Steps in Health Care:

An example of Big Data Architecture in the field of healthcare is shown below:

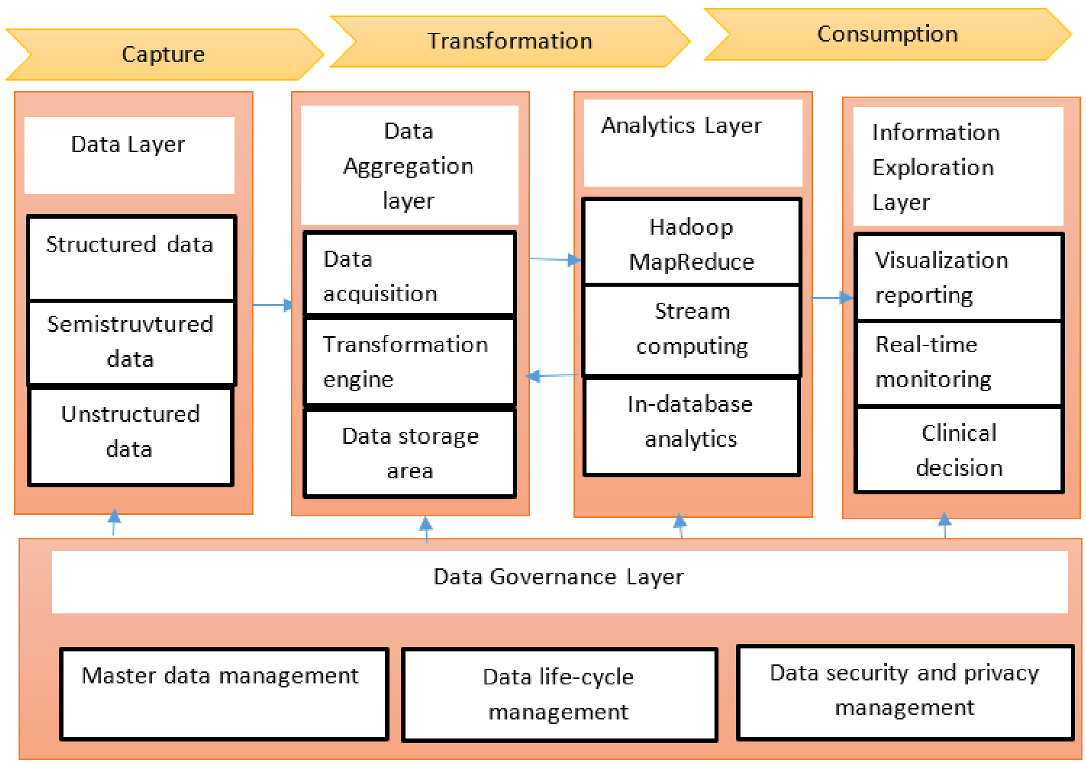

Figure 1 - a sample of big data architecture for healthcare [35]

As is clear from the figure, the stages of the Big Data analysis include three stages of capture, transformation and consumption.

Capture

The data layer is included in health care at the capture stage and includes three types of structured data, semi structured data and unstructured data that cover each category of data as follows:

- Structured Data: This type of data is displayed in predefined format, such as titled column and rows (such as database records). This type can be easily processed using traditional data processing tools (for example, using a relational database, files, etc.)
- Semi-structured Data: This type of data does not follow the relational database format. Includes tags that isolate semantic elements, including emails and HTML (Hyper Text Markup Language) and XML (Extensible Markup Language) documents.
- Unstructured Data: This type of data is not predefined and is not organized predefined. Examples of this type of data include text, video, audio, images and social media.

Transformation

The data aggregation layer is responsible for obtaining data from diverse data sources and converting it into a specific standard for data analysis. The analytics layer of the transformation phase is responsible for performing the appropriate analyzes on the aggregated data based on the analytical objectives.

Consumption

How to use and consume after a data analysis step that is visual, real-time, and clinical decisions. The data management layer is the component that manages other logical layers. Guarantees the availability of data for processing and privacy of the owners [35].

Big Data Analysis

Clustering analyzes are used to determine a high-risk group of obesity. Data mining techniques are also used to monitor abnormalities and identify effective factors. The statistical techniques, graph analysis is used to assess the quality of hospital efficacy. Also, machine learning is used to predict a disease risk, from NLP to improve the effectiveness of care and control costs, predictive future illness and treatment improvement. [36] In addition, data retrieval techniques such as text mining and semantic analyzes, as well as simulations and analyzes of survival, optimization, visualization, virtual reality, pattern recognition and expert system of AI are also used. The methods of regression, Decision Tree, NN(Neural Network), association rule mining, PSO(Particle Swarm Optimization), Deep learning and the fuzzy logic are used [35,8,37,38].

The Common platform

In some studies, Map Reduce and Hadoop have been used for health care analyzes. It is a distributed platform used to refine protein structure alignments with a high degree of precision, also to improve neural signal processing. For discovery drug, clinical research, genomics, Precision Medicine and the prognosis of diseases such as diabetes, heart disease and the outbreak of the flu, [36] and the provision of general health services were used [8,39] from other platforms, such as spark and flink, are also due to the type of analysis can be used [40,41].

Examples of Big Data Analysis

- 94% of US hospitals have adopted the EHR. McKinsey, in a report on Big Data in health care, points out that aggregated systems have an improved outcome in cardiovascular disease and cost $1 billion to reduce spending on visits and test labs.
- The Michigan healthcare system conducted blood transfusion management using the Big Data analysis, with a 31% reduction in blood transfusion and $ 200,000 in spending per month.
- In the Rizzoli Orthopedic Institute in Bologna, Italy, Big Data analysis was used to diagnose orthopedic diagnosis, resulting in a 30% reduction in hospitalization rates and a 60% reduction in imaging tests.
- At Harvard Medical School and Harvard Pilgrim Health Care, doctors have proven that EHR has the potential to identify and group diabetic patients into two categories for monitoring general health care using the Big Data analytical techniques.
- In the business processes, Blue Cross Blue Shield of Massachusetts (BCBSMA) used Big Data analysis methods to make financial and medical decisions to help with decision making. By this, risk groups were identified, the risk was minimized and the outcomes of the patient improved. For example, this preventive therapeutic protocol was introduced in the group of patients with high cholesterol and was resisted against cardiovascular disease.

- In Sick Kids in Toronto, Big Data analysis methods were used to improve the outcomes of childhood exposure to nosocomial infections. Advanced analysis methods based on collected data from bedside monitoring devices were used to identify potential symptoms of infection [2].

Here's an example of Big Data applications:

In 2012, the Obama administration announced more than $ 200 million to research and develop tools and methods for accessing, organizing and collecting data from massive digital data at National Institutes of Health, Department of Defense, Department of Energy and the United States Geological Survey [7]. It also published a report on this type of data, in which the McKinsey Global Institute (MGI) lists the job opportunities that make Big Data in some areas. These areas are as follows:

- a potential value of $300 billion in the US health care
- $149 billion in European government administration or improving the operating margin of retailer companies by 60 percent [42].

According to a report, Big Data analysis in health care saves $ 450-350 billion. $ 70-100 billion in disease management and prevention (Right living), 90-110 $ billion in medical sector and use of Evidence-base treatment for each patient to improve the outcomes of patient and reduce medical error (Right care), $ 50-70 billion to improve provider selection tailored to patient's needs and lower acceptance rates (Right provider), 50 - 100 $ billion in determining the cost-effectiveness of health care through various methods such as Outcome-patient reimbursement and elimination of fraud and misuse (Right value), $ 40-70 billion to improve and innovate in healthcare (Right innovation), which shows the necessity and importance of performing these analyzes in health care [43,44].

Challenges

Despite the great benefits, health care industry is still at the nascent of Big Data analysis. There is not enough knowledge about which data should be used for the purpose. There is no proper IT infrastructure. There is not enough knowledge about which algorithm is appropriate and what tools are suitable for analysis. An expert on interpretation of Big Data Output does not exist, which leads to an error-based conclusion. The other challenge is the great variety of data, scalability, missing data, uncertainty, fuzziness, complexity and high-speed production. The security, privacy and trust issue, cost, data access and sharing and lack of data quality are another challenge in Big Data [8, 3, 45, 46].

Conclusion

Nowadays, with the growing trend of data generation in the field of health, the Big Data Analysis has been given much attention to reduce costs and the improvement of the quality of care services. These analyzes have numerous applications in various areas of health care such as prevention, diagnosis and treatment. Many studies indicate the importance of these analyzes in the field of health care. Recently, attention has been paid to various aspects of the area of health including medical image processing, medical signal processing, medical diagnosis, primary care, rehabilitation, genetic data analysis, recommendatory systems for controlling chronic diseases, etc. [27, 47-51]. At the same time, it faces challenges such as lack of data quality, lack of expert knowledge, lack of knowledge of the method and tools that should be taken into consideration to extract useful information from this type of data. Also, this type of data has certain characteristics such as diversity, complexity, and high-speed production, it requires the use of appropriate platforms for managing these features. Future studies can examine different platforms and use appropriate tools for analysis. While the Big Data analysis is developing, rapid advances in platforms and tools can accelerate the maturity process. It is also important to use the appropriate analysis method to provide useful results. One of these techniques used to manage scalability and uncertainty in the Big Data is the Decision Fusion method, which has many applications in computer science, mathematics, statistics, Machine learning and data mining. The development of this technique is necessary to combine various information from diverse sources Also, providing frameworks in this area can be very helpful.

Reference

1. Nambiar R, Bhardwaj R, Sethi A, Vargheese R, editors. A look at challenges and opportunities of big data analytics in healthcare. Big Data, 2013 IEEE International Conference on; 2013: IEEE.
2. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health information science and systems. 2014;2(1):3.
3. Sagiroglu S, Sinanc D, editors. Big data: A review. Collaboration Technologies and Systems (CTS), 2013 International Conference on; 2013: IEEE.
4. Duggal PS, Paul S, editors. Big data analysis: challenges and solutions. International Conference on Cloud, Big Data and Trust; 2013.
5. subgroup TH. Big Data Technologies in Healthcare;Needs, opportunities and challenges. 2016.
6. Murdoch TB, Detsky AS. The inevitable application of big data to health care. Jama. 2013;309(13):1351-2.
2. 7.Chen M, Mao S, Liu Y. Big data: A survey. Mobile Networks and Applications. 2014;19(2):171-209
7. Mehta N, Pandit A. Concurrence of big data analytics and healthcare: A systematic review. International journal of medical informatics. 2018; 114:57-65.
8. Hermon R, Williams PA. Big data in healthcare: What is it used for? 2014.
9. Dexter PR, Miller DK, Clark DO, Weiner M, Harris LE, Livin L, et al. editors. Preparing for an aging population and improving chronic disease management. AMIA Annual Symposium Proceedings; 2010: American Medical Informatics Association.
10. Lewis SJ, Orland BI. The importance and impact of Evidence Based Medicine. Journal of Managed Care Pharmacy. 2004;10(5 Supp A):S3-S5.
11. Redekop WK, Mladsi D. The faces of personalized medicine: a framework for understanding its meaning and scope. Value in Health. 2013;16(6):S4-S9.
12. Harvey A, Brand A, Holgate ST, Kristiansen LV, Lehrach H, Palotie A, et al. The future of technologies for personalised medicine. New biotechnology. 2012;29(6):625-33.
13. Sox HC, Blatt MA, Higgins MC, Marton KI. Medical decision making: ACP Press; 2007.
14. Oshima Lee E, Emanuel EJ. Shared decision making to improve care and reduce costs. New England Journal of Medicine. 2013;368(1):6-8.

15. Sakkalis V, Zervakis M, Micheloyannis S, editors. Biopattern initiative: towards the development and integration of next-generation information fusion approaches. Engineering in Medicine and Biology Society, 2004 IEMBS'04 26th Annual International Conference of the IEEE; 2004: IEEE.
16. Rahman MM, Bhattacharya P. An integrated and interactive decision support system for automated melanoma recognition of dermoscopic images. Computerized Medical Imaging and Graphics. 2010;34(6):479-86.
17. Alemi F, Gustafson DH. Decision analysis for healthcare managers: Health Administration Press; 2007.
18. Rubinstein A. Modeling bounded rationality: MIT press; 1998.
19. Lee S, Lebowitz S. 20 cognitive biases that screw up your decisions. Business Insider. 2015.
20. Grandi U. Social choice and social networks. Trends in Computational Social Choice AI Access. 2017:169-84.
21. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health affairs. 2011;30(3):464-71.
22. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. International journal of medical informatics. 2008;77(5):291-304.
23. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Computerized medical imaging and graphics. 2007;31(4-5):198-211.
24. Scacchi W, editor Computer Games and Virtual Worlds: New Modalities of Rehabilitation and Therapy. Orange County Stroke Rehab Network-Continuing Education Workshop; 2011.
25. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. Pharmacy and Therapeutics. 2014;39(5):356.
26. Terveen L, Hill W. Beyond recommender systems: Helping people help each other. HCI in the New Millennium. 2001;1(2001):487-509.
27. Melville P, Sindhwani V. Recommender systems. Encyclopedia of machine learning: Springer; 2011. p. 829-38.
28. Wiesner M, Pfeifer D. Health recommender systems: concepts, requirements, technical basics and challenges. International journal of environmental research and public health. 2014;11(3):2580-607.
29. Grajales III FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. Journal of medical Internet research. 2014;16(2).
30. Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, et al. Making sense of big data in health research: towards an EU action plan. Genome medicine. 2016;8(1):71.
31. Archenaa J, Anita EM. A survey of big data analytics in healthcare and government. Procedia Computer Science. 2015; 50:408-13.
32. Wang Y, Kung L, Byrd TA. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change. 2018;126:3-13.
33. Hermon R AHWP. Big data in healthcare: What is it used for? Australian eHealth Informatics and Security Conference. 2014
34. Palanisamy V, Thirunavukarasu R. Implications of big data analytics in developing healthcare frameworks–A review. Journal of King Saud University-Computer and Information Sciences. 2017.

35. De Mauro A, Greco M, Grimaldi M, editors. What is big data? A consensual definition and a review of key research topics. AIP conference proceedings; 2015: AIP.
36. Sun J, Reddy CK, editors. Big data analytics for healthcare. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining; 2013: ACM.
37. Benke K, Benke G. Artificial Intelligence and Big Data in Public Health. International journal of environmental research and public health. 2018;15(12):2796.
38. Zhang X, Pérez-Stable EJ, Bourne PE, Peprah E, Duru OK, Breen N, et al. Big data science: opportunities and challenges to address minority health and health disparities in the 21st Century. Ethnicity & disease. 2017;27(2):95.
39. 40.Ramírez-Gallego S, Fernández A, García S, Chen M, Herrera F. Big Data: Tutorial and guidelines on information and process fusion for analytics algorithms with MapReduce. Information Fusion. 2018; 42:51-61.
40. Ferranti A, Marcelloni F, Segatori A, Antonelli M, Ducange P. A distributed approach to multi-objective evolutionary generation of fuzzy rule-based classifiers from big data. Information Sciences. 2017; 415:319-40.
41. Bifet A. Mining big data in real time. Informatica. 2013;37(1).
42. Groves P, Kayyali B, Knott D, Van Kuiken S. The ‘big data’revolution in healthcare. McKinsey Quarterly. 2013;2(3):1-22.
43. Kayyali B, Knott D, Van Kuiken S. The big-data revolution in US health care: Accelerating value and innovation. Mc Kinsey & Company. 2013;2(8):1-13.
44. Jagadish H, Gehrke J, Labrinidis A, Papakonstantinou Y, Patel JM, Ramakrishnan R, et al. Big data and its technical challenges. Communications of the ACM. 2014;57(7):86-94.
45. Bossé É, Solaiman B. Information fusion and analytics for big data and IoT: Artech House; 2016.
46. Gunay O, Toreyin BU, Kose K, Cetin AE. Entropy-functional-based online adaptive decision fusion framework with application to wildfire detection in video. IEEE Transactions on Image Processing. 2012;21(5):2853-65.
47. Li Y, Porter E, Santorelli A, Popović M, Coates M. Microwave breast cancer detection via cost-sensitive ensemble classifiers: Phantom and patient investigation. Biomedical Signal Processing and Control. 2017; 31:366-76.
48. Guo P, Banerjee K, Stanley R, Long R, Antani S, Thoma G, et al. Nuclei-based features for uterine cervical cancer histology image analysis with fusion-based classification. Biomedical and Health Informatics. IEEE Journal of, PP (99). 2015
49. Fan Y, Yin Y. Active and progressive exoskeleton rehabilitation using multisource information fusion from emg and force-position epp. IEEE Transactions on Biomedical Engineering. 2013;60(12):3314-21.
50. Bussemaker HJ, Li H, Siggia ED. Regulatory element detection using correlation with expression. Nature genetics. 2001;27(2):167.