

Gulara I. Muradova

Azerbaijan Technical University, Baku, Azerbaijan
gularamu.aztu.edu.az

DOI: 10.25045/jpit.v07.i2.10

IMPLEMENTATION OF BIG DATA IN HEALTHCARE

The paper highlights Big Data application in medicine and biomedical research. It provides examples how analysis of Big Data can significantly reduce problems in healthcare. Perspectives of implementation of Big Data in healthcare are shown.

Keywords: *Big Data, Big Data analytics, personalized medicine, biomedicine, telemedicine.*

Introduction

Informatization of healthcare has recently received remarkable attention to improve of the efficiency and quality of delivering medical services. The quality of services provided to patients, the transparency of the health system and the efficiency of the medical staff is largely dependent on the applied technology. High quality medical aid - leads to improved quality of life and increases longevity.

Outdated methods of analysis and data collection do not meet modern requirements, subsequently, now, there are a strong interest in IT solutions in Big Data. There is a need for modern equipment that allows creating new medicines much easier and faster. Therefore, it is time for the devices that read the quantitative data and do not need further decoding. Now, an increasing number of decisions would not be based on experience and intuition, but on the data and analysis. The constant accumulation and analysis of large-scale data sets allows creating a model with which experiments can be carried out. Thus, the researchers, benefiting from the data, obtain the findings of scientific value as the result. By analyzing large amounts of data, and setting the pattern, it is possible to create an object model that will be used for various purposes. With the help of the model, we can study the properties already described the real object, and make probabilistic forecasts. The development of big data technologies and adaptation to new challenges is the purpose of studying Big Data. The efficiency of using large-scaled data will improve the operation of healthcare system, particularly selecting appropriate ways of treatment, diagnosis, etc. [1].

In order to cope with the huge amount of data, there is a need for new solutions. Big data is one of such sources that provides new solutions to old problems and creates new questions. Big Data is a variety of approaches to solving problems with the use of specific tools of artificial intelligence (AI) on the large volumes of data. Big data analysis can improve the planning of the healthcare services. Big Data requires a review of the current approach to the new technologies. Big Data is the “new generation technology and architecture for an economic extraction of value from large volume of multiple data through rapid capture, processing and analysis.” Many industries use these technologies.

The following questions should be solved in order to improve the performance of the healthcare system:

- whether big data is something new in medicine?
- what are the benefits of analytics of health information systems?
- what can be said about the new field of medical research?
- how accurate the forecasting must be with the application of Big Data technologies, in order to use them as criteria to assess the data management models, in comparison with the existing criteria?
- how to realistically assess, and what constitutes a “significant” contribution to the healthcare system?

Medicine is one of the branches, which is becoming the consumer of Big Data technologies.

Numerous scientific conferences are widely publicizing scientific significance and prospects of big data technologies. Big Data challenges have been discussed at the IEEE 20th International Conference on Web Services (2013). Big Data in Healthcare Conferences (April 2015) held in Boston discussed the predictive analytics techniques. Moreover, the Conference on Big Data (2014), organized at Harvard University, mainly focused on the data confidentiality. The 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2014) covered the issues of predictive analytics in the era of Big Data: Workshop on Connected Health at Big Data Era.

Leading foreign journals describe advanced technology and efficient management solutions in a wide range of fields to handle big data. The Journal of Policy Analysis and Management (2014), a monthly journal Nature Biotechnology and many other medical journals are paying a great attention to big data in medicine. The Journal of Big Data Intelligence shows that Big Data analytics (BDA) in medicine has a great future. Information systems, such as electronic health records (EHR), computerized physician order entry (CPOE), picture archiving communications system (PACS), clinical decision support systems (CDSS) and Laboratory information systems, are being widely used in medical institutions, hospitals, clinics and health centers. Online magazines Data for Good, Journal of Electronic Resources in Medical Libraries, and others discuss the sources of data, methods of its storage and processing, the essence of analytical processing of Big Data and the areas of its practical application.

Computer technologies, particularly, large-scaled data technologies, are fundamentally changing the current medicine and creating a new systematic approach to diagnosis and treatment of patients. The case study aims at examining the current status of the use of Big Data technologies in the healthcare system throughout the world.

Big Data and medicine

Healthcare Institutes daily use large amounts of data related to health care, such as images, video, monitoring in real time. They require new methods of data collection, storage and processing. The first computer applications solved the problem of simple digital data generation. The minutely generation of a huge amount of new data from a single phone call to the doctor to the information load through social networks requires the modernization of health care, and suggests a close association of IT, science and clinical treatment to achieve the best clinical or preventive results. Decisions on the implementation of Big Data solutions are provided through the initial plan of action. Awareness of opportunities through extracting new knowledge from accumulated data includes “cultural level”. Scientific level is established by mathematical methods of data analysis that contribute to the discoveries in medicine, chemistry, biology and other disciplines. Technological level implies the reorientation of IT with the tasks as the main application for the collection, storage, processing of large amounts of data to support analytical methods [2, 3].

In this context, the following data groups can be distinguished:

1. Clinical data (recording of clinical institutions)
2. Genotype data
3. Social media data
4. Environmental data
5. Data recorded by a wearable device
6. Mobile data (from mobile devices of individuals).

Working with big data involves the following steps:

Step 1. Statement of the analysis problem. The vast amount of data requires new approaches in the control and analysis, since the use of traditional methods of data processing, in some cases, is even impossible.

Step 2. Accumulation of data from different sources discussed above.

Step 3. Initial Data processing. Data is checked for consistency and reliability. To provide data access of the analysts, the companies develop and maintain so-called storage. The data is stored in a database for subsequent analysis.

Step 4: Data Analysis. A large number of applications are used to gather statistics, which systemize the obtained data.

Step 5. Analytical results are presented to the end user as reports, charts, graphs, online recommendations or decisions.

Step 6. If the results do not make sense, the problems are reformulated, and the first step is started again.

Electronic form of work in health care facilities has made it possible to monitor the relationship between drug taking, its efficiency, and other factors that are also vital for the pharmaceutical companies. The main objectives of the use of big data are the reduction of time and costs spent for the development of new products and their delivery to the market by multifactor analysis of their effectiveness. Now, a doctor may rely not only on their own knowledge by studying the data obtained from a variety of professionals in real-time, who are involved in the treatment of patients with sporadic disease. This will enable much better analysis of the effectiveness of appointed treatments. The development of registers of patients with hereditary diseases to calculate the volume and the need for drugs, and the simplification of the diagnostics and search for hidden dependencies can be cited as the examples of using big data technologies. Many existing restrictions on the distribution of patient data impose certain restrictions on the possibility of such analysis [4].

Our lives will soon be filled with a variety of devices that constantly monitor our health. People are becoming more self-medicated. Many public information systems for self-diagnosis are developing. More and more people own wearable devices, and a variety of mobile devices are constantly evolving and increasing their functions (diabetic level, heart rate, pressure, activity, diagnostic tests, etc.). USB-ports may be embedded on healthy people for diagnostic purposes. In western countries, the tests have been already taken at home by the patients themselves. The clinical description is stored in the cloud, and the communication with a physician takes place remotely - via the Internet. Technological innovations – the gadgets enable the patients to save on costly post-operative rehabilitation [5]. With the help of a special headset that sends electrocardiogram data to the doctor, the patients, who have had a stroke or brain injury, may be provided with all-day consultations and testing sessions at home. Large amounts of data imply new demands on health care transformation.

Application areas of Big Data technologies in medical research

Big data allows implementing innovative solutions at low cost by providing coverage of a large virtual medium [6].

Modern medicine is becoming a science, in which large amounts of information about the patient provides the experts to conduct personalized treatment. This is a new model of organization of medical care, which is based on the selection of diagnostic, therapeutic and preventive medicines. These medicines are optimal for an individual patient, taking into account their genetic, physiological, biochemical, behavioral and other characteristics. On this basis, we can say that, now, any process taking place in the human body can be observed and measured. Medicine needs convenient tools for data processing, storage and transmission. Organization of medical care for people through rational drug selection is one of the vital areas of Big Data technologies. Big data analysis may facilitate the determination of the most effective and cost-effective treatment methods in the clinical aspect. Detection of the disease at an early stage makes it possible to simulate the future demand for the products of the pharmaceutical and medical sectors. Using large data capacity, the pharmaceutical companies are reducing the time and money spent for the development of new drugs when launching them at the market. These companies are also reducing

the costs spent to verify obviously unpromising components in the course of development of medicines. This is the basis for making decisions on the investments in research and development. The most important task of increasing the health rate is the analysis of the global patterns of disease in order to identify the trends of these patterns at an early stage [7, 8].

A 4-level examination with the use of bioinformatics, neuroinformatics, clinical informatics and informatization of the population is defined by the author [9] for each patient and the public as a whole to be better covered by the healthcare.

The globalization of the world has led the population to become more mobile, and geographic borders – to be increasingly blurred. Besides, for health professionals it is increasingly important to be able to access the data of visiting patients. The analysis and processing of large-scaled data, which is related to the specific regions and populations, will improve the planning of health services. Previously, medical facilities mainly dealt with individual companies, which developed and distributed hardware and software to them, and created “local” application solutions. Now, public institutions are actively invading this area, and they need access to patients’ data anywhere and anytime. The US have a program for the development of a regional healthcare infrastructure. InterSystems has participated in this program in New York. The established system includes the records of all interactions of 25-30 million patients with doctors. It is also important to recognize unstructured information of the human voice and its conversion into text. And, the data becomes very large, since a human speaks about three or four times more than he/she writes. InterSystems offers the technology to filter the raw data by extracting valuable information from it [10, 11]. This technology is also applicable to mentally disabled people. For example, it is possible to detect depression, as the depressive people tend to speak more slowly.

There is a huge interest in proteomics, both for early detection and monitoring of diseases, and for the development of more effective treatments throughout the world. The main challenge in this fast-growing area is to understand the mechanism of interaction of about 300,000 proteins in the human body. However, the amount of work to be implemented requires the use of high performance and informative methods and devices [1]. The use of Big Data technologies can provide a quick breakthrough in proteomics.

An analysis of the literature [11-27] identifies the following medical projects existing in the field of medicine:

Personal genome projects - a study in the field of personal genomics and personalized medicine.

Massachusetts Technical University - analysis of geodata for the prediction of socially significant diseases (vascular and heart disease).

InterSystems – a program for the development of a regional health infrastructure (quick access to 30 million patients’ data).

GenePool - sequencing and identifying biomarkers and systematizing received data stream.

Microsoft Research in Cambridge – technology for analysis and search of massive databases in the form of graphs.

Proteus Digital Health - research on the effectiveness of medicines, control over their intake, remote monitoring of the patient’s condition.

StationX - software for the genome decoding by searching site-markers, which are responsible for the definition of pathology.

Watson IBM Big Data system - analysis of specific symptoms, history of hereditary diseases, the synthesis of this data with the available unstructured and structured medical information, including medical books and articles.

Analysis of Big Data possibilities and prospects

Big Data analytics (BDA) is aimed at understanding the course of treatment of each patient. Progress can be expected in the field of genomics and personalized medicine for making medical

and control decisions. On the proposal of the US President Barack Obama to the leading scientific institutions of the USA adopted a ten-year research program, which is called “brain activity mapping”. 3.8 billion USD was spent on the mega-project “Human Genome”, and according to the government study, 800 billion USD has already returned to the state economy [12].

Genomics deals with decoding and determining the sequence of nucleotides in the DNA (deoxyribonucleic acid). This data is very immense and received in a variety of formats that are traditionally handled by different programs. BDA requires the allocation of supercomputer capacity and the creation of a specialized environment, which can “convert” the information into different omics data, integrate and analyze them on a single screen [13]. Automation of the main stages of “reading” DNA will provide people with cancer to undergo specialized tests that will help to appoint appropriate treatment in a short time [14, 15]. “Personal genome project” (PGP) developed at the Harvard Medical School uses a large data and sets the goal of sequencing and publishing the completed genomes and medical records of 100,000 volunteers to focus on the research of personalized genomics and personalized medicine [16]. The project is scheduled to perform decoding the personal genomes of volunteers’ sets. Their genetic data along with the medical records will be available for further analysis [17].

Computer analysis methods in comparative genomics technologies of Big Data is also used in bioinformatics. GenePool project involves sequencing and identification of biomarkers and systemizes all the huge data flow [18, 19]. Application of Big Data Technology, which decrypts this data set, also identifies the patterns in genomic world. Benefiting from Big Data technology, researchers have analyzed the genomes of many thousands of sick and healthy people. The system picks up the treatment based on the analysis of vast amounts of genetic code information. Thus, for example, Alzheimer’s disease depends on the genes involved in the development of multiple sclerosis and Parkinsonism [20-22].

In biomedicine, new big data technologies are used for the genetic architecture sequencing and aimed at improving personal health and the health of the population in general [23-25].

Currently, fundamental research in medicine focuses on the study of pathological processes of human diseases, and on the identification of new methods of diagnosis and the treatments of these diseases. Telemedicine is the use of telecommunications and information technologies to provide remote health care services. Telemedicine is convenient for the people living in remote regions. The author of [26] discusses how Big Data is transforming the telemedicine. The data obtained from smart devices allow real-time monitoring of chronic diseases to optimize the dosage of drugs and to improve the patients’ treatment results. The development of systems biology and molecular biomedicine, including hybrid and cellular technologies, has led to the need for organizing and developing the procedural aspects of the introduction (translation) of accumulated knowledge in the practice of the attending physician on the one hand, and in the field of pharmacobioeconomics - on the other [27-29]. Analysis of mHealth applications shows that mobile telemedicine has been implemented in healthcare systems. mHealth is realized in the cases of emergency response to the deterioration of the patient’s condition and when urgent measures are required [30].

Watson system of IBM Big Data platform defines the diagnosis basing on all the studied data. Watson for Clinical Trial Matching narrows the scope of all possible research and quickly identifies potential matches. It also performs the synthesis of incoming data containing available unstructured and structured medical information, and analyzes natural language [31].

Big Data technologies are also used for geopositioning the data processing, when analyzing human movement and comparing the general laws of motion for the entire population and when comparing to the health risks [32].

The main objective of the three-dimensional digital medical systems is providing the works with parameterized, tissue and cellular levels of 3D-models of normal and pathological human organs and systems [33]. The process of biomechanical analysis of any biological system, based

on the developed physical and mathematical model, requires the inclusion of the data on the spatial geometry of an object. Advanced computer tomography is equipped with special software used in the field of radiotherapy and oncology to work with obtained images [34]. The development of three-dimensional modeling for biomedical engineering is actively supported, and there are prospects for the implementation of the technologies based on supercomputing platforms [35]. The problems of clinical decision support systems, biomedical imaging, high-performance computing and biomedical modeling, analysis of bioinformatics of big data are discussed in this article [36].

Currently, healthcare organizations generate and store vast amounts of data. The amount of information that requires analysis is steadily growing. Conversion of this large data into vital practical information is challenging enough. Processing of such unstructured information reveals unique knowledge. Society of Actuaries (SOA) conducted a study, in which the leaders of the US health facilities and health insurance companies were interviewed. This report states that 66% of executives believes in a big potential of Big Data technologies in health care [37-39].

The potential benefits of big data analytics in the field of medicine have resulted in the public policy initiatives in the United Kingdom. The Prime Minister David Cameron announced the opening of the medical data of each test the patient for research. He also emphasized that the patients have the right to privacy conscious. The Prime Minister added: “it does not threaten the privacy and neither does mean that anyone can see your medical records. This implies the use of anonymous data for new medical discoveries. It is very important for research and innovation, since the healthcare sector is an arena for innovative use of big data” [40].

Conclusion

Modernization of the healthcare as one of the key priorities of the government involves the consolidation of the efforts of information technology and medical institutions. Big Data technologies offer great opportunities to combine all biological centers worldwide. The use of Big Data may contribute to addressing the problems such as investigating the structure of proteins to identify new methods for diagnosing cancer, diagnosing congenital diseases, and predicting potential growth of the population’s fertility, delivering drugs directly to the certain cell and many other problems. The forecast accuracy increases with the dimension of the data available for analysis. Analysis of large data flows from self-reports and personal medical records, including medical tests, drug prescriptions, as well as scanners, and various sensors may lead to dramatic improvements in health care. The use of up-to-date Big Data technical solutions allows us to solve complex research problems in the field of genomic, cellular and proteomic technologies. Big data solutions also turn received biomedical products and technologies into practical use as quickly as possible. Big Data technologies provide new opportunities for advanced analysis of the causes, symptoms and possible manifestations of diseases if the patient does not change the way of life. Being aware of the disease risks, we are exposed to under the supervision of a doctor, it will be possible to adjust the patient’s condition. The disease will not simply occur if the treatment is regarded this way. Availability, continuity, usability, scalability, ability to manage at different levels, privacy and security of solutions, including data quality control are essential for effective data control.

The use of big data technologies is in high demand, since they can solve urgent problems in medical and a number of related fields. The exponential growth in data volumes and new tools for their advanced analysis may transform the health care system. The use of modern computer methods for solving these problems will significantly increase the longevity of people in the future. Consequently, big data has a big future in the field of medicine.

References

1. Jee K., Kim G. Potentiality of Big Data in the Medical Sector: Focus on How to Reshape the Healthcare System / Healthcare Informatics Research, Korea, 2013, vol.19, no.2, pp.79–85.

2. Semashko A.V., newspaper “News of Medicine and Pharmacy” Gastroenterology (429) 2013.02.06.
3. Chernyak L. Seriously on technologies for Big Data // Open systems, 2014, No1, pp.26–28.
4. Koptelov A. What big data can give a business? // CNews Analytics, 2014. www.cnews.ru/reviews/bi_bigdata_2014/articles/chto_bolshie_dannye_mogut_dat_biznesu/
5. Carrera P., Dalton A. Do-it-yourself Healthcare: The current landscape, prospects and consequences//Maturitas, 2014, vol.77, no.1, pp.37–40.
6. Meyer-Shenberger V., Kukier K., “Big Data”, Moscow, 2014, pp.9–109.
7. “Big Data”, big work, big problem / ITU News <https://itunews.itu.int/Ru/Note.aspx?Note=4879>
8. Raghupathi W., Raghupathi V. Big Data analytics in healthcare: promise and potential // Health Information Science and Systems, 2014, vol.2, no.3, p.10.
9. Herland M., Khoshgoftaar T. M., Wald R. A review of data mining using big data in health informatics // Journal of Big Data, 2014, vol. 1, no.2, pp.4–7.
10. Big Data in medicine and healthcare www.osp.ru/medit/2014/04/13040834.html
11. Krechetov N.E., Kondratenkov V.A., Zelenskaya Y.V. International experience of using intersystems technologies in healthcare // Doctor and information technology, 2006, No4, pp.125–131.
12. Konstantinov A. The brain of the machine // Expert, 2013, No11, pp.8–11.
13. Genome as a collider <http://chrdrk.ru/tech/2015/7/1/emc/>
14. <https://ru.wikipedia.org>
15. Tripp S., Grueber M. Economic Impacts of the Human Genome Project / Battelle Memorial Institute, 2011, 6 p.
16. http://en.wikipedia.org/wiki/Personal_Genome_Project
17. Vasilkov A. Medicine becomes personalized affected by “big data” www.computerra.ru/88609/big-data-improve-personalized-medicine/
18. How GenePool Works. www.stationxinc.com/how-genepool-works
19. Scientific-popular portal about genetics <http://mygenome.su/articles/112/>
20. Lambert J., Ibrahim-Verbaas C, Harold D. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease // Nature genetics, 2013, no.45, pp.1452–1458.
21. Swarup V., Geschwind D.H. Alzheimer's disease: From Big Data to mechanism // Nature, 2013, vol. 500, pp.34–35.
22. Bettens K., Sleegers K., Broeckhoven C., Genetic insights in Alzheimer's disease/ The lancet neurology, 2013, vol.12, no.1, pp.92–104.
23. Muldashev E.R., Nigmatullin R.T. and et.al. Prospects of p4 concept in craniofacial surgery, News of FSBU 2012, No12 (148), pp.136–138.
24. Bradley W., Golding S., Herold C. Globalization of P4 medicine: predictive, personalized, preemptive and participatory-summary of the proceedings of the Eighth/ International Symposium of the International Society for Strategic Studies in Radiology, 2011, no.258(2), pp.571–582.
25. Shaikh A., et al. Collaborative Biomedicine in the Age of Big Data: The Case of Cancer // Journal Med Internet Res, 2014, vol.14, no.4, pp.101–109.
26. Coakley M., Crocetti G., Dressner P., Kellum W. Transforming Telemedicine Through Big Data Analytics, Cornell University Library, 2015, 10 p.
27. Kuo M., Sahama T., Andre W. Kushniruk A., Health big data analytics: current perspectives, challenges and potential solutions// International Journal of Big Data Intelligence, 2014, vol.1, no.1/2, pp.114–126.
28. Suchkov S., Rizing K. Rose N. Translational Medicine - Myth or Reality? // “Remedium” 2013, No.5, pp.10–17.

29. Health and Biomedical Big Data for Translational Research
<http://med.unsw.edu.au/event/health-and-biomedical-big-data-translational-research>
30. Shalkovsky A.G., Kuptsov S.M., Berseneva E.A. Topical issues of creating automated system for remote monitoring of human health // Doctor and information technology, 2016, No1, pp.67–79.
31. IBM Watson Health. www.interface.ru/home.asp?artId=37954
32. Trifonova O.P., Ilyin V.A. “Big data” in biology and medicine, Acta Naturae, Medicine and Health, GRNTI: Research Library, 2013, No3, pp.14–17.
33. Turlapov V., Gavrilov N., Saprykin N. Platform for digital medicine // Open systems, 2014, No 05, pp.64–66.
34. Skripcak T., Belka C., Bosch W. Creating a data exchange strategy for radiotherapy research: Towards federated databases and anonymised public datasets // Radiotherapy and Oncology, 2014, vol.113, no.3, pp.303–309.
35. Giglaviy A.V., Sokolov A.V., Abdrakhmanova G.I., Chulok A.A., Burov V.V. Long-term trends in the development of information and communication technology sector // Forsite, 2013, No3, pp.6–24.
36. Advanced Computing Solutions for Health Care and Medicine // Journal of Computational Science, 2012, vol.3, no.5, pp.247–438.
37. www.cisco.com/web/RU/news/releases/txt/2012/040612d.html
38. Alyass A., Turcotte M., Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities // BMC Medical Genomics 2015, vol.8. www.biomedcentral.com/1755-8794/8/33
39. Shinal J. If clean, big data can improve U.S. health care, Special / USA today, 2014.
40. www.usatoday.com/story/tech/columnist/shinal/2014/05/14/medical-privacy-health/9087873/
41. Tene O., Striks H. Big Data for All: Privacy and User Control in the Age of Analytic Northwestern // Northwestern Journal of Technology and Intellectual Property, 2013, vol.11, no.5, pp.239–272.