Opportunities for Big Data Analytics in Healthcare Information Systems Development for Decision Support

Blagoj Ristevski, Snezana Savoska and Natasha Blazheska-Tabakovska

Faculty of Information and Communication Technologies, University St. Kliment Ohridski, ul. Partizanska bb, Bitola, Republic of Macedonia

{blagoj.ristevski, snezana.savoska, natasa.tabakovska}@uklo.edu.mk

Abstract. Nowadays, an enormous volume of heterogeneous healthcare and medical data are generated routinely. These heterogeneous data have to be integrated and stored in a standard manner and format to perform appropriate big data analysis and visualization and improve decision-making. These data are generated from different sources such as mobile devices, sensors, national public health institutions, laboratory tests, clinical notes, social media, and various omics data that can be structured, semi-structured or unstructured. These data structure varieties necessitate these big data to be stored not only in the relational databases but also in NoSQL databases. To provide effective data analysis, besides the application of appropriate data mining techniques, excellent design and implementation of healthcare information systems are needed. These software solutions have to solve patient data security and privacy issues by employing proper big data governance policies. The design and implementation of healthcare information knowledge-based systems should provide to the patients more well-organized and economical healthcare services, and on the other hand, a boosted knowledge-based basis for decision-making to the managers in healthcare institutions and insurance companies and benefits for the all involved stakeholders. In this paper, we overview and suggest suitable development framework that will cover patient-, clinical- and population oriented approaches to decision-making and to reveal valuable knowledge and insights from these healthcare and medical big data. Moreover, on specific occasions, this knowledge should enable a rapid and reliable response to the healthcare hazards and help to decision-makers worldwide as well on the national level.

Keywords: big data, healthcare information systems, knowledge-based systems, decision-making framework, electronic health records.

1 Introduction

To obtain the optimal facilities and care for the patients, healthcare institutions in many countries have suggested numerous models of healthcare information systems. These models for personalized, predictive, preventive, patient-centric and evidence-based medicine are based on using massive amounts of complex biological, medical and healthcare data as well as electronic health records (EHRs) [15]. As the final use, data are used for decision-making in healthcare and medicine. Knowledge discovering from these medical and healthcare big data allows identifying the best practices in the hospitals, discovering the association rules and correlations in these data and unfolding the disease monitoring of particular disease and patient- and population-centric health trends.

Nowadays, smartphones are excellent platforms to deliver personal messages to patients to improve their welfare and health conditions and hence they are crucial devices for the telemedicine, which is a very important branch of medicine especially when special restrictive measures are established such as quarantines, lockdowns and curfews.

To analyze and process health and medical big datasets that come from different data sources, various techniques from many disciplines such as machine learning, pattern recognition, expert systems, statistics, applied mathematics, artificial intelligence are used. Many obstacles should be taken into account. These big data are complex, often stored in distributed databases with medical and healthcare records with a lack of integration capabilities and interoperability.

The development of new data mining techniques makes commonly used machine learning algorithms easy to be adopted by bioinformaticians and to become essential tools for the analysis of medical and healthcare big data. The integration of these data provides a clearer picture of cell functions and alterations. It will be more popular in the clinical health and disease examinations [16].

The new knowledge discovered by big data analytics techniques and developed knowledge-based healthcare information systems should afford wide-ranging and adequate advantages to the patients, clinicians, national public health organizations and institutions, healthcare policymakers, as well as the World Health Organization.

In this paper, we recommend directions for the development of knowledge-based healthcare information systems considering several aspects and functionalities.

The remainder of the paper is organized as follows. The concepts of big data and big data analytics toward a framework for decision-making are described in the second section. Section 3 describes the characteristics of healthcare information systems. The next section explains the principles of knowledge-based healthcare information systems in decision-making. The last section concludes this paper with discussion and directions for further works.

2 Big data and big data analytics

Contemporarily high throughput bioinformatics technologies generate large amounts of raw medical, biochemical and biomedical data, which are heterogeneous like EHRs data, and stored in different data formats. Health and medical big data refer to these numerous massive and complex data, which are hard to analyze and manage with traditional software and hardware resources. These big data can be categorized as structured, semi-structured or unstructured; discrete or continuous data.

Big data analytics in healthcare and medicine covers merging of heterogeneous data, control of data quality, analysis, modelling, visualization and validation. Application of big data analytics provides thorough knowledge discovering from the existing accessible large amounts of data [15]. Big data analytics in medicine and healthcare has to enable analysis of large datasets from patients, when public health of the whole population worldwide is important. Big data analytics identifies patterns and clusters from available datasets, examine existing of a data correlation, and develops predictive models using techniques from data science.

Another challenge when dealing with data mining techniques in big data is the classification of an imbalanced dataset, which appears when real-world applications produce classes with different distributions. One class is under-presented with an insignificant number of instances, while the second one has a plentiful number of instances. Identifying the minority classes is important in various fields such as medical diagnosis, drug discovery or bioinformatics [22].

The rapid development of the emerging information and communication technologies, experimental technologies and methods, cloud computing, the Internet of Things (IoT) and social networks provides the amounts of generated data that grow massively in medicine and healthcare as well as in other domains.

These high throughput data, that are so-called omics data, provide widespread insights towards different types of profiles, changes and interactions on a molecular and cellular level as well as knowledge associated to the genome, epigenome, transcriptome, proteome, metabolome, interactome, exposome, diseasome, etc. Besides these omics data, the EHRs data contain personal patients' data, clinical notes, diagnoses, administrative data, prescriptions [18], as well as charts, reports, laboratory tests, medical images, magnetic resonance imaging, ultrasound, tomography, X-ray data. Some of these data are acquired from wearable sensors or capture from medical monitoring devices, with different collection frequency that makes these data to have complex features and high dimensionality [15].

These growing amounts of various omics and EHRs data need to be collected, cleaned, stored, transformed, transferred and visualized in an appropriate fashion to be represented to the clinicians and healthcare organizations and institutions.

Thus, the development of appropriate healthcare information systems that are based on knowledge and big data is essential.

The term big data is usually described by the following features: volume, value, velocity, variety, veracity and variability, which are denoted as 6 “V’s” big data characteristics [15]. The volume of healthcare and medical data, usually measured in terabytes, petabytes and yottabytes, refers to the quantity of data, while value ascribes to the comprehensible and valuable data analysis. Velocity is associated with the shifting data in motion as well as and to the speed and frequency of their creation, processing and analysis. Complexity and heterogeneity of multiple datasets refer to the data variety. Veracity referrers to the quality, relevance, uncertainty, reliability and predictive value of the data, whereas variability is linked to the data consistency for a while.

Applications of big data analytics can improve the patient-based services, to detect spreading diseases earlier, generate new insights into disease mechanisms, monitor the quality of the medical and healthcare institutions as well as provide better treatment methods, especially for novel infectious diseases, when treatment techniques update often. If a disease occurs and cures in any part of the world, then prediction and modelling for that disease can be done competently when using big data analytics.

Big data analytics has the potential to transform the manner healthcare providers use advanced technologies to obtain knowledge from clinical and other data repositories and hence to make better decisions [12].

3 Healthcare information system

The biggest challenge today related to the analysis of big data in medicine and healthcare is the big data integration from many data sources that generate large amounts of data. Unfortunately, there is no pre-defined strategy to the healthcare data integration. Although there are many ontologies connected with healthcare and medicine as well as the factors that affect human health that attempt to integrate data on the principle of data “born interoperable”, as shown in Fig. 1. Some health data are well structured according to well-known medical and health coding systems but many of them do not have a predicted structure that could allow the ontological big data interoperability. Many commonly used ontologies, such as Gene Expression Ontology, Gene Ontology, microRNA Ontology, Protein Ontology, MONDO disease Ontology, Disease Ontology etc., as shown in Fig. 1, consider the standards established in healthcare and medicine and introduce well-chosen indicators that should help to reveal the hidden links between the data. Most medical and health data use already known coding systems in medicine, which is one-step towards improving the big data analysis. A well-known and good instance of this is the ontologically enabled big data integration into toxicology [4] that uses existing efforts and ontologies to link data using the knowledge graph to gain new knowledge. A knowledge graph is a tool used also in many domains. As an instance, Google's search algorithms since 2010 and the improved version of integration with the using of ontology since 2018 enable much more pervasive searching algorithms that allow finding hidden connections between data [5].

Another good example of data integration is the connections of biomedical data throughout most known biomedical ontologies that integrate data obtained experimentally in wet labs and reuse them via established ontological systems with predefined metadata that are more widely available. This integration of healthcare, omics, exposure and medical data is only part of the mosaic of total data sets used in medicine and healthcare, shown in Fig. 1.

The data, collected in healthcare institutions, laboratories, national and other health-related systems, should be added. Considering a huge amount of data collected from many sensors, widely used by many patients, such as Holter and other measuring sensors that show vital life signs, according to the principle of smart living with IoT, the data science techniques need to be employed to deal with these big data. In addition, the new trend of using environmental data that affect human health provides a wide range of applications in healthcare and medicine. These data are related to systems from various measuring stations for scientific research in various domains, such as measuring the number of nanoparticles, electromagnetic radiation, the concentration of pollutants in the stratosphere, chemicals that affect health and many others sensors that measure environmental pollution, radiation, various types of soil, water but also human nutrition data.

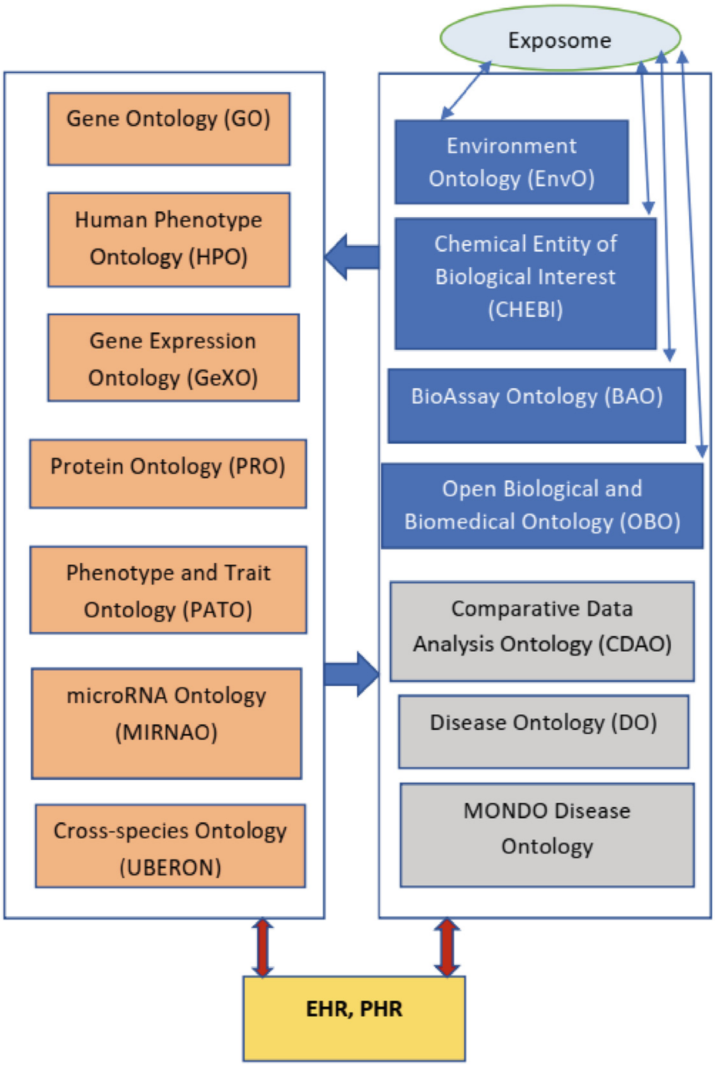

Fig. 1. A part of the mosaic of ontologies of whole data sets used in ecosystems in medicine and healthcare.

Today, the largest aggregators of healthcare data are smartphones with many applications used by users. They also do not have pre-defined structures and measurement methods but follow well-known and accepted healthcare standards. Linking of all these data with the need to risk assessment connected with genotype, phenotype and disease is another branch that has been intensively exploited since 2006 as exposome data. They are related to the impact of external factors on the genotype and phenotype and intend to provide the risk calculation for each individual based on data from his/her EHR/PHR and outer influences of the existing big data repositories for various factors. Also using artificial intelligence methods to quantify location-based risk and finding an influence of all previously mentioned data is required [10] [19].

Another important challenge includes links between the patient-based database (EHR), e-health data and all data that allow evidence-based medicine, pervasive healthcare and telemedicine [7]. The continuous migration and movement of people require from patient to bring their healthcare data with themselves when travel. Unfortunately, these data are not usually accessible for patients because data are owned by institutions that provide healthcare. So, the patients when are abroad and physicians who have to make decisions do not have enough data to practice evidence-based medicine. It suggests that medical facts are needed at every level of decision-making, especially for the treatment and care of patients for evidence-based medicine [1]. Although these data are stored in EHR, they are not available outside the institutions and the country where the patient is a citizen. Therefore, many efforts are aimed toward supporting the creation of PHR for patients and allow the patients to have access to all data related to their health statuses such as EHR from medical practitioners, institutions visits, medical research [10], laboratory and biometric results, prescriptions and referrals and other data related to the screening of patient's health. Availability and accessibility of such data are activities towards improving evidence-based medicine and making good decisions by physicians to whom the patient addresses. If the patients have their health data, the concept is PHR allowing evidence-based decision-making and medical treatment when the patients are abroad.

This concept requires using web-enabled technology with e-health and evidence-based PHR, which has many advantages. For physicians, adopting PHR is important because it allows improving the decision-making process from everywhere, based on evidence and for the patient to receive better evidence-based diagnosis and treatment from medical practitioners. IoT based health-monitoring sensors included as wearable measurement sensors also improve evidence-based decision-making. These devices connected with Bluetooth can capture and store health-related data. PHR can also contain data obtained from the equipment such as accelerometers, gyroscopes, wristband and smartwatches usually connected to smartphones and stored in PHR via Bluetooth connection [14].

Nowadays, when we are facing epidemics of various diseases associated with many autoimmune diseases that require genetic analysis but also pandemics that requires analyzing of genetic data of many microorganisms, bacteria and viruses, we are aware that data integration is an extremely important task. Patient-centric decision-making should give way to different analysis of patient groups and high dimensional data related to diseases, behaviors, treatments and many omics-data. This requires the application of global data interoperability standards such as those for PHR, proposed by ISO in 2012, introduced as EHR-ISO/TR 14292 and applied as HL7 standard [13]. Because these data have very variable formats and often they protected by private data protection standards and laws, they cannot always be used for analysis. Data decoding and aggregation are sometimes required to enable data analysis such as population-centric, epidemic-centric, clinic-disease-centric, hospital-centric, region-centric [25], country-centric, to support decision-making by healthcare stakeholders at different layers, as shown in Fig. 2.

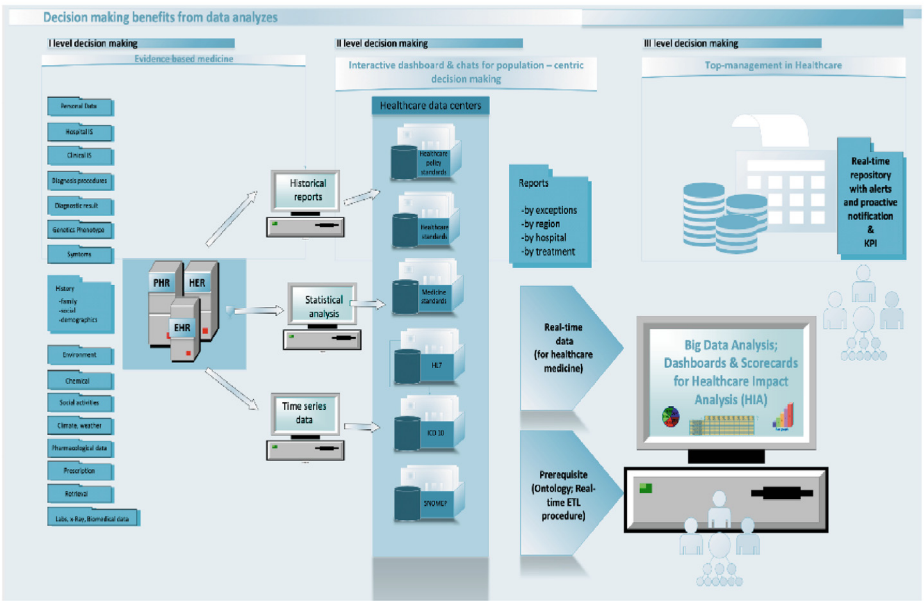

Fig. 2. Benefits from big data analysis for decision-making in three layers.

When benefits from big data analysis in healthcare and medicine are taken into account, a couple of aspects have to be considered. The first aspect, clinical conclusions as historical reports, statistical analyses as well as time-series analyses and comprehensive reports have to provide evidence-based decision making in medicine and healthcare, especially for diagnostics and healthcare treatments [24]. The second aspect is the information visualization that enables interpolation of critical big data for analysis using interactive dashboards or charts that support daily operations of physicians and nurses and helps them to make more efficient and faster decisions [17]. The third aspect is real-time reporting as warnings, alerts and proactive notifications, real-time navigation and the application of operational key performance indicators that are usually placed on dashboards in real-time.

According to the framework suggested by Shang et al. [21], there are five benefit dimensions, which include IT infrastructure benefits, operational benefits, organizational benefits, managerial benefits, and strategic benefits. This framework is suitable for a more general system model for categorizing the benefits of big data analytics. They consider the content analysis that takes place in a three-phase process: preparation, organizing and reporting. It provides a better comprehension of big data analytics capabilities and healthcare benefits. They have identified 5 big data opportunities: analytics capabilities for big data analytics, insights capability, predictive analytics capability, interoperability capability and traceability.

4 Using knowledge-based healthcare information systems in decision-making

The success of an organization depends on the knowledge quality. To support knowledge management in healthcare organizations, health information systems must provide information and guidance to the medical personnel and patients and help them in decision-making. Health information systems (IS) are complex. They cover a wide and diverse range of applications, including hospital IS, nursing IS, laboratory IS, radiological IS, pharmaceutical IS, EHR systems, patient monitoring systems, clinical decision support systems, medical education system etc.

Additionally, healthcare decision-making has to be a knowledge-driven process, so knowledge management tools in the healthcare sector are very important. Providing the right knowledge at the right time at the point of decision-making by implementing knowledge management in healthcare is crucial.

The typical architecture of a knowledge-based healthcare information system includes a knowledge base and an inference engine. Health information system backed by a rich and effective knowledge base, which contains a collection of information of the field of medical diagnosis, ensures efficiency in identification, analysis, and selection of optimal action for the patient care. Inter-organizational knowledge sharing is one of the fundamental steps in knowledge management processes and can serve as a strategic system for knowledge-intensive sectors such as healthcare [20]. The inference engine deduces insights from the information stored in the knowledge base. This improves access to the patient data and facilitates the decision-making process in the shortest possible time. The healthcare person through the interface interacts with the system during the decision-making process.

The knowledge-based healthcare information systems use appropriate tools for knowledge management and user-friendly interactions because they can significantly improve the quality and safety of care provided for patients at the hospital and home surroundings [20]. They assist in the data collection, analysis, management and sharing of knowledge between business processes for healthcare. The vital aspects of knowledge-based healthcare information systems are utilization, transfer and translation of knowledge. Knowledge utilization is the process of converting knowledge, such as evidence-based guidelines to practices. Knowledge translation moves scientific knowledge from basic discovery to testing for technical efficiency and then to acceptability for adoption in practices. The third aspect, knowledge transfer, is the diffusion of knowledge that is directed and managed by using various strategies [20].

The healthcare providers play a significant role in patient care such as enhancing the quality of care, ensuring individual based on the most up-to-date evidence, ensuring physicians to maximize the likelihood of positive outcomes as well as minimizing the existing gap between research and practice [2]. The Boateng [2] and Lapaige [8] defined four steps in decision-making and actions taken in the healthcare industry: 1) formulation of a clear clinical question related to the patient problem; 2) a search in literature for relevant clinical practices; 3) the evaluation of the available evidence for its usability and 4) implementation of the evidence in clinical practices.

There are many currently existing knowledge management tools to implement in healthcare organizations. These tools are interactive, most initiated and allow effective communication between healthcare professionals, managing their knowledge, generating discussion about new concepts or ideas, finding answers to particular problems.

Other opportunities for healthcare professionals include clinical decision support systems, EHR system, the community of practices and advanced care management. To obtain a full-fledged implementation of knowledge in healthcare, all stakeholders such as policymakers, researchers, health professionals and healthcare providers need to come together and play their part to seize the opportunities and improve healthcare quality [20].

Furthermore, public health practitioners need to be fully informed when deciding on the design of an information system, integration and using of data [3]. Accordingly, the knowledge-based information system is expected to play more roles in healthcare in the future. The design of knowledge-based information system in healthcare organizations are gaining more attention. The most important and challenging task in designing a knowledge-based information system is to organize and maintain patient information repositories securely, accurately and in a speedy manner [18].

As knowledge-based information systems' applications increase further in the healthcare sector, some privacy and security issues will arise, as well as issues related to the data transparency, drug prescription ad supply chain errors, integrity, accessibility, resource and patient management and knowledge interpretation [9]. Over time, all of the health information will be available electronically, to the patient, to the doctor and other healthcare providers. Because many organizations and people may have access to health information, there will be concerns about the privacy and security of health information [11].

Because big data flood in medicine, healthcare, and the nature of medical and healthcare data, new database management systems such as Cassandra, MongoDB, MarkLogic and Apache HBase and NoSQL database systems should be employed for the development of knowledge-based healthcare information systems [23]. The framework of developed IS should enable data analysis on a different tier, such as patient-centric, population-centric, epidemic-centric, clinical-centric, country-centric, to support decision making by the heads of the healthcare and medical institutions and organizations. Furthermore, data privacy and security issues of the patient sensitive data have to be solved in every proposed software solution by using appropriate anonymization and cryptographic protocols.

5 Conclusion and further work

The healthcare industry generates many Exabytes healthcare data, manly in the form of EHR. However, most of the achievable values of data usage do not obtain full potential. This process is still in its infancy because predictive modelling and simulation techniques for analyzing healthcare data as a whole have not yet been adequately developed [23]. Big data analytics in medicine and healthcare offers very promising possibilities for the development of decision support systems and knowledge-based healthcare information systems that will integrate explore and analyze large amounts of data. These integrated information systems should make a symbiosis among patient-, clinical- and population-centric decision-making systems [6]. As further work, the big data characteristics provide a very appropriate foundation to use promising software platforms for the development of applications that can deal with healthcare and medical big data. Moreover, the development of healthcare information systems has to consider solving the security and privacy issues of all involved parties, especially patients' sensitive data, and these software solutions have to enable patient-centric, population-centric, epidemic-centric, clinical-centric, country-centric data analysis to make decision making more effective.

References

1. Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ. Big data for health. IEEE journal of biomedical and health informatics. 2015 Jul 10;19(4):1193-208.
2. Boateng W. Knowledge management in evidence-based medical practice: does the patient matter?. Electronic Journal of Knowledge Management. 2010 Nov 1;8(3).
3. Biancone P, Secinaro S, Brescia V. A review of big data quality and an assessment method and features of data quality for public health information systems, International Journal of Management Sciences and Business Research, Jan-2018 ISSN (2226-8235) Vol-7, Issue 1
4. Boyles RR, Thessen AE, Waldrop A, Haendel MA. Ontology-based data integration for advancing toxicological knowledge. Current Opinion in Toxicology. 2019 Aug 1;16:67-74.
5. Chah N. OK Google, What Is Your Ontology? Or: Exploring Freebase Classification to Understand Google's Knowledge Graph. arXiv preprint arXiv:1805.03885. 2018 May 10.
6. Eapen BR, Archer N, Sartipi K. Public Health Information Systems: From Data to Knowledge, 2020.
7. Househ and all, Big Data, Big Challenges: A Healthcare Perspective Background, Issues, Solutions and Research Directions, Fernando Martin-Sanchez, Big Data Challenges from an Integrative Exposome/Expotype Perspective, Springer 2019, ISSN 2195-271X, ISSN 2195-2728.
8. Lapaige V. Evidence-based decision-making within the context of globalization: A "Why–What–How" for leaders and managers of health care organizations. Risk management and healthcare policy. 2009;2:35.
9. Litchfield AT, Khan A. A Review of Issues in Healthcare Information Management Systems and Blockchain Solutions. CONF-IRM. 2019.
10. López-Campos G, Merolli M, Martín-Sánchez F. Biomedical Informatics and the Digital Component of the Exposome. InMedInfo 2017 (pp. 496-500).
11. Mahmood N, Burney A, Abbas Z, Rizwan K. Data and knowledge management in designing healthcare information systems. International Journal of Computer Applications. 2012 Jan 1;50(2):34-9.
12. Milenkovic MJ, Vukmirovic A, Milenkovic D. Big data analytics in the health sector: challenges and potentials. Management: Journal of Sustainable Business and Management Solutions in Emerging Economies. 2019 Mar 19;24(1):23-33.
13. Morgenthaler J. Moving Toward an Open Standard, Universal Health Record. Smart-publications. 2007.
14. Rappaport SM, Smith MT. Environment and disease risks. science. 2010 Oct 22;330(6003):460-1.
15. Ristevski B, Chen M. Big data analytics in medicine and healthcare. Journal of integrative bioinformatics. 2018 May 10;15(3).
16. Ristevski B, Savoska S, Mitrevski P. Complex Network Analysis and Big Omics Data, ICT Innovations 2019, RN of Macedonia, 2019.
17. Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. Health affairs. 2014 Jul 1;33(7):1115-22.
18. Sarkar BK. Big data for secure healthcare system: a conceptual design. Complex & Intelligent Systems. 2017 Jun 1;3(2):133-51.
19. Savoska S, Ristevski B, Blazheska-Tabakovska N, Jolevski I. Towards Integration Exposome Data and Personal Health Records in the Age of IoT, ICT Innovations 2019, RN of Macedonia, 2019.

20. Shahmoradi L, Safadari R, Jimma W. Knowledge management implementation and the tools utilized in healthcare for evidence-based decision making: a systematic review. Ethiopian journal of health sciences. 2017;27(5):541-58.
21. Shang S, Seddon PB. Assessing and managing the benefits of enterprise systems: the business manager's perspective. Information systems journal. 2002 Oct;12(4):271-99.
22. Oussous A, Benjelloun FZ, Lahcen AA, Belfkih S. Big Data technologies: A survey. Journal of King Saud University-Computer and Information Sciences. 2018 Oct 1;30(4):431-48.
23. Wang Y, Kung L, Byrd TA. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change. 2018 Jan 1;126:3-13.
24. Wang Y, Hajli N. Exploring the path to big data analytics success in healthcare. Journal of Business Research. 2017 Jan 1;70:287-99.
25. https://ec.europa.eu/health/indicators_data/data_en