Social Science & Medicine 241 (2019) 112533

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Review article

Values, challenges and future directions of big data analytics in healthcare: A systematic review

P. Galetsi[a], K. Katsaliaki[a], S. Kumar[b,*]

[a] School of Economics, Business Administration & Legal Studies, International Hellenic University, 14th km Thessaloniki-N.Moudania, Thessaloniki, 57001, Greece
[b] Opus College of Business, University of St. Thomas Minneapolis Campus, 1000 LaSalle Avenue, Schulze Hall 435, Minneapolis, MN 55403, USA

ARTICLE INFO

Keywords:
Systematic review
Big data analytics
Health-medicine
Decision-making
Organizational and societal values
Preferred reporting items for systematic reviews and meta-analyses

ABSTRACT

The emergence of powerful software has created conditions and approaches for large datasets to be collected and analyzed which has led to informed decision-making towards tackling health issues. The objective of this study is to systematically review 804 scholarly publications related to big data analytics in health in order to identify the organizational and social values along with associated challenges. Key principles of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology were followed for conducting systematic reviews. Following a research path, we present the values, challenges and future directions of the scientific area using indicative examples from relevant published articles. The study reveals that one of the main values created is the development of analytical techniques which provides personalized health services to users and supports human decision-making using automated algorithms, challenging the power issues in the doctor-patient relationship and creating new working conditions. A main challenge to data analytics is data management and security when processing large volumes of sensitive, personal health data. Future research is directed towards the development of systems that will standardize and secure the process of extracting private healthcare datasets from relevant organizations. Our systematic literature review aims to provide to governments and health policy-makers a better understanding of how the development of a data driven strategy can improve public health and the functioning of healthcare organizations but also how can create challenges that need to be addressed in the near future to avoid societal malfunctions.

1. Introduction

Big Data Analytics (BDA) are the techniques, technologies, systems, practices, methodologies, and applications that analyse a vast amount of data to help an organization better understand its business, market, and make timely decisions (Chen et al., 2012). Particularly, healthcare industry is extremely data intensive and requires interactive, dynamic big data platforms with innovative technologies and tools to advance patient care and services. The use of BDA methods and techniques, such as optimization, forecasting, simulation, etc., is of paramount importance for the provision of meaningful recommendations and insights to managers and policy-makers when refining strategic decisions (Doumpos and Zopounidis, 2016). As a result, IT professionals constantly develop new applications with big data capabilities to help healthcare stakeholders increase value. One of the most used computing platforms for processing big data in general but in healthcare too, is Apache Hadoop (De Silva et al., 2015), a software framework enabling the distributed processing of large datasets across clusters of computers allowing storage, refinement and analysis of vast amount of data. To this extent, remarkable is also the use of the most recently developed analytics techniques of "machine learning", a data driven computational approach using algorithms capable of recognizing patterns and making predictions (Gruebner et al., 2017) and "visualization", which create tables, images, diagrams and other intuitive display methods to understand data (Chen and Zhang, 2014).

The enormous volume of gathered data could be beneficial to diverse stakeholders (patients, providers, researchers, pharmaceutical companies, medical device companies, payers, governments, and software companies) all of whom have different expectations from the evolution of healthcare data analytics (Feldman et al., 2016). For example, the management of healthcare data could be beneficial for organizations in areas such as, production of effective drugs and devices for patients' well-being, fraud detection in billing and speed of service (Raghupathi and Raghupathi, 2013) and for the society in tackling global health issues such as, disease prevention, public health surveillance, timely provision of essential medical services in emergencies.

* Corresponding author.
E-mail addresses: p.galetsi@ihu.edu.gr (P. Galetsi), k.katsaliaki@ihu.edu.gr (K. Katsaliaki), skumar@stthomas.edu (S. Kumar).

https://doi.org/10.1016/j.socscimed.2019.112533
Received 7 February 2019; Received in revised form 6 July 2019; Accepted 30 August 2019
Available online 10 September 2019
0277-9536/

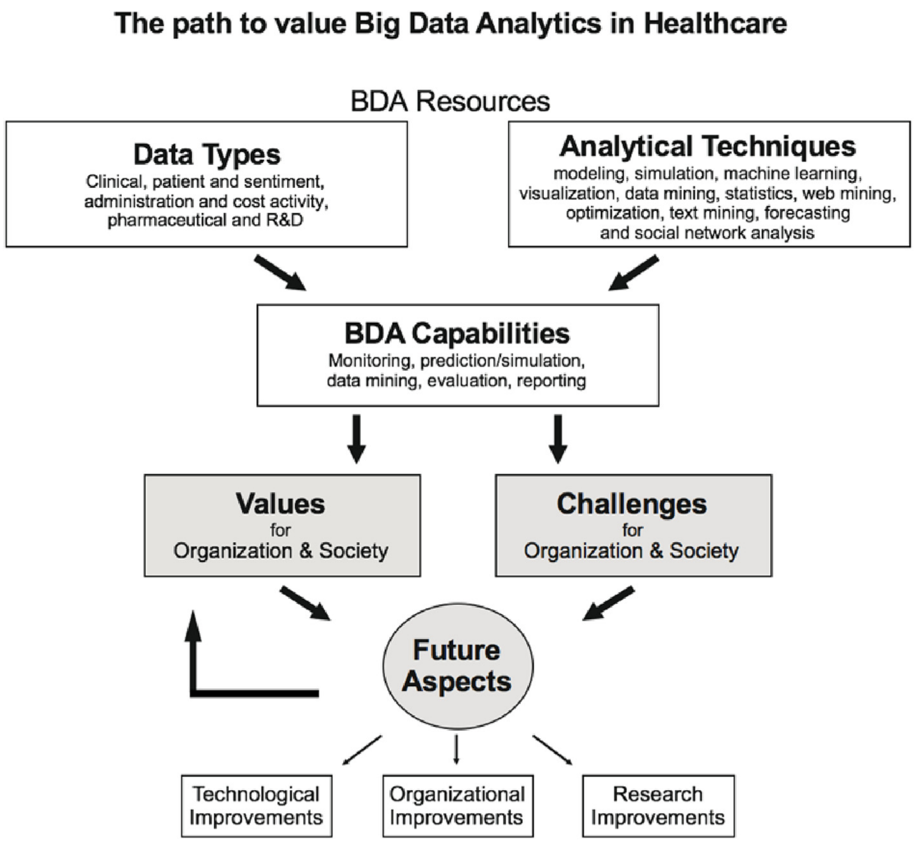

Fig. 1. Research framework.

Sociologists argue that data collected from groups of individuals using digital databases, e.g. social media groups discussing health issues, make visible aspects of individuals and groups that could not otherwise be perceptible. On the other hand, high public concern is raised too regarding big data ownership and health data privacy protection regulation (Mooney and Pejaver, 2018).

Despite these important observations, scholars have made little progress in the formulation of an approach to broadly identify organizational and social values and the related challenges of big data analytics in health. Although the published material in the field of BDA in the healthcare industry has rapidly increased in recent years (Ivan and Velicanu, 2015), there is yet no study which shows an all-encompassing organizational and social impact, positive or negative, of BDA in healthcare. "Challenging the dominant techno-utopian approach evident in digital health discourse" has been characterized as critical for future research (Lupton, 2014).

This study seeks to fill this gap in the literature. We systematically reviewed and analyzed the content of scholar publication related to health BDA in order to identify from the literature the organizational and social values that can be achieved in healthcare by the use of BDA, as well as to explore the issues that arise from its use and the future needs and directions of BDA in health, following a conceptual framework which adopts the resource-based view theory. Findings from this work will provide governments, groups and health policy-makers with a better understanding of how the development of a data driven strategy can improve public health, reduce the incidence of disease and better inform the populations, but also how such a strategy creates challenges that need to be addressed in the near future to avoid societal malfunctions.

2. Theoretical framework

According to the resource-based view, firms gain a competitive advantage by bundling resources into capabilities to create value (Gunasekaran et al., 2017). IT infrastructure is a major business resource for gaining long-term competitive advantage (Bharadwaj, 2000) along with the data gathered from IT infrastructure. The healthcare industry is a good example of the application of the resource-based view, because scientists are relying more and more on automation and applications to manage the vast amount of every day data.

The data resources that the healthcare industry needs to appropriately handle in order to create big data capabilities are categorized according to Groves et al. (2013) in (a) clinical data, (b) patient and sentiment data, (c) administration and cost activity data, and (d) pharmaceutical and R&D data. Yet, for the transformation of data into capabilities, the process of data analysis is required in-between. Based on the literature (Waller and Fawcett, 2013; Chen and Zhang, 2014), the techniques for the analysis of healthcare data are: modeling, simulation, machine learning, visualization, data mining, statistics, web mining, optimization, text mining, forecasting, and social network analysis techniques. In healthcare organizations, the five most important BDA capabilities developed from data resources and appropriate analysis infrastructure are (Groves et al., 2013): (a)"monitoring," includes efficiencies (using analytical methods) describing "what is happening now; " (b)"prediction/simulation," provides information about future outcomes (what will happen); (c)"data mining," involves methods enabling extraction and categorization of knowledge (what happened); (d)"evaluation" that demonstrates methods for testing the performance of analytical techniques or explains the outcomes of the application of BDA (why did it happen?); and (e)"reporting," includes methods that shape collected knowledge and provide it in an informational form.

The resource-based view theory asserts that those capabilities create new organizational values, which maintain their competitive advantage. This study expands the resource-based view following a "path–to value" approach to pinpoint the positive societal impact but

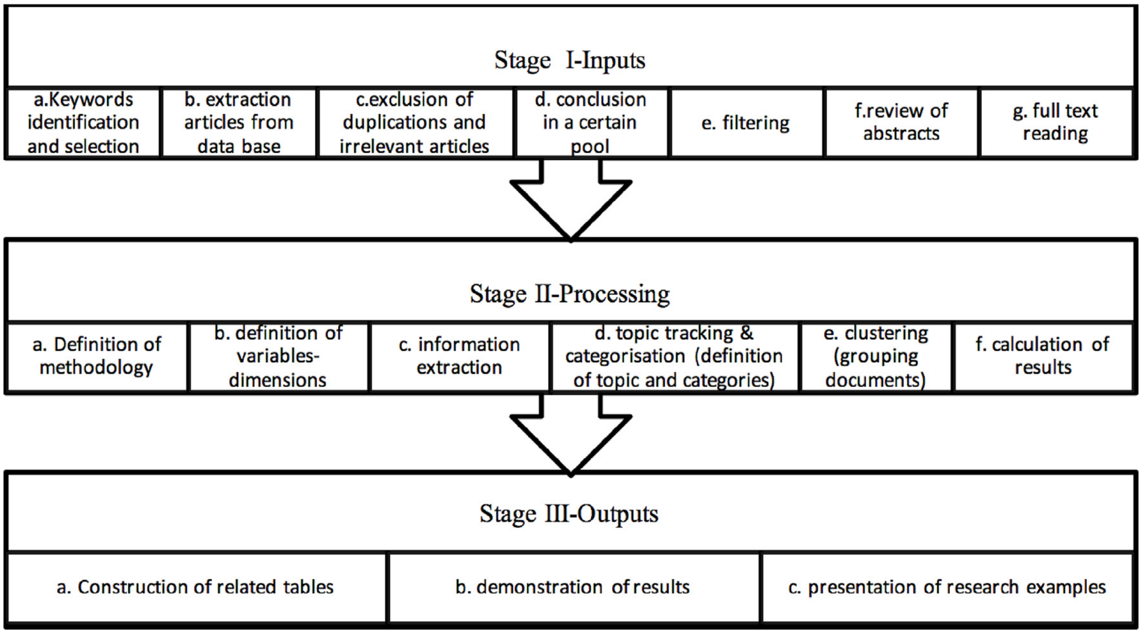

Fig. 2. Methodology process.

also to identify the challenges the acquired BDA capabilities pose both for the organizations and the society. Future research is targeted to further capitalize on these values, create new ones and overcome the challenges. Therefore, a path-to-value loop is created, and leads to changes as indicated by the literature. Following the above description, we summarized the resource-based view and our "path-to-value" concept in Fig. 1.

3. Methodology

The analysis of the elements of the research framework is based on the information gained from the synthesis of the existing literature. For conducting this systematic literature review we followed the key principles of systematic reviews (PRISMA). Our methodology is presented in three stages in Fig. 2: Stage I refers to the inputs of the review process which derived from searches in two well-known electronic databases: Web of Science® and Scopus. We reviewed "articles" and "review" papers written in English and published between 2000 and 2016 inclusive, as the terms "analytics", "business intelligence" and "big data" became popular around 2000s (Chen et al., 2012; Davenport, 2006; Mortenson et al., 2015). We consider analytics and business intelligence as similar terms since both investigate the capabilities of analytical tools in the business processes and are used in many reviews as unified terms (Chae and Olson, 2013; Chen et al., 2012). Hence, for our keywords search we used the combination of the terms a) "analytics", b) "business intelligence" and c) "big data.", in addition to the terms "health", "clinical" and "medical" with its derivatives. From our initial search, we ended up with a pool of 3241 articles excluding duplications. The remained references were screened as of their title and abstract excluding articles not analyzing/discussing big data or not relevant to the health industry. A full text screening followed for the remaining 1877 articles with the same criteria concluding to 804 relevant articles. Two reviewers independently screened the abstracts of the 804 articles for the final eligibility assessment to establish inter-reviewer reliability (O'Connor et al., 2018). The final dataset comprised of: papers testing the capabilities of a proposed new technique or technology for collecting, storing or harnessing potentially big healthcare data; papers that refer to biomarkers analysis; overviews and case study papers which are descriptive of the existence, benefits, and use of big data and its tools and technologies in the health domain.

Stage II incorporates information extraction, topic tracking and categorisation. In order to derive information from the unstructured text of the 804 articles, text classification was applied using the NVivo software. Following the research framework, we formulated variables/dimensions with subcategories taken from a specific literature source (Wamba et al., 2015), and then formulated additional ones based on the knowledge generated from reading many relevant papers. During the allocation of the 804 articles to dimensions we noted that in several occasions, a paper had to be assigned in more than one subcategory of a certain dimension.

Finally, Stage III of the research process presents the outputs of the classification process in the form of tables with article frequencies per dimension and indicative research examples per dimension. As the collected sample of the published work is large, we can assume it to be quite representative of the health analytics field. Therefore, a presentation of some proportional results of this dataset could shed some light on the research that has been conducted thus far in this area.

4. Research results

4.1. Types of value creation in healthcare

Aiming to identify the types of value creation in the healthcare industry from the analysis of big data, we formed 10 categories. The order of their presentation in Table 1 is based on their degree of popularity among the 804 articles. Five values (V2, V3, V4, V6 and V8) were taken from the literature (Wamba et al., 2015) and the remaining five (V1, V5, V7, V9, V10) were created after reading the content of all articles. Although some values may overlap with each other or be a sub-category of another, they are as inclusive as possible. For each of the 10 values we attempt to describe the gain both for the health-related organizations and for the society by also providing examples from our literature pool.

From this analysis, we conclude that the majority of researchers expect health organizations and patients to gain value from the "personalized innovative medical approaches (35.6%) (V1). For example, the analysis of specific patient's biomarkers guide disease therapy to precision medicine, which brings better diagnosis results tailored to individual perspectives (You et al., 2008). The capability of BDA to quickly monitor and analyze the health records of large patient cohorts

Table 1
Types of value creation.

	Values	N	%	Organizational impact	Social Impact
V1	New approaches to diagnosis for personalized healthcare	286	35.6	Analytic approaches that provide a personalized service to users	Enabling the collection of behavioral and social personalized data and, on an aggregate level, detecting factors with a social impact
V2	Replacing/supporting human decision-making with automated algorithms	206	25.6	Improve decision-making and reveal knowledge faster which can be made available to all stakeholders	1. Technology replacing human labor creating new employment conditions 2. Challenges the power issues in the doctor-patient relationship
V3	Innovating new business models, products, and services	197	24.5	BDA enables companies to create new products and services, enhance existing ones, and invent new business models	Innovative information tools can be transformed to socially meaningful scientific knowledge conceptualized as a public good
V4	Enabling experimentation to discover needs, expose variability, and improve performance	144	17.9	Provide experimental applications so as organizations better manage performance	Create conditions for large datasets to be analyzed and utilized for avoiding people suffering from misallocation of resources
V5	Coordination of healthcare information	122	15.2	Sharing of information and data analysis among stakeholders to gain operational efficiency	Sharing of information across health services and countries to improve decision-making for global health issues
V6	Creating efficiency	115	14.3	Collect data in a standardized format for reducing processing time and cost and enhancing data quality	Learn, in a timely and less costly manner, about population metrics that were unthinkable in the recent past
V7	Identify patient care risk	79	9.8	Avoid patient care risk through applications development that provide clinical risk prediction	Prevent social and economic vulnerabilities that challenge the health, security and sustainable growth of our society
V8	Offering customized actions by segmenting populations	72	9	Create highly specific segmentations though the exploitation of BDA and tailor products and services to meet needs	Gather vital information on targeted sensitive populations (such as HIV patients) for directing services
V9	Achieving cost-effectiveness	72	9	Discover new cost-effective ways to intervene on the determinants of health and improving health while reducing expenditures	Policy-makers with evidence from decision-support tools can allocate money to heath interventions which can save more lives and help poor communities buffer the adverse health effects of poverty
V10	Protecting privacy	41	5.1	Ethical guidelines to ensure that BDA supports the principles of respect of persons and avoids illegal acts	Protect patient confidentiality to prevent unethical targeting of groups for race, ethnicity, or sociodemographics

to learn which individuals respond to certain types of drugs, etc., help in this direction. In sociology, "the concept of identity" is a highly debated subject and concerns the understanding of a person's character, situation and experiences. This can be reflected in the individual's health condition and treatment as health is highly connected with daily life and everyday habits (Lowton et al., 2017). Personal health-related data, which can easily be acquired from the use of devices (e.g. sensors embedded in smartwatches and smartphone log data), enables the continuous collection of social and behavioral data such as, communication intensity from calls, sms, twits, etc., which characterize social interaction, physical activity (steps count), sleep patterns and heart rate monitoring which affect mood and behavior (Banos et al., 2016), identification in real-time hand-to-mouth gestures that characterize smoking pattern, and many others. The associated social value of these new approaches to diagnosis is that utilizing such data on an aggregate level could allow the detection of factors such as, stress, smoking and drugs that have a negative social impact (Kumar et al., 2015).

The next popular value type (V2) is defined by the improvement of decision-making through mining all possible knowledge from vast amounts of collected data by supporting or even replacing human decision-making with algorithms (25.6%). Such capability enhances and accelerates health professionals' diagnosis of patients and possibly mitigates errors in proposed therapy, when algorithmic results are explored and coupled with appropriate medical education. For example, Benharref et al. (2014) developed a decision–making tool, the "Fuzzy Expert System" that relies on data collected from continuous monitoring (health metrics), to produce recommendations (related to food-intake, medications, and lifestyle) for both patients and physicians by mitigating the risks of chronic diseases. The social impact of this value is related to the replacement of human labor with technology and the development of new employment conditions (e.g. new job positions related to medical technology management, increased specialization of medical staff) with debating results for the society (Loebbecke and Picot, 2015). It might also have an effect on the "sociology of diagnosis" theory related to the authority of the medical expert over the patient, and the changing power in the physician-patient encounter, since the patient will be aware that diagnosis and treatment are mainly based on automated algorithms (Lupton and Jutel, 2015), to which the patient may have access too.

Third (V3), comes the value of new business models, products, and services, identified in 24.5% of the articles. This includes articles that develop valuable tools and new products/services which assist decision-making for care and therapy for bigger populations. For example, Angulo et al. (2016) developed a new visualization software, the "BRAVIZ", that provides real-time statistical analyses of brain images for better patient diagnosis. It also refers to the context of a "rapidly developing ecosystem of digital health technologies" including dimensions such as online forums and medical and health-related apps of self-diagnosis (Lupton and Jutel, 2015) that have the potential to be transformed to socially meaningful scientific knowledge conceptualized as a public good (Evangelatos et al., 2016). It is apparent from the research examples that the last two values are closely related. For example, the development of a software/application where the user can input symptoms and get a disease diagnosis and drug recommendation (Mohan et al., 2016) entails the support (or even replacement) of physician's diagnosis and the offering of a new service available to people for everyday use.

The next value type (V4) includes papers (17.9%) that describe researchers' efforts to enable experimentation and activities to discover needs, expose inter-experiment performance variability and improve the performance of new BDA models that are used in health information systems for decision-making. This will enable large datasets to be analyzed and hopefully utilized for the social good, avoiding people suffering from resources' misallocation (Amankwah-Amoah, 2016). For example, by quickly collecting and analyzing reported cases of new diseases due to the use of new software applications for retrieving and

Table 2
Challenges from the implementation of BDA in healthcare.

	Challenges	N	%	Organizational impact	Social Impact
Ch1	Data management, security and privacy issues	161	20	Issues with data integrity and privacy lead to poor data management	Privacy violation and discrimination
Ch2	Technological issues	82	10.2	Lack of required infrastructure cannot produce safe conclusions	Social inequality, as BDA is only available to a small elite of technical specialists who have access to IT and know how to use it and interpret results.
Ch3	Further evaluation issues	57	7.1	Sample bias, inaccurate data lead to incorrect decisions	False alarms from missing data and not inclusion of population categories' data in samples
Ch4	Organizational and financing issues	25	3.1	Lack of cost-benefit analysis frameworks and skills to support IT-enabled healthcare processes. Organizational complexity	The changing labor market driven by IT may raise issues at certain employee categories within organizations
Ch5	Regulatory issues	5	0.6	Lack of connected structures and of new ways to integrate working practices across hospitals and community services	Lack of regulation about who has control on data creates confusion for taking responsibility
Ch6	Limited awareness and support	4	0.5	Lack of funding and awareness of BDA capabilities	Dependency on private funding may lead to supporting chosen players for undertaking BDA projects and reduce equality
Ch7	Political issues	3	0.4	Political barriers to adoption of IT and conflict of interests	There are trust issues regarding the responsibilities of BDA decisions' prescription in healthcare

manipulating electronic medical records (EMR), health systems of many developing countries can acquire an early-warning system which will assist public policy officials in a timely and efficient allocation of resources and treatments and provide people with essential health services (Amankwah-Amoah, 2016).

V5 is about the coordination of healthcare information (15.2%) which leads to the simplification of the data service. For example, PROACT, an application for smartphones, receives from various sources data such as, cancer patients' clinical information and drug tolerability and monitors diet and exercise levels to suggest actions to patients that reduce cancer-related problems (Hughes et al., 2016). At the organizational level this can bring operational efficiencies (e.g. better resource utilization), and at the society level the effective information sharing across health and public services can improve the quality of decision towards global health issues (Dinov, 2016), such as the guidelines issued by the world health organization (WHO) based on health metrics gathered by national governments.

V6 "creating efficiency" (14.3%), is about the capability of BDA to collect data in a standardized format for reducing data identification time, analysis time, and the cost of search and processing while maintaining/improving data quality. For example, Boytcheva et al. (2015) achieved easier knowledge extraction by storing information from large amount of clinical narratives in a structured format. Approximately 100 million outpatient care notes were used to apply the method. From the social aspect, this will also provide the opportunity to learn, with less cost and in less time, about populations that were invisible only a few years ago (e.g. in developing countries) (Grimmer, 2015), as these populations have now access to the information technology necessary for big data collection and sharing (Hilbert, 2014).

Avoiding patient risk was mentioned as a value (V7) in 9.8% of the articles. It refers to techniques (e.g. logistic regression models and regressions trees) that predict populations at high disease risk (e.g. from secondary analysis of EMRs) and the offering of preventive healthcare (Kite et al., 2015), or the likelihood of hospital re-admissions of patients and the recommendation of home monitoring, which promotes people's health but also saves money from unnecessary hospitalization and treatment (Kulkarni et al., 2016). This value also refers to the prediction of daily events that challenge the health, security and sustainable growth of our society and prevent social and economic vulnerabilities (Boulos et al., 2010), such as violent behavior, terrorism attacks, emerging infectious diseases, etc.

The value (V8) of segmenting populations to customize actions comes next (9%). Here, we refer to benefits that organizations gain by capturing the share of new markets deriving from differentiating populations' characteristics through clustering and other techniques and offering products or services tailored to the specific segments' needs, such as a cloud based solution (Software as a Service) that provides personalized recommendations about the health insurance plans according to the user specified criteria (Abbas et al., 2015). There are also societal benefits by identifying isolated or socially excluded patients, e.g. HIV patients, and offering services that can bring them together (Bram et al., 2015). The value (V9) of achieving cost-effectiveness holds 9% and describes the benefit of analytics to offer solutions for reducing organizations' expenditures by optimizing sources utilization, or capturing underpayments, etc. (Bradley and Kaplan, 2010) while upholding quality. These solutions could include business intelligence visualization and collaborative tools for identifying and eliminating non-value-added processes in patients with chronic diseases by tracking patient data during home, ambulatory and hospital care over-time. On a societal level, decision-support tools, based for example on machine learning, can provide policy-makers with more "granular information about the health of the population, the prevalence and geography of local factors that are shaping community health and where the greatest potential return on investment might lie if confirmatory research supports a causal link" (Lary et al., 2014). The hope is that policy-makers can use the freed resources to interventions that act positively to society or help poor communities buffer the adverse health effects (Lary et al., 2014), such as organize health education programmes, and use public money towards high-risk, low-income patients that cannot afford treatment.

The last value (V10) focuses on data security (5.1%), which is enabled by BDA. It may include protection from privacy breaches, securing data anonymity of EMRs, etc., suggesting the gain for organizations and society to protect people's privacy (Fabian et al., 2015). The methods of protecting patient confidentiality also helps preventing unethical targeting of groups on the basis of race, ethnicity, or socio-demographics (Clift et al., 2014). However, we believe that individuals are likely to accept the ‘dark side’ of datafication through digital traces and constant monitoring through sensors, because they are persuaded that the benefits outweigh the costs. Businesses and governments try to send to citizens the message that security is more important than privacy (e.g. for fighting terrorism or an epidemic outbreak) (Newel and Marabelli, 2015).

4.2. Challenges from the implementation of BDA in healthcare industry

Table 2 presents seven challenge categories from the use of BDA in healthcare, three of which derived from Wamba et al. (2013) review (Ch1, Ch2, Ch4) and another four were added as identified from the article pool. Most of the issues deal with data management, security and privacy (Ch1), with 20% representation (mentioned in 161 articles). In this category, we distributed all articles that mentioned ethical issues related to informed consent for sharing, aggregating, or repurposing data related to patients. This further relates to the individuals' right to maintain their privacy, and the right to be forgotten by erasing their personal data from health and other organizations' databases and from bio/nano and mobile technologies that allow pervasive computing for personalized health (Blobel et al., 2016). Concerns are raised in the literature about the possibility of re-identification of anonymized sensitive information through cross-referencing or about group-level ethical issues from the analysis of aggregate data, as research outcomes may favor populations (usually westernized) from whom data is collected. Furthermore, the capability of the algorithms and monitoring systems to identify relationships between behaviors and individuals raise concerns about the possibly use of this data for the “stigmatization” of social groups or individuals (Rich and Miah, 2017). Other ethical challenges involve the increasing complexity of big dataset analysis, the access to the technology/platforms/tools used for this purpose and the ownership of these big datasets, which raise issues of outcomes' validity and capability of replicating the findings of BDA. An issue of paramount importance is people's fallacy that BDA diagnostic applications over the internet may replace the need to be seen by a doctor for diagnosis and drug prescription. The existence of self-diagnosis apps will impact several important dimensions of patienthood and healthcare (Lupton and Jutel, 2015). This is both a challenge and a threat to population's health.

Next (Ch2) are technological issues (10.2%), which include challenges stemming from the lack of the required infrastructure to achieve the expected outcomes. On the optimistic side, technology brings advances and data that are produced more efficiently (Alyass et al., 2015). However, missing infrastructure is identified in cases where innovative analytics data-gathering platforms are absent, or information systems are poorly connected across and within healthcare organizations (Zhu et al., 2015). The challenges concerning the society are mainly referred to the inequality among those that have the privilege of the “know-how”, such as technical specialists who know how to interpret and use information technology, and those who have not access to (or knowledge of) more complex systems (Cuquet and Fensel, 2018).

Under Ch3, “further evaluation issues” (7.1%) are studies in which their authors stated issues concerning their sample, for example the sample is not representative of the general population (Dimitriadis et al., 2015) or the data is inaccurate or with noise, etc. or there is bias in the research with regard to controlling the threats to internal validity or the risk of individual observer bias (Gruebner et al., 2017). Inadequate samples can create confusing outcomes.

Organizational and financing issues (Ch4) are identified in 25 studies (3.1%). These studies report lack of cost-benefit analysis frameworks for evaluating the worthwhileness of the use of BDA for decision-making; lack of training of information analysts to apply big data analytics techniques in healthcare settings and of healthcare practitioners to comprehend the data analysis and results (Barkley et al., 2013); lack of skills to support IT-enabled healthcare processes, such as telemedicine; and finally, organizational complexity, such as handling personal healthcare data deriving from different sources (labs, home devices, mobile applications, wearables and other) (Carroll et al., 2014).

A small number of papers (5–0.6%) have been distributed in the category titled “Regulatory issues” (Ch5). These include challenges for BDA due to the lack of new norms to integrate working practices, aligned with the new technologies. For example, for the creation of effective bio-surveillance systems, the governments should provide to the healthcare agencies “appropriate jurisdiction to exhibit secure, continuous information flow with no latency across jurisdictional boundaries and to enable detection of previously difficult-to-detect events that span public health authorities” (Velsko and Bates, 2016). Four articles (0.5%) have been classified under the limited awareness and support dimension (Ch6) and discuss the lack of funding for the completion of projects related to BDA in healthcare, and the lack of awareness for BDA and their benefits to healthcare organizations and to their decision-makers (Celler et al., 2014). On the society side, dependency on private and not public funding for the execution of health big data projects will create a few big players in the field which will direct research towards their individual goals which not necessarily overlap with society's goals (Cuquet and Fensel, 2018). The last subcategory (Ch7) refers to political responsibility (noted in 3 articles, 0.4%) and the obstacles posed by political sources in the implementation of IT for withholding power and the lack of regulatory responsibility in the case of misdiagnosis and who should be liable for the adverse medical outcome: the developers of the software, the technology provider, the hospital that uses the technology, the doctor, or all of the above? (Dilsizian and Siegel, 2014). All these challenges were discussed by researchers as barriers to the implementation of their proposed approaches or to the outcomes of their research. However, more than half of the articles in our dataset do not refer to challenges.

4.3. Future perspectives as derived through the article content analysis

Based on content analysis, we mined all the future goals of every article in our dataset and we grouped the future directions of research in health analytics under three main headings: “technological”, “organizational and “research” perspectives providing a map of “the road ahead” (see Table 3). Despite differences in each research's approach, the future perspectives were in many cases similar.

Not surprisingly, since the volumes of health data will grow globally in an intense manner, and the demand for IT infrastructure will consequently increase (Abbas et al., 2015), technology has the greatest role in the future of BDA in health. Under the “technological perspective,” many researchers reported their priority to develop the technological approach, as described or evaluated in their study, so that in the future more advanced versions of their methods will become available. Researchers identified the parts of their methods that have the potential for further improvement with other experts' contribution and expect more innovative techniques for large dataset exploitation and new platforms or new mechanisms, to accrue the maximum value of data. They also expect from technology to extend systems' capabilities and to improve the accuracy of health data advancing risk adjustment. Innovative digital media technologies are positioned forward for healthcare, to provide better, more informed and more economically-efficient

Table 3
Future aspects of BDA in healthcare.

Technological perspectives	N	%	Organizational and societal perspectives	N	%	Research perspectives	N	%
Development of the specific approach	133	16.5	Need for privacy	37	4.6	More studies to prove the hypothesis	57	7.1
Creation of new mechanisms to accrue maximum value of data	101	12.6	Training and education of clinicians and public	20	2.5	Propose an approach that can be used in other healthcare applications	31	3.9
Hardware and software development -extend systems capabilities	87	10.9	Integrating environmental factors in analytics for decision-making	10	1.2	Patient involvement in effectiveness research	15	1.9
Improve risk adjustment	41	5.1	Change the protocols and define policy purposes	9	1.1	To replicate same methods in other countries	14	1.7
Identify data elements that can be automatically corrected	21	2.6	More investment in infrastructure	5	0.6	Create value to other disciplines	12	1.5
Alternate the proposed approaches	17	2.1	National investments on health monitoring	4	0.5			
			Cost effective analysis of the new tool	3	0.4			
			Create partnerships among stakeholders to establish the value of BD	2	0.2			

medical treatment for certain groups (Lupton, 2014) also reducing the misallocation of resources (e.g. track patients across service sites, aggregate a bigger amount of data, give a more comprehensive view of data, make the medical record accessible to all caregivers, etc. (Barkley et al., 2013). As it has been declared there is a definite need in health care for systems that support or improve the decision-making ability of clinical experts, specifically, to diagnose complex diseases or pathologies (López-Martínez et al., 2018). Lastly several researchers recognized the space for alternation in their computational approach in terms of proposing new modalities to successfully provide more sufficient results.

Amongst the "organizational perspectives", the most relevant are the future changes in the field of personal data privacy by the development of systems which will standardize and secure the process of extracting anonymized healthcare datasets from healthcare organizations (Al-Shaqi et al., 2016). Ensuring privacy and cybersecurity will enable healthcare organizations and researchers to manipulate these datasets for further value creation (Al-Shaqi et al., 2016). Another future direction is the need of healthcare organizations to train and educate both clinicians and the public regarding the new age of datafication and create "the digitally engaged patient" (Lupton, 2014) who have an active role in producing and consuming information about health and medicine by using digital technologies. Organizations also are expected to integrate environmental factors into the analytics process and decision-making in order to detect environmental hazards, such as gas leakage, or to reduce environmental impact by enabling automatic operation of bathroom/corridor lights, reducing trips and minimizing patient falls and other. Using innovative and computer-aided diagnosis systems and questioning old practices will bring new societal norms and will create the need for new protocols for the treatment of patients (Rodrigues et al., 2016) and the relationship between patients and clinicians. Given the gains from health BDA, it is expected that more investments will be provided to IT infrastructure and to individuals with appropriate interest/expertise from healthcare organizations or from nations for healthcare monitoring, such as automated bio-surveillance systems at a national scale (Velsko and Bates, 2016). Scientists also stated that it would be useful to focus more on the design of BDA recommendations, e.g. drug prescription patterns, which affect cost outcomes (Bjarnadottir et al., 2016). As of right now, little attention is given to organizational future perspective in the establishment of broad partnerships between manufacturers, payers, providers, and regulators in the health-care system to demonstrate and communicate the overall value in medicine that big data could bring (Szlezak et al., 2014).

The "research perspectives" focus on researchers' need for more studies to prove their hypothesis and belief that their approach could also create value to other healthcare applications. For example, real-time sensor data from patient's devices can further empower the clinical decision support, for the diabetes case, but also for other complex medical conditions, such as Alzheimer's and psychosis (De Silva et al., 2015). Other researchers envision a future where big data from large groups of patients can be pooled from across institutions so that each patient and their clinicians can find 'patients like me' to help with real-time clinical decision-making (Broughman and Chen, 2016). Many authors claim their data analytic approaches must be tested with patient demographic characteristics from other regions/countries to expand the scope of their models (Bardhan et al., 2015) and provide assumptions in a worldwide level and must also be tested in other scientific areas. For example, the interpretation of postgenomics data using certain algorithms is expected to be the center of knowledge-based innovations in various big data fields, such as precision medicine, nutrigenomics, vaccinomics, pharmacogenomics, ecogenomics (Ben-Ari Fuchs et al., 2016).

5. Limitations

This research has several limitations. For the systematic review, articles were only obtained from Web of Science® and Scopus. However, these are the world's largest multidisciplinary databases and the two mostly used in literature search (Aghaei Chadegani et al., 2013) and comprise citations from other databases, such as MEDLINE and Biological abstracts. Furthermore, this study uses as a methodology a systematic literature review approach which could be broaden to include many other aspects of health sociology, such as the effects of health data commercialization, the changing environment of labor etc. Sociology can shed light on some of the identified challenges such as data ownership, revealing where points of exploitation occur, and on issues of healthcare providers' responsibilities and their capacity to enforce or discourage certain behaviors. Thus, we hope that this study will be an inspiration for future research in the recommended fields.

6. Discussion - conclusions

Modernity has led to many changes in everyday social life. A remarkable change has been the expansion of medical activity through medical innovation in a variety of new areas (Lowton et al., 2017). In our systematic review of the literature of 804 papers, we followed the resource–based view theory to identify the created values from the big data resources and capabilities in the health field and extended our research framework to further investigate the emerging challenges and the field's future perspectives. We formulated categories based on which we mapped the literature and identified the number of articles discussing each dimension of values, challenges and future perspectives after content analysis using the NVivo software. The results were presented and discussed through representative examples which focussed both on the organizational and social impact of BDA in health.

The most highly discussed value is that analytics provide “new approaches for the diagnosis and prognosis of personalized medicine”. It includes the process of genetic profiling to offer individual health information for a variety of diseases (e.g., cardiovascular disease, cancers, diabetes, etc.), enabling the use of personalized therapeutic schemes. This value also enables early detection of factors that could create a negative social impact (e.g. people with stress issues, anti-social behavior, etc.). Another important value, is the capability to “support human decision-making with automated algorithms”. Researchers need to quantify the real value of the analysis of the available data in everyday clinical practice. This, on one hand, creates new employment conditions and the demand for the acquisition of IT skills which in turn will lead to changing educational curriculums for health professionals, and on the other hand, changes the doctor-patient relationship, giving more information to patients to challenge doctors' knowledge.

The most highly stated challenges are related to “data management, security and privacy issues” in line with the findings of Wamba et al. (2013). The topic of data privacy has become increasingly important nowadays because of the rapid development of new forms of data, and the ease of transferring and sharing data. Data Protection legislation differs between countries as each country protects medical and health-related data at different level. For example, the rising concerns about data privacy have led to the General Data Protection Regulation (GDPR), enforceable from 2018, which strengthens data protection for all individuals within the European Union, and makes the export process of personal data outside the EU more rigorous. Under this regulation, pseudonymized data are still considered personal data, which means that more BDA health projects that use pseudonymization will now require either consent or authorization (Rumbold, 2017). Issues related to technological challenges come next and mostly relate to a lack of appropriate infrastructure for supporting BDA in health-related organizations and to the inequality among specialists' knowledge of using BDA and their access to relevant IT (Cuquet and Fensel, 2018).

Conclusively, to take advantage of the use of BDA, and reduce risk from “false positives,” it is certainly important to set the focus on the exploration of the information provided from big data (i.e. what information, from which sources, for what purpose it was collected, what is the intention of the analysis, what should be explored) (Strauβ, 2015). It is unknown the extent of how digital technology and BDA are going to impact society and business in the long term. Enhancing the breadth and depth of our knowledge about the major and minor aspects of the related field assists the healthcare community in identifying new approaches to strengthen outcomes. Methods, such as machine learning, a problem-solving/pattern-matching tool, creates new perspectives in healthcare (Obermeyer and Ezekiel, 2016). This is because the intelligent elaboration of more and more data brings new evolution for the prognosis, diagnosis and treatment of diseases. After all, the desire of medical experts is to create new decision support systems that reflect upon their “intuitive thinking” and minimize or even eliminate “personal biases” (Oztekin et al., 2018).

A stated threat for the future is that algorithmic decision-making may lead to an extremely superficial understanding of why things happen, as answers will be prescribed from a “black box.” This will prohibit decision-makers (such as clinicians) to build cumulative knowledge on phenomena and diseases, which consequently, may cause them to lose their capacity to make decisions on their own (Newel and Marabelli, 2015), and therefore, be fully replaced by big data analytics. Big Data research should begin with a clear understanding of the value it can bring (Flechet et al., 2016). We believe that the demonstration of the “bright and the dark side” of the datafication in the healthcare industry can shed light to some of its dilemmas. Mapping the existing literature can facilitate health-related organizations and the society to recognize at first the impact from big data analytics and then revalue strategies, mitigate risks, and draw upon new opportunities for further development.

References

Abbas, A., Bilal, K., Zhang, L., Khan, S.U., 2015. A cloud based health insurance plan recommendation system: a user centered approach. Future Gener. Comput. Syst. 43, 99–109.
Aghaei Chadegani, A., Salehi, H., Yunus, M., Farhadi, H., Fooladi, M., Farhadi, M., Ale Ebrahim, N., 2013. A comparison between two main academic literature collections: web of Science and Scopus databases. Asian Soc. Sci. 9 (5), 18–26.
Al-Shaqi, R., Mourshed, M., Rezgui, Y., 2016. Progress in ambient assisted systems for independent living by the elderly. SpringerPlus 5, 624.
Alyass, A., Turotte, M., Meyre, D., 2015. From big data analysis to personalized medicine for all: challenges and opportunities. BMC Med. Genomics 8, 33.
Amankwah-Amoah, J., 2016. Emerging economies, emerging challenges: mobilising and capturing value from big data. Technol. Forecast. Soc. Chang. 110, 167–174.
Angulo, D.A., Schneider, C., Oliver, J.H., Charpak, N., Hernandez, J.T., 2016. A multi-facetted visual analytics tool for exploratory analysis of human brain and function datasets. Front. Neuroinf. 10, 36.
Banos, O., Amin, M.B., Khan, W.A., Afzal, M., Hussain, M., Kang, B.H., Lee, S., 2016. The Mining Minds digital health and wellness framework. Biomed. Eng. Online 15 (1), 76.
Bardhan, I., Oh, J., Zheng, Z., Kirksey, K., 2015. Predictive analytics for readmission of patients with congestive heart failure. Inf. Syst. Res. 26 (1), 19–39.
Barkley, R., Greenapple, R., Whang, J., 2013. Actionable Data Analytics in Oncology: Are We There yet? Cancer Center Business Development Group, pp. 93–96.
Ben-Ari Fuchs, S., Lieder, I., Stelzer, G., et al., 2016. GeneAnalytics: an integrative gene set analysis tool for next generation sequencing, RNAseq and microarray data. OMICS A J. Integr. Biol. 20 (3), 139–151.
Benharref, A., Serhani, M.A., Al Ramzana, N., 2014. Closing the loop from continuous M-health monitoring to fuzzy logic-based optimized recommendations. IEEE 2014, 2698–2701.
Bharadwaj, A.S., 2000. A Resource-based view Perspective on information Technology capability and firm performance: an empirical investigation. MIS Q. 24 (1), 169–196.
Bjarnadottir, M.V., Malik, S., Onukwugha, E., Gooden, T., Plaisant, C., 2016. Understanding adherence and prescription patterns using large-scale claims data. PharmacoEconomics 34 (2), 169–179.
Blobel, B., Lopez, D.M., Gonzalez, C., 2016. Patient privacy and security concerns on big data for personalized medicine. Health Technol. 6 (1), 75–81.
Boulos, M.N.K., Sanfilippo, A.P., Corley, C.D., Wheeler, S., 2010. Social Web mining and exploitation for serious applications: technosocial Predictive Analytics and related technologies for public health, environmental and national security surveillance. Comput. Methods Progr. Biomed. 100 (1), 16–23.
Boytcheva, S., Angelova, G., Angelov, Z., Tcharaktchiev, D., 2015. Text mining and big data analytics for retrospective analysis of clinical texts from outpatient care. Cybern. Inf. Technol. 15 (4), 58–77.

Bradley, P., Kaplan, J., 2010. Turning hospital data into dollars. Healthc. Financ. Manag. 64–68.
Bram, J.T., Warwick-Clark, B., Obeysekare, E., Mehta, K., 2015. Utilization and monetization of healthcare data in developing countries. Big Data 3 (2), 59–66.
Broughman, J.R., Chen, R.C., 2016. Using big data for quality assessment in oncology. J. Comp. Effect. Res. 5 (3), 309–319.
Carroll, L.N., Au, A.P., Detwiler, L.T., Fu, T.C., Painter, I.S., Abernethy, N.F., 2014. Visualization and analytics tools for infectious disease epidemiology: a systematic review. J. Biomed. Inform. 51, 287–298.
Celler, B.G., Sparks, R., Nepal, S., Jayasena, R., 2014. Design of a multi-site multi-state clinical trial of home monitoring of chronic disease in the community in Australia. BMC Public Health 14 (1), 1270.
Chae, B., Olson, D.L., 2013. Business analytics for supply chain: a dynamic-capabilities framework. Int. J. Inf. Technol. Decis. Mak. 12 (1), 9–26.
Chen, C.P., Zhang, C.Y., 2014. Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. Inf. Sci. 275, 314–347.
Chen, H., Chiang, R., Storey, V.C., 2012. Business intelligence and analytics: from big data to big impact. MIS Q. 36 (4), 1165–1188.
Clift, K., Scott, L., Johnson, M., Gonzalez, C., 2014. Leveraging geographic information systems in an integrated health care delivery organization. Perm. J. 18 (2), 71.
Cuquet, M., Fensel, A., 2018. The societal impact of big data: a research roadmap for Europe. Technol. Soc. 54, 74–86.
Davenport, T.H., 2006. Competing on analytics. Harv. Bus. Rev. 84 (11), 98.
De Silva, D., Burstein, F., Jelinek, H., 2015. Addressing the complexities of big data analytics in healthcare: the diabetes screening case. Aust. J. Inf. Syst. 19, 99–115.
Dilsizian, S.E., Siegel, E.L., 2014. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Curr. Cardiol. Rep. 16 (1), 441.
Dimitriadis, S.I., Laskaris, N.A., Bitzidou, M.P., Tarnanas, I., Tsolaki, M.N., 2015. A novel biomarker of amnestic MCI based on dynamic cross-frequency coupling patterns during cognitive brain responses. Front. Neurosci. 9, 350.
Dinov, I.D., 2016. Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data. GigaScience 5, 12.
Doumpos, M., Zopounidis, C., 2016. Editorial to the special issue “business analytics”. Omega (59), 1–3.
Evangelatos, N., Reumann, M., Lehrach, H., Brand, A., 2016. Clinical trial data as public goods: fair trade and the virtual knowledge bank as a solution to the free rider problem-A framework for the promotion of innovation by facilitation of clinical trial data sharing among biopharmaceutical companies in the era of omics and big data. Publ. Health Genom. 19 (4), 211–219.
Fabian, B., Ermakova, T., Junghanns, P., 2015. Collaborative and secure sharing of healthcare data in multi-clouds. Inf. Syst. 48, 132–150.
Feldman, B., Martin, E.M., Skotnes, T., 2016. Big data in healthcare hype and hope. Dr. Bonnie 360.
Flechet, M., Grandas, F.G., Meyfroidt, G., 2016. Informatics in neurocritical care: new ideas for Big Data. Curr. Opin. Critic. Care 22 (2), 87–93.
Grimmer, J., 2015. We are all social scientists now: how big data, machine learning, and causal inference work together. PS Political Sci. Polit. 48 (1), 80–83.
Groves, P., Kayyali, B., Knott, D., Van Kuiken, S., 2013. The ‘big data’revolution in healthcare. McKinsey Q. 2, 3.
Gruebner, O., Sykora, M.D., Lowe, S.R., Shankardass, K., Galea, S., Subramanian, S.V., 2017. Big data opportunities for social behavioral and mental health research. Soc. Sci. Med. 189, 167–169.
Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S.F., Childe, S.J., Hazen, B., Akter, S., 2017. Big data and predictive analytics for supply chain and organizational performance. J. Bus. Res. 70, 308–317.
Hilbert, M., 2014. Technological information inequality as an incessantly moving target: the redistribution of information and communication capacities between 1986 and 2010. J. Assoc. Inf. Sci. Technol. 65 (4), 821–835.
Hughes, A., Landers, D., Arkenau, H.T., et al., 2016. Development and evaluation of a new technological way of engaging patients and enhancing understanding of drug tolerability in early clinical development: PROACT. Adv. Ther. 33 (6), 1012–1024.
Ivan, M., Velicanu, M., 2015. Healthcare industry improvement with business intelligence. Inf. Econ. 19 (2), 8.
Kite, B.J., Tangasi, W., Kelley, M., Bower, J.K., Foraker, R.E., 2015. Electronic medical records and their use in health promotion and population research of cardiovascular disease. Curr. Cardiovasc. Risk Rep. 9 (1), 422.
Kulkarni, P., Smith, L.D., Woeltje, K.F., 2016. Assessing risk of hospital readmissions for improving medical practice. Health Care Manag. Sci. 19 (3), 291–299.
Kumar, S., Abowd, G.D., Abraham, W.T., et al., 2015. Center of excellence for mobile sensor data-to-knowledge (MD2K). J. Am. Med. Inform. Assoc. 22 (6), 1137–1142.
Lary, D.J., Woolf, S., Faruque, F., LePage, J.P., 2014. Holistics 3.0 for health. ISPRS Int. J. Geo-Inf. 3 (3), 1023–1038.
Loebbecke, C., Picot, A., 2015. Reflections on societal and business model transformation arising from digitization and big data analytics: a research agenda. J. Strateg. Inf. Syst. 24 (3), 149–157.
López-Martínez, F., Schwarcz, A., Núñez-Valdez, E.R., García-Díaz, V., 2018. Machine learning classification analysis for a hypertensive population as a function of several risk factors. Expert Systems with Applications.
Lowton, K., Hiley, C., Higgs, P., 2017. Constructing embodied identity in a ‘new’ ageing population: a qualitative study of the pioneer cohort of childhood liver transplant recipients in the UK. Soc. Sci. Med. 172, 1–9.
Lupton, D., 2014. The commodification of patient opinion: the digital patient experience economy in the age of big data. Sociol. Health Illn. 36 (6), 856–869.
Lupton, D., Jutel, A., 2015. ‘It's like having a physician in your pocket!’A critical analysis of self-diagnosis smartphone apps. Soc. Sci. Med. 133, 128–135.
Mohan, M., Vigneshwaran, B., Vineeth Raj, G., Harlin Jesuva Prince, S., 2016. Disease diagnosis for personalized health care using map reduce technique. IJCTA 9 (5), 2153–2164.
Mooney, S.J., Pejaver, V., 2018. Big data in public health: terminology, machine learning, and privacy. Annu. Rev. Public Health 39, 95–112.
Mortenson, M.J., Doherty, N.F., Robinson, S., 2015. Operational research from Taylorism to Terabytes: a research agenda for the analytics age. Eur. J. Oper. Res. 241 (3), 583–595.
Newel, S.l., Marabelli, M., 2015. Strategic opportunities (and challenges) of algorithmic decision-making: a call for action on the long term societal effects of “datification”. J. Strateg. Inf. Syst. 24 (1), 3–14.
O'Connor, C., Kadianaki, I., Maunder, K., McNicholas, F., 2018. How does psychiatric diagnosis affect young people's self-concept and social identity? A systematic review and synthesis of the qualitative literature. Soc. Sci. Med. 212, 94–119.
Obermeyer, Z., Ezekiel, J.E., 2016. Predicting the future - big data, machine learning, and clinical medicine. N. Engl. J. Med. 375, 13.
Oztekin, A., Al-Ebbini, L., Sevkli, Z., Delen, D., 2018. A decision analytic approach to predicting quality of life for lung transplant recipients: A hybrid genetic algorithms-based methodology. Eur. J. Oper. Res. 266 (2), 639–651.
Raghupathi, W., Raghupathi, V., 2013. An overview of health analytics. J. Health Med. Inf. 4 (132), 2.
Rich, E., Miah, A., 2017. Mobile, wearable and ingestible health technologies: towards a critical research agenda. Health Sociol. Rev. 26 (1), 84–97.
Rodrigues Jr., J.F., Paulovich, F.V., de Oliveira, M.C., de Oliveira Jr., O.N., 2016. On the convergence of nanotechnology and Big Data analysis for computer-aided diagnosis. Nanomedicine 11 (8), 959–982.
Rumbold, J.M., 2017. The effect of the general data protection regulation on medical research. J. Med. Internet Res. 19 (2).
Strauß, S., 2015. Datafication and the seductive power of uncertainty—a critical exploration of big data enthusiasm. Information 6 (4), 836–847.
Szlezak, N., Evers, M., Wang, J., Pérez, L., 2014. The role of big data and advanced analytics in drug discovery, development, and commercialization. Clin. Pharmacol. Therapeut. 95 (5), 492–495.
Velsko, S., Bates, T., 2016. A conceptual architecture for national biosurveillance: moving beyond situational awareness to enable digital detection of emerging threats. Health Secur. 14 (3), 189–201.
Waller, M.A., Fawcett, S.E., 2013. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. J. Bus. Logist. 34 (2), 77–84.
Wamba, S.F., Anand, A., Carter, L., 2013. A literature review of RFID-enabled healthcare applications and issues. Int. J. Inf. Manag. 33 (5), 875–891.
Wamba, S.F., Akter, S., Edwards, A., Chopin, G., Gnanzou, D., 2015. How ‘big data’can make big impact: findings from a systematic review and a longitudinal case study. Int. J. Prod. Econ. 165, 234–246.
You, Q., Shiaofen, F., Jake, Y.C., 2008. Gene Terrain: visual exploration of differential gene expression profiles organized in native biomolecular interaction networks. Inf. Vis. 9 (1), 1–12.
Zhu, K., Lou, Z., Zhou, J., Ballester, N., Kong, N., Parikh, P., 2015. Predicting 30-day hospital readmission with publicly available administrative database. Methods Inf. Med. 54 (06), 560–567.