International Journal of Medical Informatics 114 (2018) 57–65

Contents lists available at ScienceDirect

International Journal of Medical Informatics

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

Review article

Concurrence of big data analytics and healthcare: A systematic review

Nishita Mehta[a,*], Anil Pandit[b]

[a] Symbiosis International University, Pune, India
[b] Symbiosis Institute of Health Sciences, Pune, India

ARTICLE INFO

Keywords:
Big data
Analytics
Healthcare
Predictive analytics
Evidence-based medicine

ABSTRACT

Background: The application of Big Data analytics in healthcare has immense potential for improving the quality of care, reducing waste and error, and reducing the cost of care.
Purpose: This systematic review of literature aims to determine the scope of Big Data analytics in healthcare including its applications and challenges in its adoption in healthcare. It also intends to identify the strategies to overcome the challenges.
Data sources: A systematic search of the articles was carried out on five major scientific databases: ScienceDirect, PubMed, Emerald, IEEE Xplore and Taylor & Francis. The articles on Big Data analytics in healthcare published in English language literature from January 2013 to January 2018 were considered.
Study selection: Descriptive articles and usability studies of Big Data analytics in healthcare and medicine were selected.
Data extraction: Two reviewers independently extracted information on definitions of Big Data analytics; sources and applications of Big Data analytics in healthcare; challenges and strategies to overcome the challenges in healthcare.
Results: A total of 58 articles were selected as per the inclusion criteria and analyzed. The analyses of these articles found that: (1) researchers lack consensus about the operational definition of Big Data in healthcare; (2) Big Data in healthcare comes from the internal sources within the hospitals or clinics as well external sources including government, laboratories, pharma companies, data aggregators, medical journals etc.; (3) natural language processing (NLP) is most widely used Big Data analytical technique for healthcare and most of the processing tools used for analytics are based on Hadoop; (4) Big Data analytics finds its application for clinical decision support; optimization of clinical operations and reduction of cost of care (5) major challenge in adoption of Big Data analytics is non-availability of evidence of its practical benefits in healthcare.
Conclusion: This review study unveils that there is a paucity of information on evidence of real-world use of Big Data analytics in healthcare. This is because, the usability studies have considered only qualitative approach which describes potential benefits but does not take into account the quantitative study. Also, majority of the studies were from developed countries which brings out the need for promotion of research on Healthcare Big Data analytics in developing countries.

1. Introduction

Over the last decade, there has been a rapid digitalization across the industries. Healthcare has also undergone this digital transformation with an increase in use of Electronic Medical Records (EMRs); Healthcare Information Systems (HIS); and handheld, wearable and smart devices. As a result, a massive amount and variety of health-related data today, is in digital form, which includes −omics data, socio-demographics data and insurance claims data apart from clinical data. This high-quality healthcare data offers potential value for optimizing care delivery, but it is still “perceived as a by-product of healthcare delivery, rather than a central asset source for competitive advantages” [1]. As the electronic health data remains largely underutilized and hence wasted [2], there is a need for converting the raw data into meaningful and actionable information [3,4].

Much of the highly valuable healthcare data is in unstructured or semi-structured form. Added to it, the complex, dynamic and heterogeneous characteristics of the data [5–7] renders it difficult to extract useful information using traditional data analytical tools & techniques [8]. In fact, there is a finite human ability to process this data without effective decision support [9]. This creates the need for integration of Big Data analytics into healthcare. Big Data analytics has the ability to

* Corresponding author.
E-mail address: nishitamehta@sihspune.org (N. Mehta).

https://doi.org/10.1016/j.ijmedinf.2018.03.013
Received 22 January 2018; Accepted 23 March 2018
1386-5056/

analyze a wide variety of complex data and generate valuable insights which otherwise would not have been possible. When applied to the healthcare data, it has the potential to identify patterns and lead to improved healthcare quality & reduced costs and enable timely decision-making [6,8,10,11]. As per the report by McKinsey Global Institute [12], by utilizing Big Data effectively, US Healthcare can create a value of more than $300 billion every year, of which two-third would be in the form of reducing healthcare expenditure by about 8%. Using Big Data technology, hidden knowledge can be uncovered using automated analysis of outcomes [13].

Advancement in cloud computing and increased deployment of EMRs enable easy access to longitudinal patient data [14]. The integration of longitudinal patient data with data from disparate, structured and unstructured Big Data sources offer the potential of comprehensive understanding of diseases at a considerably higher pace [15,16]. The ability of Big Data analytics to identify disease heterogeneity allows quick and accurate diagnosis and assessment of therapies [3,17–19]. By linking data from different sources and discerning patterns, the predictive power of Big Data analytics can also be used for transforming continuous real-time data into valuable information. This is of utmost importance in emergency medical situations as it can mean the difference between life and death [10].

The promising value of Big Data technology in healthcare has created an increasing interest of academic & industry investigators. Nevertheless, there have been only a few literature reviews and the literature remains largely fragmented. The purpose of this research therefore is to gain a comprehensive understanding of current outlook on this technology. It aims at answering the research question on: *How "Big Data analytics" fits in healthcare environment to enhance its value?* Accordingly, this review explores the conceptual aspect of applying Big Data analytics to healthcare and its significance in enhancing care delivery and business worth. It also describes the challenges posed in leveraging Big Data analytics in healthcare and the need for approaches to overcome them.

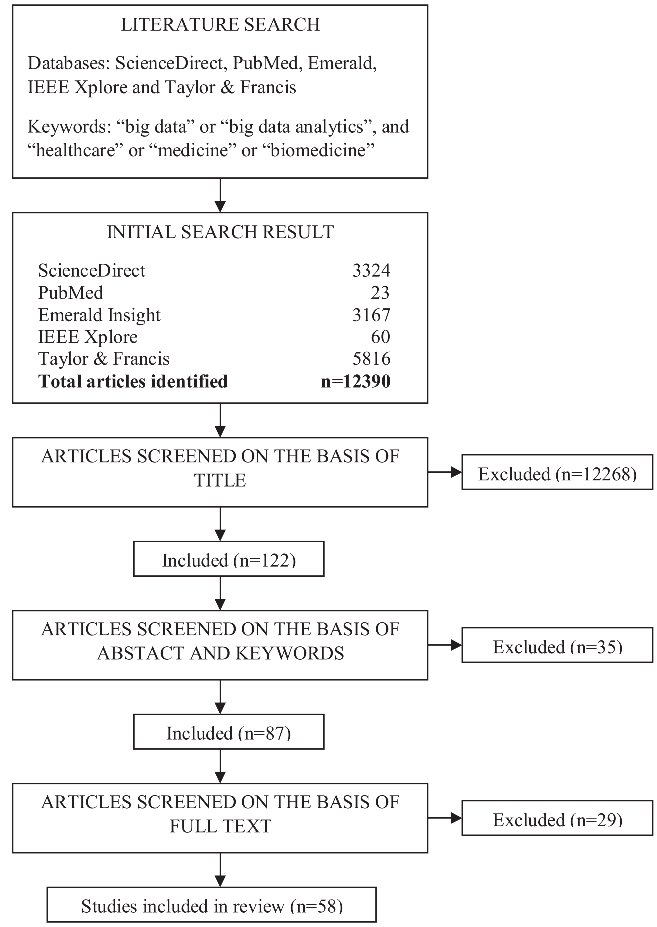

Fig. 1. Research Process.

2. Review method

A systematic review was conducted for capturing relevant literature from different sources, focusing on the following objectives:

- To determine different perspectives to definition and concepts of Big Data in healthcare
- To explore the sources of Big Health Data
- To identify Big Data analytical techniques and technologies in healthcare
- To illustrate the potential benefits and applications of Big Data within healthcare
- To present strategies for tackling the challenges of Big Data application within healthcare

By investigating these objectives in detail, this review will make a significant contribution in understanding the overall context and the future application of Big Data techniques in the healthcare domain.

2.1. Information sources

A search for articles was made on following databases: ScienceDirect, PubMed, Emerald, IEEE Xplore and Taylor & Francis. The references included in these articles were also scanned for a thorough review.

2.2. Selection criteria

To select the literature for inclusion in the literature review, following inclusion criteria were used:

IC 1: Articles that deal with Big Data analytics in healthcare

IC 2: Articles published between 2013 and 2018

IC 3: Only articles published in English

To capture the literature relevant to the research interest, the articles with their primary emphasis on traditional analytics in healthcare were excluded.

2.3. Study selection

The procedure for search and selection of research material was carried out in the following four phases (Fig. 1):

1. The search for publications on electronic databases containing keywords "big data" or "big data analytics", and "healthcare" or "medicine" or "biomedicine"
2. Scrutiny of the title, abstract and keywords of identified articles and selection of the significant articles on the basis of selection criteria
3. Perusal of articles that were not eliminated in the previous phase for the review
4. Scanning of cross-reference articles for detailed study

2.4. Quality assessment

During the review, activities ensuring the quality of the search process were undertaken. The web searches were made in incognito mode to avoid any influence of historical searches. From initial searches, the authors manually extracted relevant papers and articles. The analysis and evaluation of abstracts was carried out and the authors verified which articles were to be included or excluded from the study.

3. Results

The study selection process was followed for each of the databases. In ScienceDirect, Emerald and Taylor & Francis, the initial searches yielded a larger number of hits (3324, 3167 and 5816). While the search resulted in 23 articles on PubMed and 60 articles on IEEE Xplore. After the examination of title, a total of 122 articles were selected as per the selection criteria. The abstracts and keywords of these articles were scrutinized and 87 relevant articles were chosen for further study. The full text of these 87 articles was read. Articles discussing technical aspects of Big Data computing and statistical models for Big Data analytics in healthcare were beyond the scope of this review and were thus excluded. Of the 87 articles screened, a total of 58 articles were included in this review.

The literature included in this review comprises mainly of descriptive articles and usability studies. On the basis of the main research objectives, the content from these articles was extracted and the articles were organized into different groups: Big Data analytics definition and concepts, sources of Big Data in healthcare, Big Data techniques for healthcare analytics, application and potential benefits of Big Data in healthcare and challenges in Big Data analytics in healthcare. The following section summarizes the findings in each of these categories.

3.1. Concept of big data and its definitional perspectives

The concept of Big Data was framed in late 1990s when Michael Cox and David Ellsworth [20] considered visualization as a problem of Big Data. But one of the early definitions of Big Data was given by Francis X. Diebold [21] in 2000 when he referred to Big Data as “explosion in the quantity (and sometimes, quality) of available and potentially relevant data”. Later the key dimensions of Big Data – *volume, velocity and variety*: the 3Vs – were derived from the study of Doug Laney in 2001 [22]. Manyika et al. [12], in their report, regarded *value* as another important element of Big Data. With regards to Big Data in healthcare, Feldman et al. [23] introduced *veracity* as yet another critical feature. Several definitions of Big Data currently exist based on the characteristics of Big Data (Table 1). However, Scruggs et al. [24] state that the definition of Big Data is beyond the scope of these characteristics and should extend to its potential “to be useful and reused, accumulate value over time, and innovate a multi-dimensional, systems-level understanding”.

In healthcare, “Big Data includes heterogeneous, multi-spectral, incomplete and imprecise observations (e.g., diagnosis, demographics, treatment, prevention of disease, illness, injury, and physical and mental impairments) derived from different sources using incongruent sampling” [25]. Some of these data are structured and they focus on genotype, phenotype, genomics data, ICD codes [3,7,26]; but the unstructured data includes memos, clinical notes, prescriptions, medical imaging, EHRs, lifestyle, environmental, and health economics data [3,7,26,27]. The challenge for Big Data analytics is to deal with this heterogeneous data in order to generate insights for improved healthcare outcomes.

For defining Big Data in healthcare, Auffray et al. [26] focused on the types of healthcare data, while authors like Raghupathi & Raghupathi [10], Karen et al. [28], Tan et al. [29] emphasized on the requirement of analytical and management tools. According to Liyanage et al. [30], a quantitative definition of Big Data is difficult because the volume aspect of Big Data is relative to the time of definition and would change with the advancement in technologies. Hansen et al. [25] and Roski et al. [31] concentrate on its analytical ability. On the other hand, Bates et al. [18], Dinov [25] and Bian et al. [32] focuses on the characteristics of Big Data. These characteristics are defined as 5Vs: volume, velocity, variety, veracity and value.

The *volume* refers to the quantity of Big Data in healthcare, which is estimated to increase dramatically to 35 zettabytes by 2020 [33]. The *variety* refers to the different types of healthcare Big Data collected including their heterogeneous characteristics [8] and structured and unstructured nature of medical data [7]. The *velocity* is the speed of data generation (i.e. real-time patient data [7]) as well as data collection. The *veracity* refers to sources that influence accuracy such as inconsistencies, missing data, ambiguities, deception, fraud, duplication, spam and latency [34]. Veracity and data quality issues are of acute concern in healthcare because life or death decisions depend on having the accurate information [10]. Lastly, the *value* represents cost-benefit to the decision maker through the ability to take meaningful action based on insights derived from data [34]. These characteristics of medical data contribute to their complexity: diversity of health-related ailments and their co-morbidities; the heterogeneity of treatments and outcomes [21]; differences in clinical workflows, practice standards, patient populations, available technologies, and referral resources [35].

Specific to healthcare, Dinov [25] presented two important characteristics of Big Data: their energy and life-span. The *energy* encapsulates the holistic information content included in the data. Hence, the energy of the aggregated dataset is much higher than individual database, rendering it more beneficial for exploring associations. The *life-span* of Big Data is in terms of its value past time of acquisition which decays at an exponential rate. As per the definition of Big Data by Ghani et al. [36], Big Data is not limited by the purpose of answering a specific question. On the other hand, Baro et al. [37] defines healthcare big data in terms of the number of statistical individuals (n) and the number of variables (p). According to him, Big Data are datasets with Log (n * p) ≥ 7, and they have the properties of great variety and high

Table 1
Big Data Definitions.

Authors	Definition
Raghupathi & Raghupathi [10]	By definition, big data in healthcare refers to electronic health data sets so large and complex that they are difficult (or impossible) to manage with traditional software and/or hardware; nor can they be easily managed with traditional or common data management tools and methods
Auffray et al. [26]	“Big data in health” encompasses high volume, high diversity biological, clinical, environmental, and lifestyle information collected from single individuals to large cohorts, in relation to their health and wellness status, at one or several time points.
Bates et al. [18]	By big data, we refer to the high volume, variety, and potential for the rapid accumulation of data
Dinov [25]	Big healthcare data refers to complex datasets that have some unique characteristics, beyond their large size, that both facilitate and convolute the process of extraction of actionable knowledge about an observable phenomenon
Karen et al. [29]	Big Data is a term used to describe data sets with such large volume or complexity that conventional data processing methods are not good enough to deal with them.
Tan et al. [30]	Big data has been referred to as data that are too complex and large that cannot be processed and managed by traditional data processing tools
Hansen et al. [25]	In addition to just having more data, Big Data also generally refers to the application of machine learning for analyzing the data sets.
Roski et al. [31]	Big data–that is, the sophisticated and rapid analysis of massive amounts of diverse Information
Bian et al. [32]	Big data is commonly defined through the 4 Vs: volume (scale or quantity of data), velocity (speed and analysis of real-time or near-real-time data), variety (different forms of data, often from disparate data sources), and veracity (quality assurance of the data).
Ghani et al. [36]	The term big data refers to ultra-large bodies of data that have not been prospectively limited in size or scope by the intent to address specific research questions or disease conditions, and that grow continuously and rapidly.
Baro et al. [37]	Big data can be defined as datasets with Log (n∗p) ≥7. Properties of big data are its great variety and high velocity.

velocity.

By analyzing the literature, it is evident that although the significance of big data in strengthening healthcare is recognized and understood, there is still a lack of consensus on the operational definition of big data in healthcare. Thus, the examination of definitions from previous studies allows discernment of the common elements.

3.2. Sources of healthcare big data

Data in healthcare are disorganized and distributed, coming from various sources and having different structures and forms [38]. Healthcare Big Data includes data on physiological, behavioral, molecular, clinical, environmental exposure, medical imaging, disease management, medication prescription history, nutrition, or exercise parameters [26]. Some of the primary sources of Big Data in healthcare are administrative databases (insurance claims and pharmaceuticals), clinical databases, electronic health record data [11], and laboratory information system data [39]. The other sources of data [11] are biometric data (wearable or sensor generated [35]), patient-reported data (standardized health surveys), data from social media [35], medical imaging data, and biomarker data, including all the spectrum of 'omics' data (that is, genomic, proteomic, and metabolomic data). Miller [40] identifies two main sources of health Big Data to be genomics-driven Big Data (genotyping, gene expression, sequencing data) and payer–-provider Big Data (electronic health records, insurance records, pharmacy prescription, patient feedback and responses). On the other hand, Swan [41] categorized Big Data streams into (a) Traditional medical data obtained from EMRs, medication history and lab reports which assist in a better understanding of disease outcomes and optimizing healthcare delivery; (b) "Omics" data including genomics, microbiomics, proteomics, and metabolomics, which helps in understanding the mechanisms of diseases and accelerate the individualization of medical treatments (c) Data from social media, wearables & sensors which provides the information about behavior and lifestyle of individuals. Thus, the healthcare data comes from internal sources such as EMRs, CPOE, imaging data and biomedical data, as well as external data sources, such as government, insurance claims/billing, R&D laboratories, and social media [10,42].

According to Belle et al. [2], healthcare data is spread among different healthcare systems, health insurers, researchers, government entities. Huang et al. [43] recognizes that Big Data in precision medicine comes from four different stakeholders: (a) Government and large companies, (b) Smaller stakeholders such as academic groups and technology, biotech, and device startups, (c) Health care providers and payers, and (d) Not-for-profit foundations and patient advocacy groups.

Clinical data such as vital signs, past medical history, medications, immunizations and medical imaging can be derived from electronic health records, CPOE, clinical decision support systems medication administration records, laboratory and pharmaceutical records [10,44], cohort studies, government surveys & clinical trials [20]. Administrative data, on the other hand, contains patient demographic data and visit information, admit date, discharge date, ICD diagnosis & procedure codes, admit source, discharge disposition and claims data such as charges for the visit, payer and reimbursement [44]. Table 2 summarizes the sources of different types of healthcare data.

3.3. Big data analytical techniques and technologies in healthcare

The multi-dimensional healthcare data – medical images (X-ray, MRI images), biomedical signals (EEG, ECG, EMG etc.), audio transcripts, handwritten prescriptions and structured data from EMRs [45] − and its dynamicity and complexity makes it difficult to analyze them. There is paucity of analytical strategies that can handle such heterogeneous data and facilitate decision-making [46]. The literature mentions some of the analytical approaches which can apply to healthcare and medicine.

As described by Asante-Korang & Jacobs [3] and Groves et al. [47], by incorporating descriptive and comparative analytics, healthcare organizations have seen improved quality of care. However, they state that the long-term tangible benefits can be accrued with utilization of predictive analytics. According to the literature, predictive analytics can be used for prediction of high-cost patients, readmissions, triage, decompensation (when a patient's condition worsens), adverse events, and treatment optimization for diseases affecting multiple organ system [18,27,36,48]. Some of the Big Data Analytical Techniques used in healthcare are shown in Table 3.

Furthermore, Mohammed et al. [39] highlighted some of the applications of Big Data Technologies like MapReduce and Hadoop for healthcare analytics which was supported by other researchers:

- MapReduce has the ability to improve the performance of common signal detection algorithms for pharmacovigilance at approximately linear speedup rates [59].
- Algorithms based on the Hadoop distributed platform can refine protein structure alignments more accurately than existing algorithms [60].
- MapReduce based algorithms can improve the performance of neural signal processing [61].
- Image reconstruction algorithms accelerate the reconstruction process [62].

MapReduce framework has been also been used by Markonis et al. [63] for finding optimal parameters for lung texture classification and to increase the speed of medical image processing.

Peek et al. [64], in their study discussed about some of the Hadoop-based Big Data processing tools like Oozie and Pig which can be used for batch processing; and non-Hadoop processing tools like Storm, Spark, Hive and GraphLab which can be used for streaming data analysis. Regardless of these potential applications, there is a need for analytical tools to offer parallelization, in order to enable the timely processing of data [5].

3.4. Application of big data analytics in healthcare

Big Data analytics has the potential to transform business and clinical models for smart and efficient delivery of care [34]. It enables integration of de-identified health information to allow secondary uses of data [65]. Also, by recognizing patterns and deciphering associations it can facilitate autonomous-decision making [66]. In clinical practice, Big Data analytics can help early detection of disease, accurate prediction of disease trajectory, and identification of deviation from healthy state, changed disease trajectories and detection of fraud. By providing this information, it helps the healthcare organizations in personalization of predictions, targeted-treatment and cost-effectiveness of care, and reduction in waste of resources; and by giving actionable recommendations to individuals it encourages them maintain themselves in good health [10,18,26]. Big Data presents an opportunity to detect relatively low-frequency events that nonetheless can have significant clinical impact. Apart from that, clinical data integration and its effective usage support a vast range of applications, such as disease surveillance, clinical decision support systems, and individual healthcare management; improvement of health-process efficiency; enhancement of healthcare quality and reduction of healthcare cost [8,67]. Sukumar et al. [34] reveals that integrating Big Data analytics into healthcare can provide answer to eight important questions in healthcare − 1. How are costs for various aspects of health care likely to rise in the future? 2. How are certain policy changes impacting cost and behavior? 3. How do health care costs vary geographically? 4. Can fraudulent claims be detected? 5. What treatment options seem most effective for various diseases? 6. Why do some providers seem to have better health outcomes? 7. Why do patients choose one provider over another? 8. Are there early signs of an epidemic?

Table 2
Sources of Healthcare Data (Adapted from [15]).

Type	Description		Source
Clinical	Electronic Medical Records (EMRs)	Detailed patient-related information (physician prescriptions, medications, medical history)	Hospitals & Clinics
	Diagnostic	Diagnostic Results (imaging results, laboratory reports)	Laboratories Radiology Departments
	Biomarkers	Molecular data (genomic, proteomic, transcriptomic, metabolomic)	Diagnostic Companies
	Ancillary	Administrative data (admission, discharge, transfer) & financial data (claims)	Hospitals & Clinics Data Aggregators
Claims	Medical Claims	Medical reimbursement data (procedures, hospital stay, insurance policy details)	Payers Data Aggregators
	Prescription Claims	Prescription reimbursement data (drugs, dose, duration)	Payers Data Aggregators
Clinical Research	Clinical Trails	Design parameters (compound, size, end points)	Pharma Companies Medical Journals
Patient-generated Data	Social Media	Community discussions	Web Health Portals Social Media Websites
	Wearable & Sensors	Wellness & lifestyle data (smartphones, fitness monitors)	Device Data Systems

Table 3
Big Data Analytical Techniques in Healthcare.

Big Data Analytical Technique	Healthcare Application	Studies By
Cluster Analysis	Determination of obesity clusters for identifying high-risk groups;	Clark et. al. [49]
	Determination of population clusters with specific health determinants for treatment of chronic diseases	Swain [50], Schatz [51]
Data Mining	Bio-signal monitoring for health-related abnormalities;	Forkan et. al. [52]
	Determination of epidemics;	Ghani et. al. [36]
	Inductive reasoning and exploratory data analysis in healthcare	Roski et. al. [31]
Graph Analytics	Analysis of hospital performance across various quality measures	Downing et. al. [53]
Machine Learning	Prediction of disease risk;	Chen et. al. [54]
	Assessment of the hospital performance;	Downing et. al. [53]
	Determination of epidemics	Ghani et. al. [36]
Natural Language Processing (NLP)	Improvement of efficiency of care and controlling costs;	Wang et. al. [27]
	Providing training, consultation and treatments;	Khalifa and Meystre [55]
	Identification of high-risk factors;	Martin-Sanchez et. al. [5]
	Extraction of information from clinical notes;	Roski et. al. [31]
	Reducing likelihood of morbidity & mortality	
Neural Networks	Diagnosis of chronic diseases;	Al-Jumeily et. al. [56]
	Prediction of patients' future disease	Martin-Sanchez et. al. [5]
Pattern Recognition	Improvement of public health surveillance	Luxton [57]
Spatial Analysis	Extracting meaningful population-level insights by using visual, spatial and advanced analytics	Amirian et al. [58]

Table 4
Big Data Applications in Healthcare.

Application Area	Studies By:
Genomics	Maia et al. [70]
Drug Discovery & Clinical Research	Szlezák et al. [15]; Taglang & Jackson [46]; Wong et al. [72]
Personalized Healthcare	Viceconti et al. [73]
Precision Medicine	Leff & Yang [9]; Weng & Kahn [35]; Huang et al. [43]
Elderly Care	Jiang et al. [74]
Mental Health	Geerts et al. [75]
Cardiovascular Disease	Asante-Korang & Jacobs [3]; Rumsfeld et al. [11]; Mandawat et al. [76]; Kim [77]
Diabetes	Bellazi et al. [68]; Kumar Sarvana [48]
Gynecology	Erekson & Iglesia [78]
Nephrology	Nandkarni et al. [79]
Oncology	Mandawat et al. [76]; Maia et al. [70]; Naqa [80]
Ophthalmology	Clark et al. [49]
Urology	Ghani et al. [36]

Belle et al. [2] identified three major areas for the application of Big Data analytics in Healthcare: Image Processing, Signal Processing and Genomics. On the other hand, the eight areas of application of Big Data analytics to improve healthcare as per Rumsfeld et al. [11] include: 1) predictive modelling for risk and resource use; 2) population management; 3) drug and medical device safety surveillance; 4) disease and treatment heterogeneity; 5) precision medicine and clinical decision support; 6) quality of care and performance measurement; 7) public health; and 8) research applications. Raghupathi & Raghupathi [10] state that, 'the areas in which advanced analytical techniques yield the greatest results include: pinpointing patients who are the greatest consumers of health resources or at the greatest risk for adverse outcomes; providing individuals with the information they need to make informed decisions and more effectively manage their own health as well as more easily adopt and track healthier behaviors; identifying treatments, programs and processes that do not deliver demonstrable benefits or cost too much; reducing readmissions by identifying environmental or lifestyle factors that increase risk or trigger adverse events and adjusting treatment plans accordingly; improving outcomes by examining vitals from at-home health monitors; managing population health by detecting vulnerabilities within patient populations during disease outbreaks or disasters; and bringing clinical, financial and operational data together to analyze resource utilization productively and in real time'. Electronic phenotyping is another area which can successfully exploit Big Data technologies for ascertaining a clinical condition or characteristic (phenotype) [68,69]. These studies show that there is a vast potential of Big Data analytics in Healthcare. It is beyond the scope of this review to encompass each of these potential applications. Table 4 summarizes some of the usability studies according to their application areas.

Apart from these clinical benefits and applications mentioned above, the literature also presents operational and financial benefits of

Big Data analytics. From among all the articles examined, three articles [6,27,70] unveil the business value in healthcare. Findings from the study of Wang & Hajli [6], exhibits that benefits of Big Data analytics are improved IT effectiveness and efficiency, and optimization of clinical operations.

3.5. Challenges in big data analytics in healthcare and strategies to overcome them

Even with huge potential benefits, the healthcare industry is in its nascent stage for adoption of Big Data analytics. With the huge amount of data available, there is a lack of knowledge about which data to use and for what purpose [15]. Another major challenge that healthcare faces is the lack of appropriate IT infrastructure [8,27,81] and transition from use of paper-based records to use of distributed data processing [64,82]. The resistance for redesigning processes and approving technology that influences the health care system [39,70] and need for huge initial investment [15,83], makes it more difficult to utilize Big Data technology. Studies show that because of the lack of knowledge about the best algorithm and tool for analysis [71] and unavailability of trained clinical scientists and Big Data managers for interpretation of Big Data outcomes [3,26,81,83], healthcare remains far from realizing the potential of Big Data analytics. A major concern with the use of Big Data analytics in Healthcare is the processing of information without human supervision which might lead to erroneous conclusions [84,85]. According to Raghupathi and Raghupathi [10], there is a need for a simple, convenient and transparent Big Data analytics system which can be applied for real-time cases.

From the technical point of view, challenges include integration of structured, semi-structured and unstructured data from a variety of resources [26]. Studies show that the main technical issues in Big Data analytics include siloed/fragmented data [10,15,45], limitations of observational data [11,43], validation [11], data structure issues, data standardization issues [10,32,43,75,86], data inaccuracy and inconsistency (veracity) [15,20,32,34], data reliability [87], semantic interoperability [25,64,87], network bandwidth, scalability, and cost [2]. The problems such as missing data and the risk of false-positive associations [11,88] also add to it. Security issues such as Big Data breaches can be significant threat in healthcare [3,35].

Patient privacy and confidentiality are of utmost importance in healthcare. But data sharing between various stakeholders for deriving insights, can deepen the concern for privacy [8,39,43,70,83,85]. According to Mittelstadt et al. [83] informed consent and privacy are the key areas of concern. Lack of data protocols and standards are some of the governance issues faced by Big Data analytics in healthcare [2]. One of the studies by Lee and Yoon [20] states that one of the prominent reasons for the lack of clinical integration of Big Data technology is the dearth of evidence of practical benefits of Big Data analytics in healthcare.

In order to overcome these challenges, various strategies were found in the literature. The strategies for curbing the aforementioned issues include:

- *Implementing (big) data governance*: Due to poor governance, healthcare organizations incur huge financial costs in IT investment [23]. With appropriate data governance, the enterprise-wide data resources can be leveraged effectively to create business value [27,83].
- *Developing an information sharing culture*: Information sharing and aggregation of data can address the issue of interoperability and enable effective utilization of the Big Data analytical and predictive capabilities [27,45].
- *Employing security measures*: Strong encryption of data, validation of source of data, access control and authentication [89] and de-identification [90] are some of the measures for securing the data and maintaining confidentiality.
- *Training key personnel to use Big Data analytics*: In order to extract meaningful insights and valuable information from Big Data, healthcare professionals should be trained with Big Data analytics competencies. This is critical for healthcare, because incorrect interpretation of the reports generated could lead to unanticipated consequences [27].
- *Incorporating cloud computing into the organization's Big Data analytics*: The challenge of storage of voluminous data can be tackled by making use of cloud computing. This would enable small and medium sized hospitals and care organizations to eliminate cost and data storage issues [27].

According to Wang et al. [27], a shift of focus from technology tools to the managerial, economic, and strategic impacts of Big Data analytics and exploration of effective path for acquiring healthcare business value would enable exploit the benefits of Big Data analytics.

4. Discussion

4.1. Main findings

This systematic review assessed the emerging landscape of Big Data analytics for healthcare. Specifically, it identified the best available literature about the concept of Big Data analytics, sources of Big Health Data, Big Data analytical techniques for clinical data, implementation of Big Data analytics in healthcare, causes for underutilization of Big Data analytics in healthcare and strategies to mitigate them.

The concept of Big Data covers a wide range of definitions extending from the data that is difficult to manage using traditional analytical tools, to the characterization of big data as having large volume, high velocity, huge variety and varied veracity. While most of the studies define Big Data in terms of the aforementioned characteristics, one of the studies state that Big Data in healthcare is also characterized by having energy and life-span [25], but the literature lacks detailed description of these characteristics especially with regards to healthcare. Studies have demonstrated that Big Data analytics differ from traditional analytical approach in terms that instead of tracking care quality and outcomes in retrospective view by using deductive reasoning, it uses inductive reasoning for prospective analysis of data [91]. These techniques of data analysis are hypotheses-generating rather than hypotheses-testing since they focus on finding association and correlation in the observational data and not on the casual relationship between variables. But it needs for the testing of hypotheses before applying results into a clinical practice [92]. Few of the studies state that there is a need for the application of human judgement and supervision on the insights obtained from application of Big Data analytics in healthcare [84,93], as it would prevent the occurrence of adverse events which result from relying solely on Big Data analytics. Big Data analytics can, thus, lead to a shift from experience-based to evidence-based decision-making in healthcare [10,94].

Most studies have shown that there are a large number of sources of healthcare Big Data. Clinical and administrative data in healthcare comes from various sources including healthcare providers, laboratories, diagnostic companies, insurance companies, pharmaceutical firms, not-for-profit organizations, government and web-health portals [38–44]. The literature on Big Data techniques and technologies that are applied to clinical data is largely fragmented. Most of this literature highlights the use of natural language processing for clinical as well as operational applications. As identified from the review, other Big Data techniques which find application in healthcare are cluster analysis, data mining, graph analytics, machine learning, neural networks, pattern recognition and spatial analysis. Studies showed that in most of the cases, Hadoop and tools which run on top of Hadoop are used for processing of patient-data [59–64], but since they are batch-processing tools, newer tools like Storm, Spark and GraphLab have started finding their application for streaming and real-time data [64].

Majority of the studies reviewed, concerned the application of Big Data analytics in different areas of healthcare. According to them, Big Data analytics finds application in clinical decision support; personalized medicine; and optimization of clinical operations and cost-effectiveness of care. Thus, the integration of Big Data technology into healthcare can not only improve the quality of care, but enable early identification of high-risk patients by making use of real-time analytics and hence can benefit by saving lives. Studies on the use of Big Data analytics for cardiovascular diseases, diabetes, oncology, elderly care, gynecology, and clinical research have shown that it can enable delivery of timely care and cost-saving by eliminating inefficiencies [70–80].

Despite the tremendous value-addition with use of Big Data analytical tools, healthcare industry still lags in adoption of this technology due to many challenges. The non-availability of appropriate IT infrastructure, huge investment costs associated with implementing analytical tools, data privacy & security issues, fragmented data ownership and technical challenges such as data quality & multi-dimensionality of data are some of the issues. One of the studies identified lack of evidence of practical benefits as a major cause behind the reluctance for using Big Data analytics in healthcare [20]. Few studies highlighted that there is a shortage of skilled Big Data analysts with knowledge of healthcare, who have the ability to identify right data and right tools for analysis of health-related data and interpret insights obtained after analysis, which makes the use of technology difficult. The complexity of Big Data analytical systems, is also one of the factors for limited use of technology for healthcare applications [10]. Strategies for mitigating these challenges are hence required in order to realize its full potential. Some of these strategies include change in organizational culture; health-information exchange; training of key healthcare personnel; development of simple-transparent Big Data systems; use of cloud storage and distributed data processing; and strengthening of IT security.

4.2. Gaps and implications for future research

- Most of the studies reviewed, have a relatively narrow scope with limited practical application. The published work discussed about the potential application of Big Data analytics but there was no evidence of real-world cases about the application in healthcare.
- None of the usability studies on Big Data analytics included in this review discussed about quantitative results by the usage of technology. Most of these studies used qualitative approach to explain the benefits and challenges of using Big Data technology for healthcare applications, while the use of quantitative approach will provide evidence for the practical benefits and will enable wide-scale adoption of technology.
- Majority of the studies included in review, were from the developed countries. It is essential to promote the research on Big Data analytics in healthcare in the developing countries, since that will enable delivery of better quality care.

4.3. Limitations

While the literature covers information about Big Data analytics and its role in healthcare and medicine, current research has few limitations. First, the contents of this study consists of a systematic review of the current status of Big Data technology in healthcare, but it does not take into consideration the technical details regarding the implementation and results obtained in each of the study reviewed. Second, there is a heterogeneity in documentation since the literature contains disparate sources of information on definition of Big Data, techniques of Big Data analytics and their application and challenges in healthcare. Finally, despite the use of a systematic approach for review, the inclusion of studies on 'big data analytics' in 'healthcare' for this review was based on subjective judgement, hence the cross-reference articles were also considered for this review.

5. Conclusion

Big Data analytics has emerged as a new frontier for enhancing healthcare delivery. With the opportunities created by digital and information revolution, healthcare industry can exploit the potential benefits of leveraging Big Data technology. Big Data analytics increasingly provides value to healthcare by improving healthcare quality and outcomes and providing cost-effective care. The predictive nature and pattern-recognition aspect of Big Data analytics enable the shift from experience-based medicine to evidence-based medicine. Through its systematic review, the study presents a useful starting point for the application of Big Data analytics in future healthcare research. In addition, the study reflects that once the scope of Big Data analytics is defined; its characteristics and features are understood; and challenges are properly tackled, its application will maximize the healthcare value through promoting the extensive usage of insights.

Author contributions

Study conception and design: Nishita Mehta
Acquisition of data: Nishita Mehta, Dr. Anil Pandit
Analysis and interpretation of data: Nishita Mehta, Dr. Anil Pandit
Drafting of manuscript: Nishita Mehta, Dr. Anil Pandit
Critical revision: Nishita Mehta, Dr. Anil Pandit

Conflict of interest

None.

Summary Points

What was already known on the topic:

- With digital evolution in healthcare, there is a huge amount of health data available
- which remains unexploited and hence is wasted.
- Big Data analytics (BDA) has the ability to process health data and generate valuable insights which can help improve the clinical, operational and financial outcomes.
- Despite the potential applications of BDA, healthcare industry is slow in leveraging Big Data initiatives.

What this study has added:

- Technology readiness for BDA is very low in healthcare majorly because of the non availability of skilled and trained clinical Big Data analysts.
- There is minimal evidence on how BDA can improve the quality of care, nor are there any economic assessment studies regarding its cost-effectiveness.
- There is a need for a user-friendly and transparent Big Data system for wider adoption of BDA in healthcare.
- Shift in organizational culture and training of key healthcare personnel, will enable effective utilization of BDA for better healthcare outcomes

References

[1] T.B. Murdoch, A.S. Detsky, The Inevitable Application of Big Data to Health Care 309 (2016), pp. 5–6.
[2] A. Belle, R. Thiagarajan, S.M.R. Soroushmehr, F. Navidi, D.A. Beard, K. Najarian, Big data analytics in healthcare, Hindawi Publ Corp. 2015 (2015) 1–16, http://dx.doi.org/10.1155/2015/370194.
[3] A. Asante-Korang, J.P. Jacobs, Big Data and paediatric cardiovascular disease in the

era of transparency in healthcare, Cardiol. Young 26 (2016) 1597–1602, http://dx.doi.org/10.1017/S1047951116001736.
[4] Z. Goli-Malekabadi, M. Sargolzaei-Javan, M.K. Akbari, An effective model for store and retrieve big health data in cloud computing, Comput. Methods Programs Biomed. 132 (2016) 75–82, http://dx.doi.org/10.1016/j.cmpb.2016.04.016.
[5] F.J. Martin-Sanchez, V. Aguiar-Pulido, G.H. Lopez-Campos, N. Peek, L. Sacchi, Secondary use and analysis of big data collected for patient care, IMIA Yearbook 26 (2017) 1–10, http://dx.doi.org/10.15265/IY-2017-008.
[6] Y. Wang, N. Hajli, Exploring the path to big data analytics success in healthcare, J. Bus. Res. 70 (2017) 287–299, http://dx.doi.org/10.1016/j.jbusres.2016.08.002.
[7] B. Cyganek, M. Graña, B. Krawczyk, A. Kasprzak, P. Porwik, K. Walkowiak, M. Woźniak, A survey of big data issues in electronic health record analysis, Appl. Artif. Intell. 30 (2016) 497–520, http://dx.doi.org/10.1080/08839514.2016.1193714.
[8] F.F. Costa, Big data in biomedicine, Drug Discov. Today 19 (2014) 433–440, http://dx.doi.org/10.1016/j.drudis.2013.10.012.
[9] D.R. Leff, G.-Z. Yang, Big data for precision medicine, Engineering 1 (2015) 277–279, http://dx.doi.org/10.15302/J-ENG-2015075.
[10] W. Raghupathi, V. Raghupathi, Big data analytics in healthcare: promise and potential, Heal. Inf. Sci. Syst. 2 (2014) 3, http://dx.doi.org/10.1186/2047-2501-2-3.
[11] J.S. Rumsfeld, K.E. Joynt, T.M. Maddox, Big data analytics to improve cardiovascular care: promise and challenges, Nat. Rev. Cardiol. 13 (2016) 350–359, http://dx.doi.org/10.1038/nrcardio.2016.42.
[12] J. Manyika, M. Chui, B. Brown, et al., Big Data:The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, 2011, 2018.
[13] L. Haar, Big Data Expected to Have Big Impact on Diagnostic Imaging, (2014) http://www.diagnosticimaging.com/siim-2014/big-data-expected-have-big-impact-diagnostic-imaging.
[14] M. Grindle, J. Kavathekar, D. Wan, A new era for the healthcare industry–cloud computing changes the game, Accenture (2013).
[15] N. Szlezák, M. Evers, J. Wang, L. Pérez, The role of big data and advanced analytics in drug discovery, development, and commercialization, Clin. Pharmacol. Ther. 95 (2014) 492–495, http://dx.doi.org/10.1038/clpt.2014.29.
[16] W.D. Kirsh, Why Longitudinal Data Is Crucial to Making Better Care Decisions, (2017) https://www.healthdatamanagement.com/opinion/why-longitudinal-data-is-crucial-to-making-better-care-decisions.
[17] E. Capobianco, Systems and precision medicine approaches to diabetes heterogeneity: a Big Data perspective, Clin. Transl. Med. 6 (2017) 23, http://dx.doi.org/10.1186/s40169-017-0155-4.
[18] D.W. Bates, S. Saria, L. Ohno-Machado, A. Shah, G. Escobar, Big data in health care: using analytics to identify and manage high-risk and high-cost patients, Health Aff. 33 (2014) 1123–1131, http://dx.doi.org/10.1377/hlthaff.2014.0041.
[19] H. Krumholz, Big data and new knowledge In medicine: the thinking training, and tools needed for a learning health system, Health Aff. 33 (2014) 1163–1170.
[20] M. Cox, D. Ellsworth, Application-controlled demand paging for out-of-core visualization, Proceedings. Vis. '97 (Cat. No. 97CB36155) (1997) 235–244, http://dx.doi.org/10.1109/VISUAL.1997.663888.
[21] F.X. Diebold, Big data dynamic factor models for macroeconomic measuring and forecasting, Adv. Econ. Econom. Eighth World Congr. Econom. Soc. (2003) 115–122.
[22] D. Laney, META delta, Appl. Deliv. Strateg. 949 (2001) 4, http://dx.doi.org/10.1016/j.infsof.2008.09.005.
[23] B. Feldman, E.M. Martin, T. Skotnes, Big data in healthcare–hype and hope, Dr. Bonnie 360 degree, Bus. Dev. Digit. Heal. 2012 (2013) 122–125 http://www.riss.kr/link?id=A99883549.
[24] S.B. Scruggs, Karol Watson, Andrew I. Su, Henning Hermjakob, J.R. Yates 3rd, M.L. Lindsey, P. Ping, Harnessing the heart of big data, Circ. Res. 116 (2015) 1115–1119.
[25] I.D. Dinov, Volume and value of big healthcare data, J. Med. Stat. Inf. 4 (2016), http://dx.doi.org/10.7243/2053-7662-4-3.
[26] C. Auffray, R. Balling, I. Barroso, L. Bencze, M. Benson, J. Bergeron, E. Bernal-Delgado, N. Blomberg, C. Bock, A. Conesa, S. Del Signore, C. Delogne, P. Devilee, A. Di Meglio, M. Eijkemans, P. Flicek, N. Graf, V. Grimm, H.-J. Guchelaar, Y.-K. Guo, I.G. Gut, A. Hanbury, S. Hanif, R.-D. Hilgers, Honrado Á, D.R. Hose, J. Houwing-Duistermaat, T. Hubbard, S.H. Janacek, H. Karanikas, T. Kievits, M. Kohler, A. Kremer, J. Lanfear, T. Lengauer, E. Maes, T. Meert, W. Müller, D. Nickel, P. Oledzki, B. Pedersen, M. Petkovic, K. Pliakos, M. Rattray, J.R. i Màs, R. Schneider, T. Sengstag, X. Serra-Picamal, W. Spek, L.A.I. Vaas, O. van Batenburg, M. Vandelaer, P. Varnai, P. Villoslada, J.A. Vizcaíno, J.P.M. Wubbe, G. Zanetti, Making sense of big data in health research: towards an EU action plan, Genome Med. 8 (2016) 71, http://dx.doi.org/10.1186/s13073-016-0323-y.
[27] Y. Wang, L.A. Kung, T.A. Byrd, Big data analytics: understanding its capabilities and potential benefits for healthcare organizations, Technol. Forecast. Soc. Change (2015), http://dx.doi.org/10.1016/j.techfore.2015.12.019.
[28] K.Y. He, D. Ge, M.M. He, Big data analytics for genomic medicine, Int. J. Mol. Sci. 18 (2017) 1–18, http://dx.doi.org/10.3390/ijms18020412.
[29] S.S.-L. Tan, G. Gao, S. Koch, Big data and analytics in healthcare, Methods Inf. Med. 54 (2015) 546–547, http://dx.doi.org/10.3414/ME15-06-1001.
[30] H. Liyanage, S. De Lusignan, S. Liaw, C. Kuziemsky, F. Mold, P. Krause, D. Fleming, Big data usage patterns in the health care domain: a use case driven approach applied to the assessment of vaccination benefits and risks contribution of the IMIA primary healthcare working group big data for assessing vaccination benefits and risks: a, IMIA, Yearbook Med. Inf. (2014) 27–35.
[31] J. Roski, G.W. Bo-Linn, T.A. Andrews, Creating value in health care through big data: opportunities and policy implications, Health Aff. 33 (2014) 1115–1122, http://dx.doi.org/10.1377/hlthaff.2014.0147.
[32] Jiang Bian, David Maslov, Mayer Angel Mayer, Siven Seevanayanagam, L. Toldo, Challenges and opportunities of big data in health care: a systematic review, J. Med. Internet Res. 4 (2016), http://dx.doi.org/10.2196/medinform.5359.
[33] Hao Gui, Rong Zheng, C. Ma, An architecture for healthcare big data management and analysis, Nternational Conf. Heal. Inf Sci. (2016) 154–160.
[34] R. Sreenivas Sukumar, Ramachandran Natarajan, R. Ferrell, Quality of Big Data in health care, Int. J. Health Care Qual. Assur. 28 (2015) 621–634.
[35] C. Weng, M.G. Kahn, Clinical research informatics for big data and precision medicine, Yearbook Med. Inf. (2016) 211–218, http://dx.doi.org/10.15265/IY-2016-019.
[36] K.R. Ghani, K. Zheng, J.T. Wei, C.P. Friedman, Harnessing big data for health care and research: are urologists ready? Eur. Urol. 66 (2014) 975–977, http://dx.doi.org/10.1016/j.eururo.2014.07.032.
[37] Emilie Baro, Samuel Degoul, Regis Beuscart, E. Chazard, Toward a literature-driven definition of big data in healthcare, BioMed. Res. Int. (2015).
[38] N. Rouse, B. William, Serban Understanding and Managing the Complexity of Healthcare, (2014).
[39] E.A. Mohammed, B.H. Far, C. Naugler, Applications of the MapReduce programming framework to clinical big data analysis: current landscape and future trends, BioData Min. 7 (2014) 22, http://dx.doi.org/10.1186/1756-0381-7-22.
[40] K. Miller, Big data analytics in biomedical research, Biomed. Comput. Rev. (2012) 15–21.
[41] M. Swan, The quantified self: fundamental disruption in big data science and biological discovery, Big Data 1 (2013).
[42] Michael J. Ward, Keith A. Marsolo, C.M. Froehle, Applications of business analytics in healthcare, Bus. Horiz. 57 (2014) 571–582.
[43] B.E. Huang, W. Mulyasasmita, G. Rajagopal, The path from big data to precision medicine, Expert Rev. Precis. Med. Drug Dev. 1 (2016) 129–143, http://dx.doi.org/10.1080/23808993.2016.1157686.
[44] P.S. Bradley, Implications of big data analytics on population health management, Big Data 1 (2013) 152–159.
[45] D.V. Dimitrov, Medical internet of things and big data in healthcare, Healthcare Inf. Res. 22 (2016) 156–163, http://dx.doi.org/10.4258/hir.2016.22.3.156.
[46] G. Taglang, D.B. Jackson, Use of big data in drug discovery and clinical trials, Gynecol. Oncol. 141 (2016) 17–23, http://dx.doi.org/10.1016/j.ygyno.2016.02.022.
[47] S. Groves, Basel Van Kayyali, David Knott, P. Kuiken, The Big Data Revolution in Healthcare: Accelerating Value and Innovation, (2013).
[48] N.M. Saravana Kumar, T. Eswari, P. Sampath, S. Lavanya, Predictive methodology for diabetic data analysis in big data, Procedia Comput. Sci. 50 (2015) 203–208, http://dx.doi.org/10.1016/j.procs.2015.04.069.
[49] A. Clark, J.Q. Ng, N. Morlet, J.B. Semmens, Big data and ophthalmic research, Surv. Ophthalmol. 61 (2016) 443–465, http://dx.doi.org/10.1016/j.survophthal.2016.01.003.
[50] A. Swain, Mining big data to support decision making in healthcare, J. Inf. Technol. Case Appl. Res. 18 (2016) 141–154, http://dx.doi.org/10.1080/15228053.2016.1245522.
[51] B. Schatz, National surveys of population health: big data analytics for mobile health monitors, Big Data 3 (2015) 219–229, http://dx.doi.org/10.1089/big.2015.0021.
[52] A. Forkan, I. Khalil, M. Atiquzzaman, ViSiBiD, A learning model for early discovery and real-time prediction of severe clinical events using vital signs as big data, Comput. Networks 113 (2017) 244–257, http://dx.doi.org/10.1016/j.comnet.2016.12.019.
[53] N. Downing, A. Cloninger, A. Venkatesh, A. Hsieh, E. Drye, R. Coifman, et al., Describing the performance of U. S. hospitals by applying big data analytics, PLoS One 12 (2017) e0179603, http://dx.doi.org/10.1371/journal.pone.0179603.
[54] M. Chen, Y. Hao, K. Hwang, L. Wang, L. Wang, Disease prediction by machine learning over big data from healthcare communities, IEEE Access. 5 (2017) 8869–8879, http://dx.doi.org/10.1109/access.2017.2694446.
[55] A. Khalifa, S. Meystre, Adapting existing natural language processing resources for cardiovascular risk factors identification in clinical notes, J. Biomed. Inf. 58 (2015) S128–S132, http://dx.doi.org/10.1016/j.jbi.2015.08.002.
[56] D. Al-jumeily, A. Hussain, C. Mallucci, C. Oliver, Applied Computing in Medicine and Health, (2015).
[57] D. Luxton, Artificial Intelligence in Behavioral and Mental Health Care, (2015).
[58] P. Amirian, F. van Loggerenberg, T. Lang, A. Thomas, R. Peeling, A. Basiri, S.N. Goodman, Using big data analytics to extract disease surveillance information from point of care diagnostic machines, Pervasive Mob. Comput. (2017), http://dx.doi.org/10.1016/j.pmcj.2017.06.013.
[59] W. Wang, K. Haerian, H. Salmasian, R. Harpaz, H. Chase, C. Friedman, A drug-adverse event extraction algorithm to support pharmacovigilance knowledge mining from PubMed citations, AMIA Annu. Symp. Proc. 1464-70 (2011) (22195210).
[60] C.-L. Hung, Y.-L. Lin, Implementation of a parallel protein structure alignment service on cloud, Int. J. Genomics. 2013 (2013) 1–8, http://dx.doi.org/10.1155/2013/439681.
[61] L. Wang, D. Chen, R. Ranjan, S.U. Khan, J. Kołodziej, J. Wang, Parallel processing of massive EEG data with MapReduce, Proc. Int. Conf. Parallel Distrib. Syst.–ICPADS. (2012) 164–171, http://dx.doi.org/10.1109/ICPADS.2012.32.
[62] B. Meng, G. Pratx, L. Xing, Ultrafast and scalable cone-beam CT reconstruction using MapReduce in a cloud computing environment, Med. Phys. 38 (2011) 6603–6609, http://dx.doi.org/10.1118/1.3660200.
[63] D. Markonis, R. Schaer, I. Eggel, H. Muller, A. Depeursinge, Using MapReduce for large-Scale medical image analysis, 2012 IEEE Second Int. Conf. Healthcare Informatics, Imaging Syst. Biol. (2012) 1, http://dx.doi.org/10.1109/HISB.2012.8.

[64] N. Peek, J. Holmes, J. Sun, Technical challenges for big data in biomedicine and health: data sources, infrastructure, and analytics, IMIA Yearbook 9 (2014) 42–47, http://dx.doi.org/10.15265/iy-2014-0018.

[65] I. Cano, A. Tenyi, E. Vela, F. Miralles, J. Roca, Perspectives on Big Data applications of health information, Curr. Opin. Syst. Biol. 3 (2017) 36–42, http://dx.doi.org/10.1016/j.coisb.2017.04.012.

[66] A. Asante-Korang, J.P. Jacobs, Big Data and paediatric cardiovascular disease in the era of transparency in healthcare, Cardiol. Young 26 (2016) 1597–1602, http://dx.doi.org/10.1017/S1047951116001736.

[67] T. Schultz, Turning healthcare challenges into big data opportunities: a use-case review across the pharmaceutical development lifecycle, Bull. Assoc. Inf. Sci. Technol. (2013).

[68] Riccardo Bellazzi, Arianna Dagliati, Lucia Sacchi, D. Segangi, Big data technologies: new opportunities for diabetes management, J. Diabetes Sci. Technol. 9 (2015) 1119–1125.

[69] L.J. Frey, L. Lenert, G. Lopez-Campos, EHR big data deep phenotyping. Contribution of the IMIA Genomic Medicine Working Group, Yearbook Med. Inf. 9 (2014) 206–211, http://dx.doi.org/10.15265/IY-2014-0006.

[70] Y. Wang, L. Kung, W.Y.C. Wang, C.G. Cegielski, An integrated big data analytics-enabled transformation model: application to health care, Inf Manag. (2017) 1–16, http://dx.doi.org/10.1016/j.im.2017.04.001.

[71] A.T. Maia, S.-J. Sammut, A. Jacinta-Fernandes, S.-F. Chin, Big data in cancer genomics, Curr. Opin. Syst. Biol. 4 (2017) 78–84, http://dx.doi.org/10.1016/j.coisb.2017.07.007.

[72] H.T. Wong, Q. Yin, Y.Q. Guo, K. Murray, D.H. Zhou, D. Slade, Big data as a new approach in emergency medicine research, J. Acute Dis. 4 (2015) 178–179, http://dx.doi.org/10.1016/j.joad.2015.04.003.

[73] M. Viceconti, P. Hunter, D. Hose, Big Data, big knowledge: big data for personalised healthcare, IEEE J. Biomed. Heal. Inf. 2194 (2015) 1, http://dx.doi.org/10.1109/JBHI.2015.2406883.

[74] P. Jiang, J. Winkley, C. Zhao, R. Munnoch, G. Min, L.T. Yang, An intelligent information forwarder for healthcare big data systems with distributed wearable sensors, IEEE Syst. J. (2014) 1–9, http://dx.doi.org/10.1109/JSYST.2014.2308324.

[75] H. Geerts, P.A. Dacks, V. Devanarayan, M. Haas, Z.S. Khachaturian, M.F. Gordon, S. Maudsley, K. Romero, D. Stephenson, Big data to smart data in Alzheimer's disease: the brain health modeling initiative to foster actionable knowledge, Alzheimer's Dement 12 (2016) 1014–1021, http://dx.doi.org/10.1016/j.jalz.2016.04.008.

[76] A. Mandawat, A.E. Williams, S.A. Francis, Cardio-oncology: the role of big data, Heart Fail. Clin. 13 (2017) 403–408, http://dx.doi.org/10.1016/j.hfc.2016.12.010.

[77] J. Kim, Big Data, Health informatics, and the future of cardiovascular medicine, J. Am. Coll. Cardiol. 69 (2017) 899–902, http://dx.doi.org/10.1016/j.jacc.2017.01.006.

[78] E.A. Erekson, C.B. Iglesia, Improving patient outcomes in gynecology: the role of large data registries and big data analytics, J. Min. Invasive Gynecol. 22 (2015) 1124–1129, http://dx.doi.org/10.1016/j.jmig.2015.07.003.

[79] G.N. Nadkarni, S.G. Coca, C.M. Wyatt, Big data in nephrology: promises and pitfalls, Kidney Int. 90 (2016) 240–241, http://dx.doi.org/10.1016/j.kint.2016.06.003.

[80] I. El Naqa, Perspectives on making big data analytics work for oncology, Methods 111 (2016) 32–44, http://dx.doi.org/10.1016/j.ymeth.2016.08.010.

[81] S. Fodeh, Q. Zeng, Mining Big Data in biomedicine and health care, J. Biomed. Inf. 63 (2016) 400–403, http://dx.doi.org/10.1016/j.jbi.2016.09.014.

[82] F. Zhang, J. Cao, S. Khan, K. Li, K. Hwang, A task-level adaptive MapReduce framework for real-time streaming data in healthcare applications, Future Gener. Comput. Syst. 43-44 (2015) 149–160, http://dx.doi.org/10.1016/j.future.2014.06.009.

[83] J. Wu, H. Li, S. Cheng, Z. Lin, The promising future of healthcare services: when big data analytics meets wearable technology, Inf. Manag. 53 (2016) 1020–1033, http://dx.doi.org/10.1016/j.im.2016.07.003.

[84] G. Asokan, V. Asokan, Leveraging big data to enhance the effectiveness of one health in an era of health informatics, J. Epidemiol. Global Health 5 (2015) 311–314.

[85] M. Grossglauser, H. Saner, Data-driven healthcare: from patterns to actions, Eur. J. Prev. Cardiol. 21 (2014) 14–17, http://dx.doi.org/10.1177/2047487314552755.

[86] R. Budhiraja, R. Thomas, M. Kim, S. Redline, The role of big data in the management of sleep-disordered Breathing, Sleep Med. Clin. 11 (2016) 241–255, http://dx.doi.org/10.1016/j.jsmc.2016.01.009.

[87] S. Salas-Vega, A. Haimann, E. Mossialos, Big data and health care: challenges and opportunities for coordinated policy development in the EU, Health Syst. Reform 1 (2015) 285–300, http://dx.doi.org/10.1080/23288604.2015.1091538.

[88] T. McNutt, K. Moore, H. Quon, Needs and challenges for big data in radiation oncology, Int. J. Radiation Oncol., Biol., Phys. 95 (2016) 909–915, http://dx.doi.org/10.1016/j.ijrobp.2015.11.032.

[89] R. Naydenov, D. Liveri, L. Dupre, E. Chalvatzi, C. Skouloudi, Big Data Security: Good Practices and Recommendations on the Security of Big Data Systems, European Union Agency for Network and Information Security (ENISA), Greece, 2015https://www.enisa.europa.eu/publications/big-data-security/at_download/fullReport.

[90] J. Adler-Milstein, A. Jha, Healthcare's big data challenge, Am. J. Manage. Care 19 (2013) 537–538.

[91] C.H. Lee, H.-J. Yoon, Medical big data: promise and challenges, Kidney Res Clin. Pract. 36 (2017) 3–11, http://dx.doi.org/10.23876/j.krcp.2017.36.1.3.

[92] H. Care, B. Outcomes, Leveraging Big Data and Analytics in Healthcare and Life Sciences: Enabling Personalized Medicine for High-Quality Care, Better Outcomes Developing Technologies to Meet Big Data Challenges, (2012).

[93] Ernst, Young, Is the next evolution of big data big judgment? (2015) http://www.ey.com/gl/en/issues/business-environment/ey-analytics (accessed August 14, 2017).

[94] H.J. Watson, Tutorial: big data analytics: concepts, technologies, and applications, Commun. Assoc. Inf. Syst. 34 (1) (2014) 1247–1268.