Big data analytics in healthcare: a systematic literature review

Sayantan Khanra [a,b], Amandeep Dhir [c,d,e], Najmul Islam [f] and Matti Mäntymäki [a]

[a]Turku School of Economics, University of Turku, Turku, Finland; [b]School of Business, Woxsen University, Hyderabad, India; [c]School of Business and Management, LUT University, Lappeenranta, Finland; [d]Department of Management, School of Business & Law, University of Agder, Kristiansand, Norway; [e]Optentia Research Focus Area, North-West University, Vanderbijlpark, South Africa; [f]Department of Future Technologies, University of Turku, Turku, Finland

ABSTRACT

The current study performs a systematic literature review (SLR) to synthesise prior research on the applicability of big data analytics (BDA) in healthcare. The SLR examines the outcomes of 41 studies, and presents them in a comprehensive framework. The findings from this study suggest that applications of BDA in healthcare can be observed from five perspectives, namely, health awareness among the general public, interactions among stakeholders in the healthcare ecosystem, hospital management practices, treatment of specific medical conditions, and technology in healthcare service delivery. This SLR recommends actionable future research agendas for scholars and valuable implications for theory and practice.

ARTICLE HISTORY

Received 3 March 2020
Accepted 15 August 2020

KEYWORDS

Big data; analytics; big data analytics; healthcare; hospital management; medical informatics; personalised patient care; systematic literature review; Covid-19

1. Introduction

Healthcare enterprises search for suitable technologies to streamline resources for the sake of improving the patient experience and organisational performance (Tang et al. 2019; Wang, Kung, and Byrd 2018; Tandon et al. 2020). Healthcare can be conceptualised as a system comprising three constituent parts: (a) core providers of medical care services, such as physicians, nurses, technicians, and hospital administrations (Boudhir, Ben Ahmed, and Soumaya 2017; Zhang, Simon, and Yu 2017); (b) critical services that are associated with medical care services, such as medical research and health insurance (Austin and Kusumoto 2016; Chandola, Sukumar, and Schryver 2013); and (c) beneficiaries of medical care services, i.e., patients and the public (Salomi and Balamurugan 2016; Weng and Kahn 2016). This study considers that a healthcare system includes contact-based and technology-based remote monitoring services extended by constituent service providers to promote, maintain, or restore the health of beneficiaries (George, Chacko, and Kurien 2019; Kaur, Sharma, and Mittal 2018). Big data analytics (BDA) has had a considerable influence across healthcare functions (Gu et al. 2017; Sáez and García-Gómez 2018), including clinical decision support, disease surveillance, and health management, among others (Raghupathi and Raghupathi 2014).

to guide future research on the topic (Dhir et al. 2020; Talwar et al. 2020). This SLR aims to address four research questions (**RQs**) as follows: **RQ1**. What is the current status of research on the application of BDA in healthcare? **RQ2**. In which contexts within healthcare are the applications of BDA being studied? **RQ3**. What are the key takeaways from prior research on BDA in healthcare? **RQ4**. What future agendas may advance research on BDA in healthcare?

The present SLR adopted a protocol for planning, performing, and presenting a review, as followed in prior studies (Behera, Bala, and Dhir 2019; Tandon et al. 2020). The current status of research on any topic under investigation can be represented in terms of key information, such as the annual distribution of publications, average citation count, top contributors, and methodologies adopted in the reviewed studies (Behera, Bala, and Dhir 2019; Tandon et al. 2020). Therefore, the current SLR answered **RQ1** by reporting a sample profile of the reviewed studies. A total of 39 out of 41 studies reviewed herein were published between 2015 and 2019, indicating that the research topic is relevant to the recent literature on BDA and healthcare. **RQ2** is answered by analysing the contexts of the reviewed studies. The synthesis of the findings of prior studies addresses **RQ3**. Furthermore, a comprehensive framework for the use of BDA in healthcare is developed based on the findings of the present study. Actionable future research agendas, the object of **RQ4**, emerged from the insights offered by the comprehensive framework. Two major contributions of the current study are as follows: (a) synthesis of the literature on BDA in healthcare, and (b) guidance for future researchers interested in the topic by providing them with a framework on the application of BDA in healthcare.

The rest of the paper is organised as follows. The next section presents a brief overview of the characteristics of big data, particularly in the context of healthcare. The third section outlines the methodology followed in this SLR. This is followed by a section on this study's findings. The fifth section discusses the outcomes and implications of this study. The sixth section is dedicated to acknowledging the limitations of the present study, suggesting future scopes of research, and presenting the concluding remarks of this SLR.

2. Background

2.1. Characteristics of big data

The concept of BDA overarches several data-intensive approaches to the analysis and synthesis of large-scale data (Galetsi, Katsaliaki, and Kumar 2020; Mergel, Rethemeyer, and Isett 2016). Such large-scale data derived from information exchange among different systems is often termed 'big data' (Bahri et al. 2018; Khanra, Dhir, and Mäntymäki 2020). Although it is referred to as 'big' data, its importance is associated with its ability to capture small details about the subject being studied (George, Haas, and Pentland 2014; McAfee et al. 2012). Kitchin (2014) summarised the characteristics of big data with seven 'V's as follows:

A) Volume (size). A large amount of data is a primary characteristic of big data.

B) Variety (complexity). Big data includes structured, semi-structured, and unstructured data in different formats, such as text, image, audio, video, and sensor data, among others.

c) Velocity (speed). Big data handles high rates of data inflow and processes the data in real-time.

d) Veracity (quality). Big data accumulates detailed data that is exhaustive in scope.

e) Value (knowledge). Big data offers in-depth information about a topic of discussion.

f) Variability (flexibility). Big data provides support for the constantly changing nature of data by offering extensionality (the addition of new data fields) and scalability (expansion in size).

g) Valence (connectedness). Big data connects common fields to conjoin different data sets.

2.2. Big data in healthcare

Applications of BDA in healthcare are gradually increasing with the growing volume of big data in this context (Galetsi and Katsaliaki 2019; Kamble et al. 2019). Among the possible sources of big data in healthcare are heterogeneous and multi-spectral observations, such as patient demographics (Malik, Abdallah, and Ala'raj 2018), treatment history (Ozminkowski et al. 2015), and diagnostic reports (Amirian et al. 2017). Mehta and Pandit (2018) suggest that such data may be structured (e.g., genotype, phenotype, or genomics data) or unstructured (e.g., clinical notes, prescriptions, or medical imaging). Implementing data in healthcare often requires the generation and collection of real-time data (Tang et al. 2019) of high quality (Wang, Kung, and Byrd 2018). Decision-makers in healthcare organisations are able to take meaningful action based on valuable insights derived from big data (Prasser et al. 2019; Wang, Kung, and Byrd 2018). Healthcare organisations deploy technologies to cope with the changing nature of big data (Harerimana et al. 2018; Zhang et al. 2015). Moreover, big data in healthcare can be employed to connect different fields to comprehensively study a disease (Zhang, Simon, and Yu 2017). In sum, all of the characteristics of big data mentioned above are observable in the context of healthcare.

2.3. Opportunities for BDA in healthcare

The applications of descriptive, predictive, and prescriptive analytical techniques when using big data offer opportunities to enhance the quality of various aspects of healthcare (Kaur, Sharma, and Mittal 2018). The literature proposed different opportunities offered by BDA in healthcare sector, such as the following:

a) Medical diagnosis

A data-driven diagnosis may detect diseases at an early stage and reduce complications during the treatment (Gu et al. 2017; Raghupathi and Raghupathi 2014).

b) Community healthcare

Authorities may take preventive steps against the predicted risks of chronic disease among a population (Lin et al. 2017) and contagious disease outbreaks (Antoine-Moussiaux et al. 2019).

c) Hospital monitoring

Real-time monitoring of hospitals can help government authorities ensure optimal service quality (Archenaa and Anita 2015).

d) Patient care

Customised patient care facilitated by BDA has the potential to provide rapid relief (Salomi and Balamurugan 2016) and reduce readmission rates in hospitals (Gowsalya, Krushitha, and Valliyammai 2014).

2.4. Challenges of BDA in healthcare

The application of BDA to healthcare may face various challenges (Aiello et al. 2019; Amalina et al. 2019). Common challenges in this area include the following:

a) Initial investment

The deployment of the requisites to leverage the benefits of big data incurs huge initial costs for organisations providing healthcare (Szlezak et al. 2014; Wu et al. 2016).

b) Quality of data

The lack of trained personnel and resistance to change in organisational routines may affect the quality of big data accumulated by the organisation (Wang, Kung, and Byrd 2018; Zhang et al. 2015).

c) Quality of insights

The poor quality of heterogeneous biomedical data has the potential drawback of yielding inadequate insights and misleading suggestions (McNutt, Moore, and Quon 2016; Sáez and García-Gómez 2018).

d) Privacy and security

Scholars warn about the privacy and security concerns of patients regarding exposure to unauthorised data access during intersystem exchanges (Mohammed, Far, and Naugler 2014; Weng and Kahn 2016).

3. Methodology

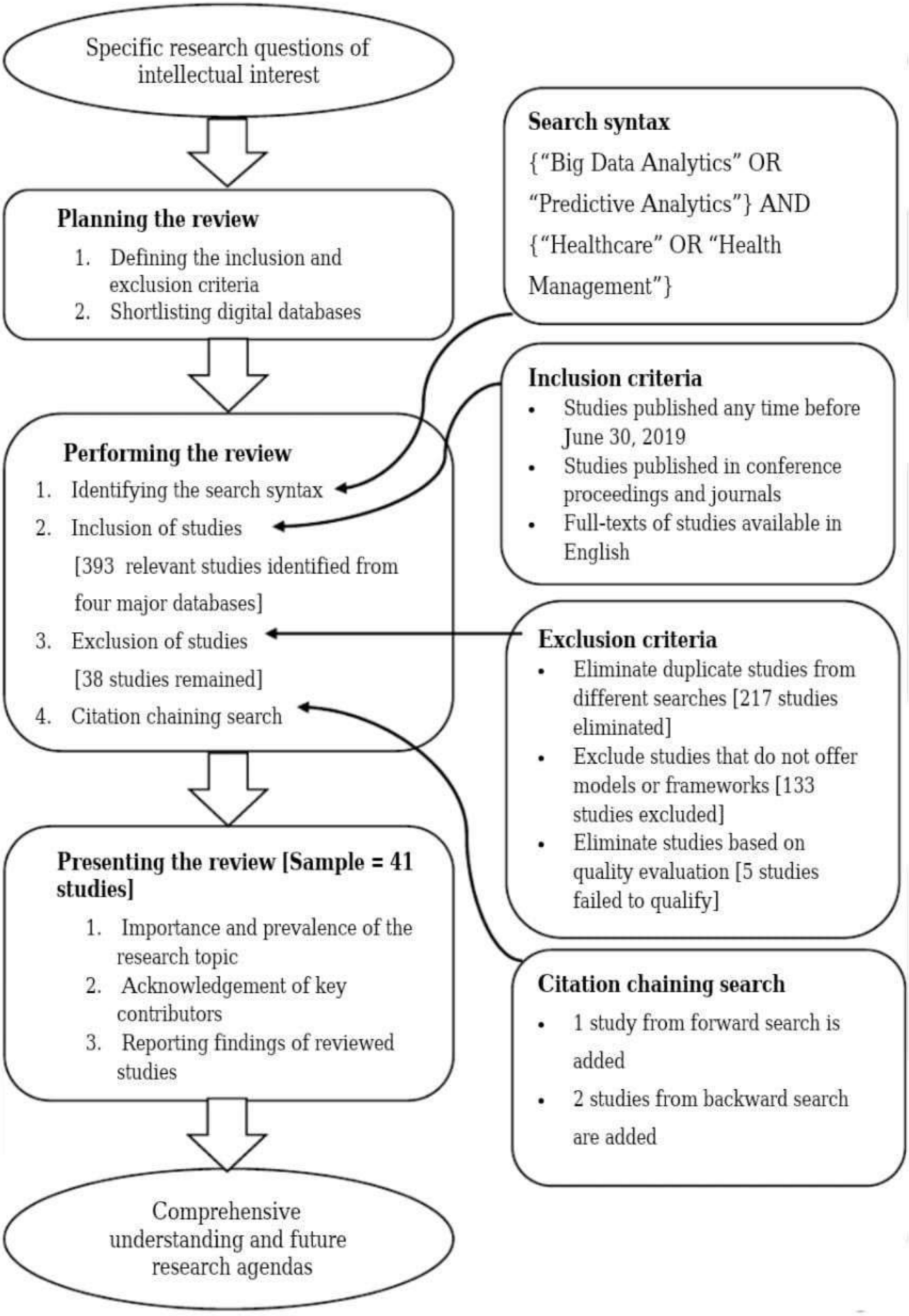
The protocol for the current SLR, as presented in Figure 1, is comprised of three sequential processes: planning the review, performing the review, and presenting the review (Behera, Bala, and Dhir 2019; Tandon et al. 2020). The present SLR includes preset inclusion and exclusion criteria (see Figure 1), as recommended by prior literature (Behera, Bala, and Dhir 2019; Tandon et al. 2020).

BDA can be conceptualised as the analysis of detailed, dynamic, low-cost, massive, and varied data sets to deliver sophisticated solutions (Kamble et al. 2019; Kaur, Sharma, and Mittal 2018). The primacy of BDA has often been attributed to its ability to convert data-scarce decisions into data-rich decisions and to provide competent simulations for problems in various fields (Kitchin 2014). Numerous studies have probed the potential application of BDA in healthcare (Harerimana et al. 2018; Hussain et al. 2019; Palanisamy and Thirunavukarasu 2017). For instance, patient care, patient monitoring, disease diagnosis, treatment methods, and other areas may benefit from BDA applications (Kaur, Sharma, and Mittal 2018). Enhanced risk-profiling based on a Bayesian multitask learning approach has the potential to revolutionise clinical practices by helping to minimise failures and reduce delays in providing preventive interventions (Lin et al. 2017). Wang, Kung, and Byrd (2018) posited that the adoption of a strategic approach by healthcare organisations in deploying BDA may also produce business benefits. However, scholars have argued that the trade-off between harvesting efficient data-driven healthcare solutions and the associated privacy risks in the process is yet to achieve an equilibrium (Kim, Lee, and Chung 2017; Li et al. 2015).

A discussion on the use of BDA in healthcare has been developing recently in the literature (Galetsi and Katsaliaki 2019; Prasser et al. 2019). Consequently, the need to summarise the insights emerging from the discussions has gained prominence (Mehta and Pandit 2018; Zhang, Simon, and Yu 2017). In the quest to meet this need, prior research has followed two approaches to reviewing the literature. The first approach has been to review narrow areas within the literature on the use of BDA in healthcare. For instance, Zhang, Simon, and Yu (2017) highlighted the promising opportunities that BDA offers to advance research on Alzheimer's disease. Malik, Abdallah, and Ala'raj (2018) reviewed the use of BDA in supply chain management in healthcare. Saheb and Izadi (2019) reviewed the use of big data sourced from Internet-of-Things devices in the healthcare industry. Such review studies are not designed to provide a comprehensive review of the literature on BDA in healthcare.

The second approach to reviewing this body of literature has focused on summarising broad topics related to the use of BDA in healthcare. For instance, studies often attempt to summarise the sources of big data (Galetsi and Katsaliaki 2019; Kaur, Sharma, and Mittal 2018), the technologies used in the analysis of big data (Bahri et al. 2018; Harerimana et al. 2018), the benefits offered by BDA (Galetsi, Katsaliaki, and Kumar 2020; Kamble et al. 2019), and the challenges involved in harnessing those benefits in healthcare (Amalina et al. 2019; Mehta and Pandit 2018). We appreciate the valuable knowledge about the use of BDA in healthcare offered by these studies. However, none of these studies have evaluated the quality of the documents in the sample under review. Consequently, the findings of these studies are subject to a critical limitation of their sample design. Therefore, the unavailability of a comprehensive summary of key takeaways from quality articles is a major research gap in the literature on the use of BDA in healthcare. Furthermore, there is a paucity of research aimed at identifying the contexts within healthcare where BDA is commonly applied.

The present study aims to address the research gaps in the literature on the use of BDA in healthcare by conducting a systematic literature review (SLR) (Dhir et al. 2020; Tandon et al. 2020; Talwar et al. 2020). SLRs have been recognised for their ability to summarise valuable knowledge about a topic of importance (Dhir et al. 2020; Talwar et al. 2020) and


Specific research questions of intellectual interest
Planning the review
1. Defining the inclusion and exclusion criteria
2. Shortlisting digital databases
Performing the review
1. Identifying the search syntax
2. Inclusion of studies [393 relevant studies identified from four major databases]
3. Exclusion of studies [38 studies remained]
4. Citation chaining search
Presenting the review [Sample = 41 studies]
1. Importance and prevalence of the research topic
2. Acknowledgement of key contributors
3. Reporting findings of reviewed studies
Comprehensive understanding and future research agendas
Search syntax
{"Big Data Analytics" OR "Predictive Analytics"} AND {"Healthcare" OR "Health Management"}
Inclusion criteria
• Studies published any time before June 30, 2019
• Studies published in conference proceedings and journals
• Full-texts of studies available in English
Exclusion criteria
• Eliminate duplicate studies from different searches [217 studies eliminated]
• Exclude studies that do not offer models or frameworks [133 studies excluded]
• Eliminate studies based on quality evaluation [5 studies failed to qualify]
Citation chaining search
• 1 study from forward search is added
• 2 studies from backward search are added