Opportunities and Challenges in the Use of Big Data in Healthcare: A Literature Review

Grazia Dicuonzo, Graziana Galeone, and Matilda Shini
University of Bari Aldo Moro, Italy

ORCID:
Grazia Dicuonzo: http://orcid.org/

Abstract
Digitalisation and the use of technology are pushing the spread of new business models and improving the efficiency of processes. The demand for innovative and revolutionary applications is increasing, along with the use of big data (BD). The proliferation of large quantities of data is receiving considerable attention in all sectors due to the possibility of using these data in decision-making processes. In the healthcare sector, the role of BD is prominent, especially regarding patient diagnostics, fast epidemic recognition and patient management improvement. To ensure personalised care, the health system must transform individual medical services into electronic forms and favour complete and systemic automation based on the advanced technologies of Industry 4.0. This paper consists of a systematic literature review of the use of BD in the healthcare sector, focusing on the opportunities and challenges. To this end, we selected articles from the Scopus and Web of Science databases. Providing a deep understanding of the state of the art, this paper aims to reveal the implications of the use of BD and offer valuable insights to address future research and identify emerging issues.

Keywords: big data, healthcare, digitalisation, internet of things, artificial intelligence

1. Introduction
The accelerating progress of information and communication technology (ICT) in healthcare, known as Healthcare 4.0 [1, 2], has made large amounts of data, known as big data (BD), available [3]. Some of these data are generated by the spread of the Internet of Things (IoT), which allows the collection of various kinds of data that are generated by a large number of devices [4]. Specifically, these are networks of biomedical sensors, smart devices and processors that enable the measurement of a patient’s vital signs and other biometric information to provide comprehensive care in a variety of settings, including acute care (in hospitals), long-term care (nursing homes) and the patient’s home [5, 6].

To support the growing volume of data available and overcome the inability to store and analyse it using traditional databases, new technologies have been developed that can handle its processing and provide insights for decision making [7–11]. The growing focus on BD has led several organisations to invest heavily in implementing the same [3, 12].
The use of large datasets is especially relevant for healthcare institutions [13], which operate in an uncertain environment and are increasingly engaged in defining the most appropriate therapies for the treatment of each patient’s disease [9, 14].

The use of BD is one element that drives the transition of healthcare systems towards a sustainable path [15]. The storage and sharing of large amounts of clinical data facilitate predictive, timely diagnosis, personalised treatment and the development of precision medicine (patient-centred care) [16, 17].

The modes of operation in the healthcare industry are no longer volume-based but rather value-based; therefore, practitioners’ activities must be geared towards productivity and efficiency requirements [18].

The extremely high availability of each patient’s health data produced by a variety of devices and new technologies [17,19–21], if properly coded and interpreted, allows for improved efficiency and quality of patient care [9, 22].

There are many fields of application for BD as well as technological tools to collect and process it within the healthcare sector. Regarding application, there is the possibility of sequencing the DNA of the patients themselves [23], the real-time electronic recording of patients’ health data and the creation of ‘storable’ databases, which can be accessible to staff members on demand [24]. As for tools, BD analytics (BDA) techniques are applied, including business intelligence methodologies (e.g. data mining and process mining) [19] and cloud computing [25].

The increasing influence of BD has prompted healthcare organisations to use artificial intelligence (AI) and development new professional profiles to effectively exploit BDA (e.g. Data scientists, Big data analysts, Big data specialists) [26, 27].

The use of digital solutions in healthcare has the potential to completely innovate both healthcare and organisational processes. However, BD analysis requires restructuring the technological infrastructure to allow for better screening, more reliable diagnosis, more means and more capillarity in services [28]. In this sense, it is essential to innovate the acquisition, use and storage of information in digital archives (electronic medical records [EMR], health dossiers, electronic health records) [29].

In particular, the importance of BD has grown exponentially with the outbreak of the Covid-19 pandemic—an unknown challenge that has affected healthcare systems worldwide [30]. This emergency situation has put even more emphasis on the need to develop a BD information system to increase data acquisition and integration speeds to generate knowledge in an uncontrolled and ever-changing context [12]. Indeed, in this new phase of the pandemic, BDA could empower and support healthcare systems and organisations to properly monitor Covid patients [22].

Based on the above, this paper aims to examine the main areas of BD application in the healthcare sector to answer the following research question:

*What are the advantages and challenges of practical applications based on BD use in the healthcare sector?*

This paper is structured as follows: section 2 presents the methodology adopted; section 3 describes the main results of the analysis; and section 4 presents concluding remarks and future research directions.
2. Methodology

To answer the research question and identify the main strands of literature around the role of BD as a support tool for the healthcare system, we conducted a systematic literature review [31], which differs from a narrative review in that it is an objective and replicable analysis. Typically, a systematic literature review has four phases [32]:

1. Search: Identification of studies
2. Appraisal: Quality assessment of studies
3. Synthesis: Data extraction
4. Analysis: Data analysis

Starting from 2021, a literature analysis was conducted on the previous decade (2011–2021) to better understand the evolution of the phenomenon [32].

We queried the databases with the following word combinations in the title, abstract and/or keywords of any published article to select the maximum number of articles:

- 'BIG DATA' and 'HEALTHCARE'
- 'BIG DATA' and 'HEALTHCARE' and 'OPPORTUNITY' or 'OPPORTUNITIES'
- 'BIG DATA' and 'HEALTHCARE' and 'CHALLENGE'
- 'BIG DATA' and 'HEALTHCARE' and 'WEAKNESS' or 'WEAKNESSES' and 'STRENGTH' or 'STRENGTHS'

The research was delimited with respect to the year of publication and the research area ‘business, management and accounting’. The search was also limited to articles published in journals and written in English.

As shown in Table 1, records were retrieved from the initial keyword search in the two databases. After excluding duplicates, 444 items remained. From the initial screening, based on the content of the title and abstract, we excluded 361 articles that were not considered relevant to either the healthcare sector or BDA. The remaining 83 papers met all inclusion criteria.

<table>
<thead>
<tr>
<th>Search criteria</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+) Search results: Scopus</td>
<td>411</td>
</tr>
<tr>
<td>(+) Search results: Web of Science</td>
<td>402</td>
</tr>
<tr>
<td>(=) Total papers from Scopus and Web of Science</td>
<td>813</td>
</tr>
<tr>
<td>(-) Duplicates</td>
<td>369</td>
</tr>
<tr>
<td>(-) Non-relevant papers</td>
<td>361</td>
</tr>
<tr>
<td>(=) Total papers considered</td>
<td>83</td>
</tr>
</tbody>
</table>

Source—Author’s own work

The analysis of the time interval showed that the areas of BDA applications in healthcare have been gradually increasing since 2014 (Fig. 1).
3. Results

3.1. Health and m-Health in the challenge of networked patient management

The fields of medicine and health, which are becoming increasingly invested in digital transformation, will be disrupted in the future by significant changes. These changes will be driven by measures aimed at cost containment and effectiveness of the treatment pathway [33]. In this regard, the spread of technology and digitisation is encouraging increased investments by healthcare organisations in the development of electronic health (e-health) and mobile health (m-health) [34]. Generally, e-health refers to seeking health information via desktop or laptop computers that are enhanced by an AI system [35]. Data enhancement and the extraction of actionable information for decision making has prompted healthcare organisations to use AI and the skills required to effectively exploit BD [26, 27].

BD in healthcare can comprise a large collection of data from various healthcare sources, and these data can enable increasingly personalised treatments, evaluate the treatments’ effectiveness and reduce clinical risk through innovative ways of managing and controlling processes [36, 37].

Combined with BD, AI in healthcare, if appropriately implemented, would enable improved healthcare services through more precise diagnoses [38], discoveries in the pharmacology and epidemiology fields [39, 40], precision and personalised healthcare treatments in support of the shift to a patient-centred model [41–43] and operational effectiveness. This efficiency results from the automation of certain tasks in terms of support to physicians in carrying out clinical activities and the decision-making process (considered important by healthcare providers and scientific societies) and in healthcare and administrative back office processes [44].

According to the Stanford University study ‘Artificial Intelligence and Life in 2030’, healthcare is one of eight industries where the impact of AI will be most significant.

Currently, digital technologies that allow the use of health-relevant data from portable medical devices and so-called ‘wearables’ (wearable devices, such as fitness bracelets)
are particularly widespread. This new model of social healthcare, which is aimed at the health and well-being of the citizen/patient, is better known as the ‘m-Health’ paradigm.

M-health, according to the World Health Organization, is ‘a practice of public and medical healthcare supported by mobile devices, such as smartphones, patient monitoring devices, personalised digital assistance and other wireless devices. In addition to providing constant contact with healthcare providers, these devices enable better treatment adherence and provide aggregate data on patient health status [45].

Therefore, m-health (i.e. apps related to the health status and lifestyles of citizens/patients) and health IoT (i.e. vital signs and/or bio-images resulting from the connection of medical devices or other sensors, such as bracelets, watches, detectors, etc.) collect data favouring the exchange of information between doctors and citizens/patients remotely and/or on the move [21]. Patients will be increasingly inclined to use AI due to the improved access to healthcare introduced by such systems in terms of simplicity, speed and efficiency (e.g. autonomous management of access to first or follow-up visits via telemedicine). The patient will become more involved in all phases of care and will be able to understand the importance that the quality of their digitised clinical data will assume in relation to their care processes [46].

For example, fitness monitors, such as Fitbit, Jawbone, Germain and Suunto, are designed to monitor and track symptoms for better symptom control, ongoing support for disease management [21, 47, 48] and access to timely communication with healthcare providers [49]. Several applications have also been developed for monitoring the health status of patients, such as MobiHealth or AlarmNet. The former collects data from devices worn by patients, providing the opportunity to intervene in a timely manner in case of accidents or ill health, while the latter is a biosensor network system consisting of heart rate, oxygen saturation and electrocardiograms [50].

The use of data from IoT technologies has also reduced wait times and enabled remote monitoring of patients’ conditions, diagnosis and the prediction of diseases in their early stages [17, 22, 51, 52].

### 3.2. Opportunities and challenges of big data in healthcare systems

New technologies and innovation have brought methodological improvements to the analysis of large amounts of data [53]. BDA is now an important competitive advantage [14, 54] that has significant benefits regarding efficiency in the healthcare industry [1, 2, 55, 56], including reducing healthcare costs and comparing quality of care amongst different healthcare organisations [6].

New technologies and the use of BD in healthcare have enabled health information systems to change drastically over time (Bharati, 2020). New database management systems facilitate the integration of clinical information [6]. Among these, there are some that ensure data integration and retrieval (e.g. MongoDB, Marklogic and Apache Cassandra) and some that store large volumes of data and have different formats (e.g. Apache HBase and NoSQL) for collecting, processing and analysing data using machine learning algorithms (e.g. Hadoop) [57, 58].
The digitisation of health information has enabled a shift from paper medical records to EMR as well as the digital developments in diagnostic imaging [3, 56]. EMR can electronically record details about a patient’s health, such as vital signs, medication intake, radiology reports or lab data [3]. The implementation of EMR allows for improved healthcare, creating a link between different healthcare facilities and facilitating personalised care and decision making [3, 50].

The increasing push for prevention and the need for early identification of signs of disease are gradually eliminating older machines and leading to the emergence of new powerful digital diagnostic technologies. The proliferation of these tools has led to the creation of the Digital Imaging and Communication in Medicine standard, which defines the rules for archiving and sharing images [59, 60].

The use of BD is relevant, especially in unforeseen situations, such as the current health emergency of Covid-19 [1, 30]. Following initial difficulty in containing the emergency, high-tech technologies, such as BDA and AI, have played important roles in the battle against Covid-19 [10, 14, 61]. For example, advanced analysis techniques, such as monitoring people’s movements, understanding health trends and managing and administering drugs, have been developed to predict the spread of the virus and reduce the number of positive patients [6, 14, 62].

For BD to produce the desired benefits, it will be necessary to have not only highly configured machines and improved data quality but also qualified personnel to manage the data accumulated by the system [63, 64]. Despite the great potential of BD, technology development does not always run parallel with organisational and administrative upgrades. Costs are reduced in the long term, but it is necessary to invest in the implementation of appropriate tools as well as in adequate training of healthcare professionals [65]. Ongoing training of health providers and establishing protocols in the different stages of health analytics (data acquisition, data storage, data management, data analysis and data visualisation and reporting) should be central to ensuring adequate health and social care and should support the restructuring of healthcare processes [66]. The lack of professionals with appropriate technical skills and multidisciplinary knowledge can increase data entry errors and lead to information loss [67]. To fully exploit data and convert it into a strategic resource, it is therefore necessary to make a significant investment in healthcare structures, such as by hiring analytics experts (i.e. professionals who can identify problems from data and propose the most appropriate solutions) [65].

Further priority actions concern the restructuring of the technological infrastructure, which is certainly linked to the search for the financial capacity to support this investment [68, 69].

The large amount of data to be managed has highlighted the inadequacy of traditional methods and the need for restructuring of the technological infrastructure to integrate traditional data analysis tools and techniques with a computational technology capable of enhancing and extracting useful information for decision making [68–70].

Furthermore, data are processed in different places and transmitted through protocol networks, which is a difficult process to control [71, 72]. BD can be subject to transfer by healthcare providers (e.g. patient biometric data) to the cloud [18, 52, 72]. In this
field, studies highlight the advantage of storing BD without the need to install physical storage units, but they simultaneously object to excessive dependence on network connection speed and the frequent unreliability of the cloud regarding initial expectations [53, 72]. Therefore, there is a need to investigate new protocols for more efficient and secure transmissions of sensitive health data [73]. By definition, BD involves the collection of large amounts of data, which raises ethical concerns because the storage and processing of personal data can result in privacy violations [48, 54, 55, 74]. To protect these aspects, data governance and data compliance activities are necessary [47, 57]. Some authors have drawn attention to the emergence of appropriate data governance regulations (e.g. U.S. Health Insurance Portability and Accountability Act [HIPAA], Personal Information Protection and Electronic Documents Act [PIPEDA]) [75]. These were established by public and private entities to safeguard patients’ rights to consent to the transmission of data to third parties and to request changes to data [45, 76]. Based on existing regulations, Jibai and Najdi (2019) suggested a specific framework for implementing the same data governance regulations in the healthcare sector. The framework includes absolute consideration of privacy practices and the ad hoc creation of specific bodies or individual figures (e.g. data council, data immuniser) [75].

**Figure 2:** Big data: Opportunities and challenges (Source—Author’s own work)

4. Conclusions and Future Research

According to the literature review, in recent years, the focus on new and more advanced technological solutions has enabled the diffusion of innovative and revolutionary applications in all sectors, including healthcare [27]. The application of ICT is becoming a reality, even in healthcare organisations that are moving towards Healthcare 4.0 [77, 78], such as the IoT paradigm [79] and BDA [80].
This technological revolution enables timely decision making for appropriate therapy diagnosis and specific treatment for particular or rare pathological cases [81], enabling the revitalisation of health systems that have been challenged, especially during Covid-19. In particular, the current pandemic calls for the mitigation of privacy associated with patient risks and the protection of data privacy [82].

The creation of new skills in the field of data analysis is necessary to take advantage of the benefits of BD, especially in terms of reducing therapeutic errors, identifying chronically ill patients and forecasting preventive actions, which can also prevent overflows in emergency channels, such as the emergency room. Human resources are an indispensable strategic resource in transforming consolidated data into valuable information for business decision makers [83]. The literature review aims to analyse the revolutionary role of BDA in the healthcare system and highlight the benefits and challenges that need to be addressed, which, if properly managed, can translate into an important competitive advantage for healthcare providers, especially in terms of error reduction and improved personalisation of care.

Therefore, the future research agenda is aimed at further uncovering issues within previously discussed topics, especially in turbulent contexts, such as the current pandemic.

References

[3] Chen PT. Medical big data applications: Intertwined effects and effective resource allocation strategies identified through IRA-NRM analysis. Technological Forecasting and Social Change. 2018;130:150-164.


El Sayed EMAM. How AI, data science and technology is used to fight the pandemic COVID-19: Case study in Saudi Arabia environment. Research in World Economy. 2020;11(5):409-419.


[78] Oliver N, Arnesh T, Tak I. Smart hospital services: Health 4.0 and opportunity for developing economies. 29th International Conference of the International Association for Management of Technology: Towards the Digital World and Industry X.0, IAMOT 2020; 2020 Sep 13-17; Cairo, Egypt.


