Journal of Big Data

CASE STUDY Open Access

The power of big data mining to improve the health care system in the United Arab Emirates

Khawla Eissa Alhajaj[1,2] and Immanuel Azaad Moonesar[1,3*]

*Correspondence:
immanuel.moonesar@mbrsg.ac.ae;
immanuel.moonesar@outlook.com

[1] Mohammed Bin Rashid School of Government, Dubai, United Arab Emirates
[2] Dubai Health Authority, Dubai, United Arab Emirates
[3] Academy of International Business- Middle East North Africa Chapter, Dubai, United Arab Emirates

Abstract

Context: Collecting and analyzing data has become crucial for many sectors, including the health care sector, where a hefty amount of data is generated daily. Over time, the amount and complexity of this data increase substantially. Consequently, it is considered *big data* that cannot be stored or analyzed conveniently unless advanced technologies are incorporated. Latest advances in technology have divulged new opportunities to use big data analysis to track a patient's record and health. Still, it has also posed new challenges in maintaining data privacy and security in the healthcare sector.

Purpose: This systematic review aims to give new researchers insights into big data use in health care systems and its issues or to advise academics interested in investigating the prospects and tackling the challenges of big data implementation in rising nations like the UAE. This study uses a systematic methodology to examine big data's role and efficacy in UAE health care.

Methods: The research follows the methodology of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) for reporting the reviews and evaluating the randomized trials. Furthermore, the Critical Appraisal Checklist for PRISMA 2009 was applied for the research.

Findings: The study concludes that the healthcare systems in the United Arab Emirates can be improved through big data; however, the country authorities must acknowledge the development of efficient frameworks for performance, and quality assessment of the new health care system is significant. The said goal can be achieved via integrating big data and health informatics with the help of IT specialists, health care managers, and stakeholders. Data privacy, data storage, data structure, data ownership, and governance were the most often expressed concerns.

Contribution to knowledge: By discussing numerous issues and presenting solutions linked with big data, the current study contributes substantially to the knowledge of big data and its integration into health care systems in the UAE.

Keywords: Big data, Health care system, United Arab Emirates, PRISMA systematic reviews, Big data opportunities, Big data challenges

Introduction

Technology has become a vital part of our daily lives, and adopting new technologies has drastically altered our daily lives. Technology has made our lives more comfortable and efficient, regardless of age, level of knowledge, or even the reason for utilizing it. We are experiencing technological wonders in the shape of smartphones, the Internet of Things (IoT), robotic surgery, and applications that use Artificial Intelligence (AI). IoT transforms healthcare by improving patient care, treatment results, costs, provider workflows, performance, and patient experience. Healthcare IoT has problems. IoT-enabled devices gather a lot of data, including sensitive information, raising security issues [41]. Rich healthcare data frequently includes survival information. Healthcare data analysis is crucial since it can save lives and improve the quality of life [41]. IoT revolutionized health care systems and administration. IoT promises to revolutionize healthcare.

We are shifting from regular technology use to more complicated technologies connected via a robust internet and frequently generate vast amounts of data. With the growth of data from numerous mobile networks, cloud computing systems, health applications, and electronic medical records, there is an increased need for a comprehensive approach to maintaining and updating information. The expanding data and knowledge of the patients and the relevant health care activities are getting more challenging due to the data's speed, amount, and complexity. Kumar and Manjula reported that health care facilities generate an abundant amount of data each day that is centered around the patients, medicines, treatments, diseases, research, and other similar factors [29]. To manage this data more efficiently, modern health care units choose to digitize the data related to patients. Worldwide, medical institutes shifted from the traditional paper-based medical file to the electronic medical record, providing help in managing patient information, lab tests, medications, and medical imaging.

Electronic medical records (EMR) are considered to be an essential, rich platform containing patient information. EMR captures all demographic data, lab results, radiology images, and free-text notations. That collective information is beneficial as a database for many longitudinal studies. Mining data from EMR can help understand disease signs and symptoms and the progression of a particular disease. It also improves clinical knowledge and understanding of a specific phenomenon and assists in clinical trials, disease management, and therapeutic trials [15, 19].

Further, it assists in predicting disease progression, comorbidities, and mortalities [32]. Data comes from various sources, including electronic medical files, home sensors, and wearable devices. As such, it will generate a massive amount of data known as *big data*. *Big data* refers to massive data, although the term has no universally accepted definition. The oldest definition is provided by Laney, who observed that (big) data was growing in three different dimensions, namely: volume, velocity, and variety (known as the *three V's*) [24]. This definition has been expanded by Demchenko et al., who define big data by *five V's*: volume, velocity, variety, veracity, and value [18]. Volume refers to the amount of massively generated data that requires a unique storage format. Data velocity means the high speed of data generated from different resources. Variety of data implies the complexity of data that varies from numerical data to text notation or from numerical data to text notation or a (radiological) image. Finally, veracity refers to the accuracy of the data, and value evaluates the quality of data [35].

Wang expands on big data and defines it as a data set that cannot be analyzed by a standard computerized method [44]. Big data is segregated into structured, unstructured, and semi-structured forms of the data. The structured data can be stored, accessed, and processed in a specific format. It is an already-segregated and dedicated form of easily retrievable and readable data. The unstructured data is not explicit in its structure, as it was discussed for the structured layout of the data. As stated by Wu and Lin, this type of data possesses multiple challenges in terms of processing as well as retrieving valuable information from it [45]. The processing resources' limitations include data-related, language, relationship identification, and technical issues [2, 3]. The durability and efficiency of data variety and data skew were evaluated using a broad range of simulated and real-world healthcare datasets [3–5, 22]. A typical example of this data is the data that comes from heterogeneous sources, meaning a combination of text files, images, and videos. Data heterogeneity is due to the mixing of structured and unstructured data, having its roots in various quantitative or qualitative platforms. The quantitative data sources include laboratory tests, images, sensor data, and gene arrays. The qualitative data sources include demographics and textual information [34]. One of the critical challenges in this regard is related to the accuracy and trustworthiness of the data since the credibility of the data may be challenged as it is from unmanaged sources. First, to preprocess multi-database medical record linkage, employ approximation query processing. Second, aggregate queries may get approximations.

Preprocessing may be used in multi-database systems to locate patient records. This is the initial stage in connecting. If gathering aggregate data, imprecise replies may be adequate. At least they may spark additional inquiry. This estimate is critical for multi-database query design and optimization [3, 4, 48]. Health care systems need to analyze the unstructured and semi-structured data to get the ultimate benefits of big data technology. The extraction and retrieval of big data may be subject to challenges related to social and legal technicalities. These social and legal issues might be generated due to problems associated with data ownership, privacy, identification, and governance [27].

Big data and health care system within the United Arab Emirates

The United Arab Emirates (UAE) health care system is operated by government-funded health services and the rapidly growing private health sector. The standards of the health services provided by both sectors are acceptable. The healthcare industry of the UAE is realizing the potential of big data analysis which can transform the health care system (refer to Fig. 1). According to Bani-issa et al., such developments are the inactive lifestyle among the residents, leading to an increase in chronic diseases such as diabetes [11]. Several regions in the Middle East, including the UAE, have undergone or are considering implementing health care insurance, which then needs to analyze the large volume of health data generated from claims. The UAE introduced a standardized insurance coding system to deal with the situation and improve process efficiency. The insurers in the UAE are pricing premiums based on little historical data due to the lack of big data analysis tools and the sophisticated nature of the big data. The availability of big data will enable insurers to paint a clear picture of health care in the region. It will allow them to accurately predict the validity of the claims [8, 33].

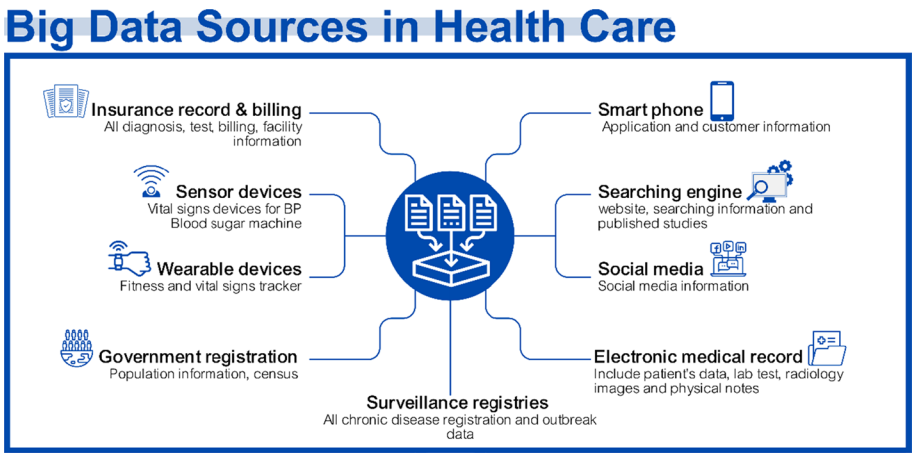

Fig. 1 Big data sources in health care (Source: Authors)

The UAE's vision is to provide world-class healthcare by 2021, and the government's direction is to foster innovation in the healthcare system to achieve its vision. Many strategies have been explored to ensure that people are provided with a high-quality care system and to implement SDGs, particularly Goal 3 (ensuring healthy lives and promoting well-being for all ages) [39].

With advanced technology in the UAE, smart government, and public service, big data helps provide a big database within the country, especially in the healthcare sector, which can assist in a better understanding of the population's health and provide the necessary service. To meet the government's objective of providing world-class healthcare and ensuring sustainability in delivering health and well-being to everyone in the UAE, big data mining and data analysis may help improve services and health programs to create a healthier, happier population. However, despite the availability of big data in the UAE and the potential to use big data as a government seeking innovation and exploiting big data, there are constraints and a shortage of published research on big data mining in the UAE, particularly in the health care system. Although, there is no standardized government approach and policy regarding big data mining or storing. However, a large amount of data is generated daily among different entities within the UAE. Regarding consensus, the UAE open data policy was launched in 2018 as per the UN eGovernment Survey to help access data without restriction [43]. However, by 2020, not all data was accessible, and there remain restrictions on available data from the entities [38, 40].

Much research in the literature review provides agreement about big data mining and its beneficial role in enhancing the health care system. Yet, there is still no unified process or solutions for big data mining and how to make it possible. This research will help understand the importance of harnessing big data and utilizing it to enhance the health care system and identify its challenges and limitations.

Big data and sustainable development goals

The UN developed a 2030 strategy to fight poverty, ensure equity among people and address global challenges through 17 sustainable goals [42]. According to Wu,

policymakers, decision-makers, and investors need factual, accurate, and real-time data to adopt the appropriate policy decisions to accomplish the Sustainable Development Goals (SDGs). They then need to be able to check the impact of the policy, which can be achieved through the analysis of big data from different sources. Similarly, a report released by United Nations states that the big data revolution can contribute to Sustainable Development Goals (SDG) by providing accurate and reliable data and analyzing the data to develop policy and plans to achieve SDG 2030 [43]. However, the main concerns were the inadequacy in technology adoption among all countries and data privacy and transparency regardless of the industry [8, 10]. For instance, medical records are vital to medical care and include sensitive personal data; therefore, keeping electronic medical records private is a crucial difficulty [8, 33].

Furthermore, blockchain can hold accessible, immutable, tamper-proof medical data [16]. Therefore, doctors and nurses would use a Big data analysis can provide the ability to monitor the progress toward achieving SDGs by 2030. Big data analysis can be more cost-effective and faster in tracking SDGs than, for example, monitoring poverty by traditional methods such as questionnaires or interviews, which can be ineffective and time-consuming and require significant effort [13, 36].

The focus on SDGs Is stated in Goal 3, to "ensure healthy lives and promote well-being for all at all ages" [42]. Big data can help in providing precise and clear information about health. Barrett et al. makes this point by adopting big data analysis to understanding population behavior, social and environmental factors (2013) by adopting big data analysis [12]. This will help in population health management, prevent the disease, and target subpopulations by having accurate and real-time data. Accurate or approximate processing in health care systems has been associated with hospital statistics deemed critical for assessing performance and ensuring safe and dependable healthcare delivery [6–8]. Data quality is described by correctness, validity, reliability, completeness, legibility, timeliness, accessibility, usefulness, and secrecy [5, 20]. All data are susceptible to missing values, bias, measurement inaccuracy, and human input and processing errors. These difficulties include technical, behavioral, and organizational [20]. Therefore, big data analysis can help achieve SDGs by promoting well-being and chronic disease prevention through big data analysis.

Research gap

As discussed in the literature, big data is emerging as a great source of improvement in different sectors of the world, especially for countries adopting advanced health care systems. Most developed countries have recognized the importance of big data and have shown interest in improving the health care system through the collection and analysis of big data [14, 32]. The UAE is an example of a nation whose healthcare systems are up to date and equipped with modern health facilities. However, there is limited research in this context that has considered that despite existing challenges of data security, data classification, data modeling, data storage, data accommodation, and technology incorporation, whether the integration of big data and health care can emerge as a sustainable system. Implementing big data analysis will be difficult in countries similar to the UAE, with a high population and complex health care systems. The motivation for this systematic review is to provide new researchers insights into the field of big data usage in health

care systems, together with its associated challenges, or as a guide for researchers interested in exploring the opportunities and solving the challenges of big data implementation in emerging countries such as the UAE.

This research study aims to gain further insight using a systematic approach to review the role, effectiveness, and evaluation of big data in the area of health care within the UAE. The research questions of the study using the systematic review of the methodology of the article are:

- RQ1: What is the role of big data in the health care system?
- RQ2: What are the potential opportunities to enhance quality-of-care services through integrating big data in the health care system?
- RQ3: How to best understand the challenges of implementing and using big data technologies?

With RQ1, we can investigate the role of big data usage in improving health care systems and their industry. Moreover, RQ2 allows us to find the existing solutions to enhance the quality of care services in health care systems using big data. RQ3, on the other hand, gives an insight into how various challenges can be mitigated to enhance the implementation of big data technologies in health care systems.

Methodology

The current research follows the methodology of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) for reporting the reviews and evaluating the role and usefulness of big data in health care systems for emerging countries such as the UAE. According to Tricco et al. [32], to assess the quality of the selected research articles for current research, the critical appraisal PRISMA checklist 2009 was used [26, 28, 30]. Figure 2 shows the flow of information through the different phases of a systematic review for this study. However, due to limitations in publication or access control, access to a wide range of journals will limit the overview of available literature during the study time. Furthermore, restrictions on published papers in English will restrict the information to publish valuable articles in other languages.

Inclusion criteria

The article searching methodology discussed above resulted in thousands of research articles. Instead of considering all of them, the most relevant research articles were segregated from those less relevant. The titles, abstracts, and keywords of the research articles were screened, and those articles which discussed the relation of big data to health care systems were separated for full-text review. The screening process was made more efficient by removing duplicate research articles at the eligibility phase of PRISMA.

A systematic review of the studies was conducted, and the articles were judged according to their objectives, methodology and study design, use of authentic data sources, validity and reliability of the study method, analysis, and consideration of ethical issues. Further, comprehensiveness of the description of the findings and outcomes, the appropriateness of the tools used for data mining, the suitability of the qualitative methodology, the use of valid research designs, and clearly stated research findings were

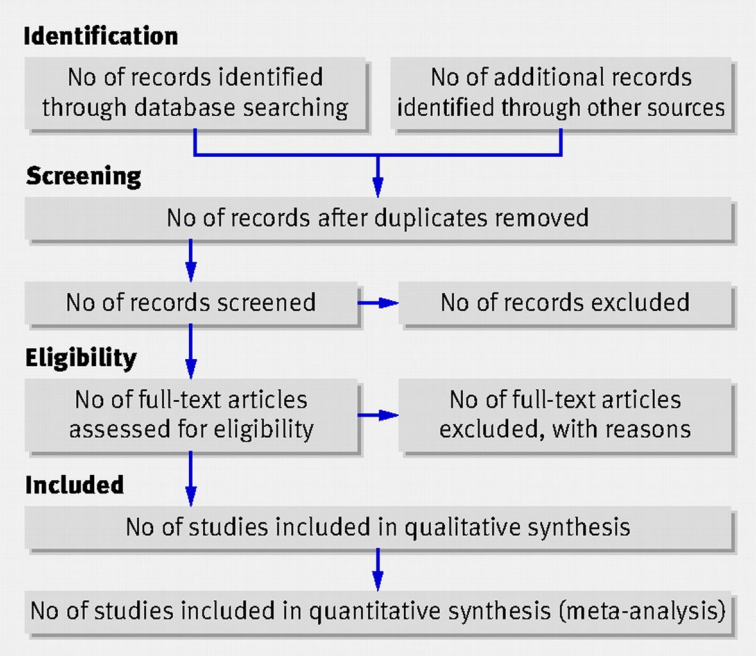

Fig. 2 Flow of information through the different phases of a systematic review [26]

considered in the review. Linkage to the current study questions and objective was taken into consideration. Only one study was conducted on big data within the context of the UAE. The study used quantitative data to better understand the topic, with the same context of the UAE as the current study.

Exclusion criteria

The different research articles were excluded from consideration in the research based on the criteria of quantitative studies, surveys, focus groups from the research other than health care, feasibility studies, work environments other than health care, data collection techniques adopted, editorials, and short reports, and articles which were not reputably published, such as in international journals. The assessment of the research articles based on the criteria mentioned above further narrowed the number of articles to be included in the current research, and it was ensured that the remaining articles are the most relevant and high-quality manuscripts, which make the findings and outcomes of the study authentic and most reliable. The research that made mention of computerized or digital tools to analyze big data and technical details about the transfer of big data, processing of big data, storing the data, cleaning the data, and analyzing the data were excluded as well as these are not one included in the research objectives and technical details will not add value to the current study. Artificial Intelligence (AI) algorithms and the role of big data in AI studies were further excluded.

Quality assessment and processing steps

To evaluate each article in terms of quality, the Critical Appraisal PRISMA Checklist 2009 (refer to Table 1) was applied in the current research for qualitative studies. In addition, the quality of papers was judged according to their objectives, as was the

Table 1 PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both	Supplementary file – Microsoft Excel
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number	Supplementary file – Microsoft Excel
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known	Supplementary file – Microsoft Excel
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS)	Supplementary file – Microsoft Excel
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number	Supplementary file – Microsoft Excel
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale	Supplementary file – Microsoft Excel
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched	Supplementary file – Microsoft Excel
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated	Supplementary file – Microsoft Excel
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis)	Supplementary file – Microsoft Excel
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators	Supplementary file – Microsoft Excel
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made	Supplementary file – Microsoft Excel

Table 1 (continued)

Section/topic	#	Checklist item	Reported on page #
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis	Supplementary file – Microsoft Excel
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means)	Supplementary file – Microsoft Excel
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis	Supplementary file – Microsoft Excel

approach of the research methods, objectives, and abstract use of authentic data sources. The validity and reliability of the research design approaches adopted were evaluated. There was also a consideration of ethical issues. A thorough evaluation was done of the comprehensive description of findings and outcomes, and of the appropriateness of the qualitative methodology, use of valid research designs, and clearly stated results of the research., The use of valid research designs and clearly stated research findings were analyzed. The flow diagram of the PRISMA methodology adopted for the research is illustrated in Fig. 3, where the parameter 'n' represents the number of articles identified and the PRISMA checklist in Table 2. Extended results are attached in Table 3.

Results and discussion

Big data in health care

The findings flow diagram of the PRISMA methodology adopted for the research is illustrated in Fig. 3, where the parameter 'n' represents the number of articles identified, and the PRISMA checklist write-up is in Table 2. The extended results of the articles are attached in Table 3. The selection process is visualized in Fig. 3 as a PRISMA flow diagram, showing the number of papers (n) obtained from each stage of the review process, i.e., search results, duplicate removal, title and abstract screening, full-text screening, and final selected papers.

RQ1: What is big data's role in the health care system?

Big data is now considered the gold standard of the new technological era, especially for an institute that encounters great quantities of data daily, such as in the health care sector. Considering the importance of big data in healthcare, the content is analyzed to showcase how data is utilized in the healthcare sectors of the UAE and the certain challenges faced by the healthcare industry in this regard.

Big data can improve the efficiency and effectiveness of health strategy and policy. They can shift the policymaking from the patient visit to more advanced (value-based) policies with big data analysis, which can accurately reflect the population, as Gamache mentioned [21]. It was also noted by Auffray et al. and Balsam et al. that big data could help formulate a prevention and prediction strategy, in addition to improvement in the

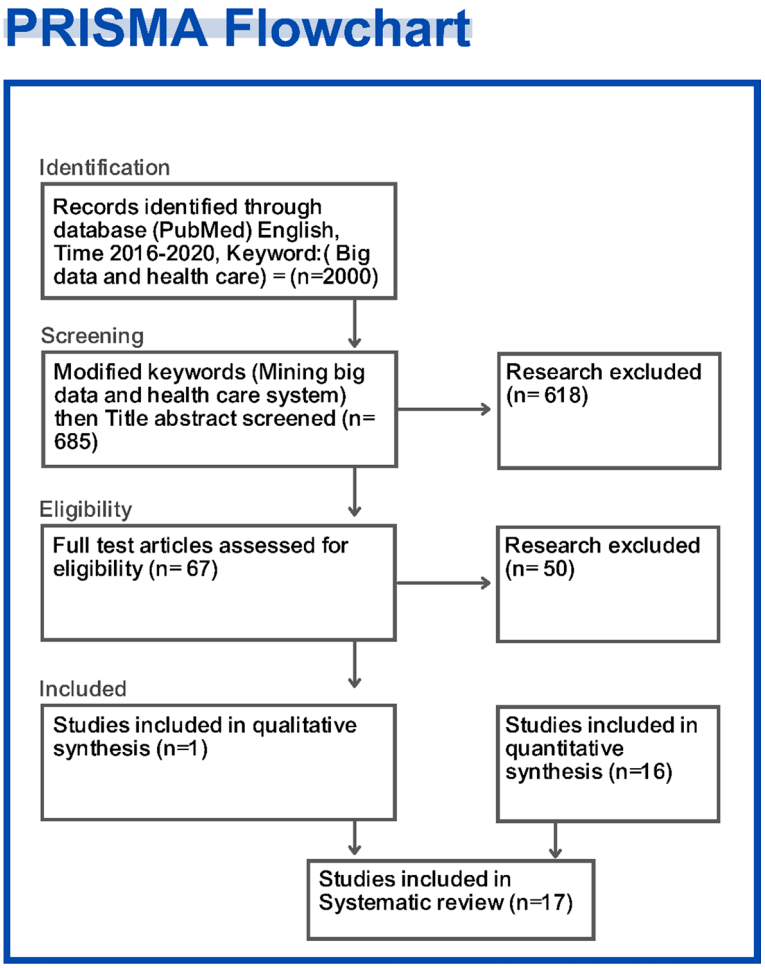

Fig. 3 PRISMA Results of Flow of information through the different phases of a systematic review

health of the population [9, 58]. Big data can help improve individual health progress and shift toward personalized medicine. It can predict personal health and improve clinician decisions [25]. In addition, big data can assist in a clinical trial in choosing suitable participants and make the process less expensive. Big data can also give more insight into drug safety, early detection, and trace adverse drug side effects. Big data can also help monitor infectious disease trends and track the cases among the population, which can assist in making the right decision and act immediately to limit the spread of contagious diseases.

It is mentioned by Zeng that big data integration with different behavioral and social factors can help in better understanding of health care and health disparities, to enhance the population health and reduce such disparities [47]. Zeng discusses two vital areas where big data can play an essential role in the health care system: it can help integrate and understand social factors impacting population health to improve the population's quality of care [47].

The USA HITECH Act helps adopt the EMR, contributing to data generation and reducing health disparities. The same applies to the context of the UAE, where an EMR is adopted among various public and private sector authorities. The EMR is

Table 2 PRISMA 2009 Checklist- Big Data Write-up

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Systematic review for "Big data & health care system."	
ABSTRACT			
Structured summary	2	Research method limited to secondary data through a published journal on "PubMed" website was done, and limitation of the studies to the keyword "Big data mining and Healthcare" was done. The period was selected between January 2016 and September 2020, limited to English language and free full text	Supplementary file – Microsoft Excel
INTRODUCTION			
Rationale	3	Many studies describe the importance of harvesting big data in different sectors, including health care; however, the authors did not locate a single research within the context of the United Arab Emirates (UAE). Therefore, the authors are verifying the applicabilities of various studies finding within the UAE health sector. The motivation for this systematic review is to provide new researchers insights into the field of big data usage in health care systems, together with its associated challenges, or as a guide for researchers interested in exploring the opportunities and solving the challenges of big data implementation in emerging countries such as the UAE. Thus the following research questions are designed, which this paper aims to answer: With RQ1, we can investigate the role of big data usage in improving health care systems and its' industry. Moreover, RQ2 allows us to find the existing solutions to enhance the quality of care services in health care systems using big data. RQ3, on the other hand, gives an insight into how various challenges can be mitigated to enhance the implementation of big data technologies in health care systems	Supplementary file – Microsoft Excel

Table 2 (continued)

Section/topic	#	Checklist item	Reported on page #
Objectives	4	This research study aims to gain further insight using a systematic approach to review the role and effectiveness of big data in the area of health care within the UAE. The objectives of the study using PRISMA are: •RQ1: What is big data's role in the health care system? •RQ2: What are the potential opportunities to enhance quality-of-care services through integrating big data in the health care system? •RQ3: How to best understand the challenges of implementing and using big data technologies?	Supplementary file – Microsoft Excel
METHODS			
Protocol and registration	5	The research was done through a single web engine, "PubMed" and all the collection criteria like language, full free text, and period were entered through the website and the keywords	Supplementary file – Microsoft Excel
Eligibility criteria	6	Studies were included in qualitative studies, published journal from January 2016 to September 2020, limited to English language, full free text	Supplementary file – Microsoft Excel
Information sources	7	The database was "PubMed", no authored as contacted as the study limited to full free text only	Supplementary file – Microsoft Excel
Search	8	The search was done through the website" PubMed" through keywords (Big data and Healthcare), which resulted in thousands of studies that were irrelevant to the research aim and objective. Then the keywords were more specified to (Mining bigdata and Healthcare), restricted to the English language and time (2016–2020) resulted in 685 reviews	Supplementary file – Microsoft Excel
Study selection	9	A systematic review of the studies was conducted, and the articles were judged according to their objectives, the approach of the description methodology and study design, use of authentic data sources, validity and reliability of the study method, analysis, consideration of the ethical issues, comprehensive description of finding and outcomes, the appropriateness of the tools used for data mining, the suitability of the qualitative methodology, use of valid research designs, and clearly stated findings of the research	Supplementary file – Microsoft Excel

Table 2 (continued)

Section/topic	#	Checklist item	Reported on page #
Data collection process	10	The search was done through the website" PubMed" through keywords (Big data and Healthcare), which resulted in thousands of studies that were irrelevant to the research aim and objective. Then the keywords were more specified to (Mining bigdata and Healthcare), restricted to the English language and time (2016–2020) resulting in 685 reviews Then scanning of the abstract was done to check the relevant studies relevant to the objective was highlighted for further screening. Sixty-seven studies at the first stage were included, and the rest were excluded. Then a quick review of the studies was done and checked if it was relevant to the study aim and objective. Quantitative studies, and interview were excluded as well. Studies that highlight a different tool for data mining are excluded as well. This resulted in exhaustive studies, which included 22 reviews	Supplementary file – Microsoft Excel
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made	Supplementary file – Microsoft Excel
Risk of bias in individual studies	12	Describe methods used for assessing the risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis	Supplementary file – Microsoft Excel
Summary measures	13	State the principal summary measures (e.g., risk ratio, the difference in means)	Supplementary file – Microsoft Excel
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis	Supplementary file – Microsoft Excel

considered a new opportunity to create a vast amount of data to understand better other demographic factors and their effects on a large population. Still, there is an issue related to the standardization of medical notes worldwide, which play an essential role in a better understanding and data analysis. Big data analysis can improve health care delivery and reduce unnecessary costs. Big data can help better understand disease progression and the side effect of medications and ensure that health care delivers equally among the population. In addition to EMR data, integration with other devices such as home monitoring devices and smartphone applications may provide better insight into the population data.

Table 3 PRISMA 2019 checklist – Results

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Ristevski B, Chen M. Big Data Analytics in Medicine and Healthcare [49]	• To understand the big data • To find the challenges mining • To understand data privacy and security	Qualitative	Qualitative	One	1, China	A hardware innovations in processor technology, newer kinds of memories/network architecture will minimize the time spent in moving the data from storage to the processor in a distributed setting	Big data characterstics, big data analtyics, chanllanges in big data analytics and big data privacy and security	Big data analytics in medicine and healthcare is an auspicious process of integrating, exploring, and analysing large amounts of complex heterogeneous data: biomedical data, experimental data, electronic health records data and social media data. Furthermore, integrating such diverse data makes big data analytics intertwine several fields, such as bioinformatics, medical imaging, sensor informatics, medical informatics, health informatics and computational biomedicine
Gamache R, Kharrazi H, Weiner JP. Public and Population Health Informatics: The Bridging of Big Data to Benefit Communities [21]	To summarize the recent public and population health informatics literature with a focus on the synergistic "bridging" of electronic data to benefit communities and other populations	Search of the literature from July 1, 2016 to September 30, 2017	Systmatic review	One	3	• The newly emerging public health informatics vision and infrastructure • The alignment of informatics aims, goals, and outcomes across the oftentimes separate fields of public health and population health • The increased incorporation by both public and population health informatics professionals of SDH data	Several categories were observed in the review focusing on public health's socio-technical infrastructure: evaluation of surveillance practices, surveillance methods, interoperable health information infrastructure, mobile health, social media, and population health. Common trends discussing socio-technical infrastructure included big data platforms, social determinants of health, geographical information systems, novel data sources, and new visualization techniques. A common thread connected these categories of workforce, governance, and sustainability: using clinical resources and data to bridge public and population health	Both medical care providers and public health agencies are increasingly using informatics and big data tools to create and share digital information. The intent of this "bridging" is to identify proactively, monitor, and improve a range of medical, environmental, and social factors relevant to the health of communities. These efforts show a significant growth in a range of population health-centric information exchange and analytics activities

Table 3 (continued)

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Zhang X, Pérez-Stable EJ, Bourne PE, Peprah E, Duru OK, Breen N, Berrigan D, Wood F, Jackson JS, Wong DWS, Denny J. Big Data Science: Opportunities and Challenges to Address Minority Health and Health Disparities in the 21st Century [50]	• To find opportunities to Use Big Data Science to Advance Minority Health and Reduce Health Disparities • To understand Potential Challenges of Using Big Data for Minority Health and Health Disparities Research	Quantitaive		6	One	Ethics, Privacy, and Trust Missing Data and Statistical Uncertainty Data Access and Sharing Data Science Training and Workforce Diversity	This article focuses on three priority opportunities that Big Data science may offer to the reduction of health and health care disparities. One opportunity is to incorporate standardized information on demographic and social determinants in electronic health records in order to target ways to improve quality of care for the most disadvantaged populations over time. A second opportunity is to enhance public health surveillance by linking geographical variables and social determinants of health for geographically defined populations to clinical data and health outcomes. Third and most importantly, Big Data science may lead to a better understanding of the etiology of health disparities and understanding of minority health in order to guide intervention development	Specific recommendations include investing in data collection on small sample populations, building a diverse workforce pipeline for data science, actively seeking to reduce digital divides, developing novel ways to assure digital data privacy for small populations, and promoting widespread data sharing to benefit under-resourced minority-serving institutions and minority researchers. With deliberate efforts, Big Data presents a dramatic opportunity for reducing health disparities but without active engagement, it risks further widening them
Madanian S, Parry DT, Airehrour D, Cherrington M. mHealth and big-data integration: promises for healthcare system in India. BMJ Health Care Inform [51]	• Review healthcare challenges in India, their context and investigate its current status—with a specific focus on mHealth • Identify opportunities in Indian healthcare system for deploying mHealth and big-data • Propose a mHealth model and provide recommendations for its efficient use to improve healthcare and action policies by authorities	Systmatic review	A critical review was conducted using electronic sources between December 2018 and February 2019, limited to English language articles and reports published from 2010 onwards	One	One, India	This paper describes trending relationships in mHealth with big-data as well as the accessibility of national opportunities when specific barriers and constraints are overcome. The paper concentrates on the healthcare delivery problems faced by rural and low-income communities in India to illustrate more general aspects and identify key issues. A model is proposed that utilises generated data from mHealth devices for big-data analysis that could result in providing insights into the India population health status. The insights could be important for public health planning by the government towards reaching the Universal Health Coverage	This paper describes trending relationships in mHealth with big-data as well as the accessibility of national opportunities when specific barriers and constraints are overcome. The paper concentrates on the healthcare delivery problems faced by rural and low-income communities in India to illustrate more general aspects and identify key issues. A model is proposed that utilises generated data from mHealth devices for big-data analysis that could result in providing insights into the India population health status. The insights could be important for public health planning by the government towards reaching the Universal Health Coverage	Biomedical, behavioural and lifestyle data from individuals may enable customised and improved healthcare services to be delivered. The analysis of data from mHealth devices can reveal new knowledge to effectively and efficiently support national healthcare demands in less developed nations, without fully accessible healthcare systems

Table 3 (continued)

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, Boccia S. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives [31]	To find the initiatives utilize big data in health care system in EU countries To study the potential benefits of Big Data for healthcare in the European Union	Systmatic review	Review the EU supported initiatives concerning activities that involve the use of Big Data in public health in Europe from 2012 to 2018		12, European Countries	The potential benefits of Big Data for healthcare in the European Union Big Data in public health Ethical and legal issues for the effective use of Big Data in healthcare	The implementation of precision medicine remains contingent on significant data acquisition and timely analysis to determine the most appropriate basis on which to tailor health optimization for individual prevention, diagnosis and disease treatment. Achieving effective and proportionate governance of health-related data will be essential for the future healthcare systems, and it requires that stakeholders collaborate and adapt the design and performance of their systems to reach the maximum innovative potential of information and innovation technology on health in the EU In this context, EU Member States should agree on international technical standards, taking also into account openness that is considered as the basic paradigm for digital transformation. Additionally, new approaches must be found for translating the vast amount of data into meaningful information that healthcare professionals can use. Further efforts must be made to make information for doctors and health professionals more accessible and understandable To achieve this, existing training and education programmes for healthcare professionals should integrate the issues of data handling in the curricula to ensure the development of the necessary skills and competencies. This is one of the objectives of the 'European network staff eXchange for integrAting precision health in the health Care sysTems' consortium (ExACT)20 project that aims to integrate precision health in European health systems by training a new generation of healthcare professionals across and outside of the EU	we are living in fast-moving times, not least in terms of healthcare innovation. Whilst there are pressing needs for more personalized and sustainable health services, science and technology are offering a host of potentially invaluable new tools to deliver them. A cooperation at the EU level is needed to facilitate investments both in new technology and in the human capital, in order to guide citizens into this new frontier of human health and well-being where data are becoming a significant corporate asset, a vital economic input and the foundation of new business models

Table 3 (continued)

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Adibuzzaman M, DeLaurentis P, Hill J, Benneyworth BD. Big data in healthcare - the promises, challenges and opportunities from a research perspective: A case study with a model database [1]	• Identify and present the initial work addressing the relevant challenges of utilizing big data in clinical care, in three broad categories: data, accessibility, and translation	Qualitative	A case study with a model database		2		• Identify tow initiatives of Medical Information Mart for Intensive Care (MIMIC III) and "Informatics for Integrating Biology and the Bedside (i2b2)" where big data and the electronic medical record used to find a clinical outcome or conduct clinical research • Identify the challenges of big data utilization in three broad categories: data, accessibility, and translation	The big data is considered as the future promise to enhance the health care system but still the available software system, polices doesn't support the improvement in these areas. The study focused on data, accessibility and translation. To make data more understandable and valuable. We need a larger cohort of institutions to share complete, precise, and time-stamped data as well as with greater willingness to invest in technologies for de-identifying private patient data for it to be shared broadly for scientific research. At the same time, as more and more "big data" systems are developed, the scientific and regulatory communities need to figure out new ways of understanding causal relationship from data captured during routine health care, that would complement current gold standard methods such as RCTs as well as identify the relationship between clinical practice and outcomes, as there is a wide disparity in the quality of care across the country
Murphy S, Castro V, Mandl K. Grappling with the Future Use of Big Data for Translational Medicine and Clinical Care [52]	• Identify the reasons for the lack of integration of the big data by most the electronic medical record system • Identify the area of extracting the big data integration into clinical care				One, USA		• Data are often positioned outside the EMRS with their own distinct semantic and technical structures. Integrating big data into clinical systems requires web services organized around a common information model to reach out to the Big Data Repositories as they exist in cloud-native infrastructures • Harassing of big data can be used in phenotyping, hypothesis generation, predication and clinical decision support • Features that are important to healthcare often need to be computed from, rather than being readily available from, the Big Data	• Apps that will run inside and outside the EMRS can provide support for complex decisions and workflows that involve genomics, imaging, and personal health repositories. Accurate phenotyping can become a routine part of clinical care. An infrastructure based on the SMART API will allow an ecosystem of apps to be shared across healthcare institutions using web service standards such as FHIR. This provides a vision for a new type of EMRS, which is expected to play out over the next five to ten years

Table 3 (continued)

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Roca J, Tenyi A, Cano I. Paradigm changes for diagnosis: using big data for prediction. [53]	(i) Regional deployment of Integrated Care services for chronic patients with a personalized medicine approach; (ii) Development of a test-bed, willing for international leadership, for the use of ICT in novel services that generate value in the healthcare system of Catalonia; and, (iii) Development and monetization of novel products and services with a high level of transferability to other healthcare systems, contributing to strengthen Catalan industrial competences		Case study		One, Spain	ACSV	• The study describes how to implement innovation and big data analysis in managing chronic disease, a case of 'chronic pulmonary obstructive disease' prescribed as an example. Then the concept of "Integrated care" utilizing a different source of data stored in cloud platform to analyse population conditions and predicted using modelling method in the outcome • The study illustrated four challenges: Enhanced clinical predictive modelling, Technology related and security and data storing issue and application of evaluation and adoption of decision support system (DSS)	The cloud-based data analytics platform has been proposed to successfully address the implicated potentials of health risk assessment and stratification and to facilitate large-scale adoption of Integrated Care of chronic patients [17, 39], contributing to enhance healthcare outcomes and patient experience of care while reducing costs and improving the health of populations. Applying holistic strategies for subject-specific risk prediction and stratification, that consider multilevel covariates influencing patient health, would increase the predictive accuracy and facilitate clinical decision-making based on sound estimates of individual prognosis

Table 3 (continued)

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Thompson ME, Dulin MF. Leveraging Data Analytics to Advance Personal, Population, and System Health: Moving Beyond Merely Capturing Services Provided [54]	Explore the facets of the promise, perils, trends, and trajectories of health informatics and analytics Explore the causes of underutilizing big data in health care Identify the role and opportunity of introducing the new policy (North Carolina's new modernized health information exchange policy)		Qualitiaive study	One	One		The causes of underutilization of big data in North Carolina categorized into 5 points: • the Health Information Technology Act and the Affordable Care Act (ACA), he payment-for-services side of the process is well developed and functional. The shift to value-based payments has yet to materialize. limit investment to utilize EMR data • Data governance is lacking in consistency and rigor, within and across organizations. As Shannon Fuller notes in his article, without assurances of data quality, integrity, and security, any conclusions drawn from analyses are suspect at best • uniform standards for capturing and reporting the largely unstructured data found in medical records are absent. Aside from a few high-level data elements, the interoperability and easy aggregation and pooling of analytic data promised by the federal Health Information Exchange (HIE) are undeveloped • The sharing, linking, aggregating, and disaggregating of patient, system, and community data, while now technically feasible, are simultaneously required and prohibited by a variety of conflicting laws and policies • The processes needed to link and analyze these disparate data, and make them accessible to patients, providers, payers, and researchers, also create the risks (and liabilities) of breach and abuse	Big data and analytics hold much promise, but are under-developed in health care when compared to other sectors. Given the many pitfalls, security concerns, and risks engendered by our lack of a universal coverage/single risk pool system, perhaps we are fortunate to have the added breathing space to more fully develop our governance structures, ensure data quality and security, and align our policies before charging ahead into the brave new electronically integrated world. Building systems and networks based upon trust and transparency among health care's many stakeholders will take time and effort outside the scope of the technology that makes it possible. The move toward valued-based payments will reinforce the need for health care industries to adapt to a systems-thinking-driven model and will further foment a culture of data-driven learning organizations

Table 3 (continued)

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Carney TJ, Kong AY. Leveraging health informatics to foster a smart systems response to health disparities and health equity challenges [55]	• Identified health disparities and health equity • Identify big data challenges in health disparities and health equity • Identify the domains of health informatics and how to apply smart data analysis within each domains		Qualitiaive study	One	one,USA		• Define the health disparities and health equality and the value to apply this in country to understand health related issues • Identify four groups to apply smart health and big data which are: Public health, population health, community health and consumer health informatics • Identify the challenges and available resources for each domains and what required to improve health informatics in each domains	• This presentation of collective intelligence and the corresponding terms of smart health, knowledge ecosystem, enhanced health disparities informatics capacities, knowledge exchange, big-data, and situational awareness are a means of demonstrating the complex challenges informatics professional face in trying to model, measure, and manage an intelligent and smart systems-level response to health disparities. Study outlined public and population health disparities challenges across four distinct domains of informatics. Study also introduced a concept of performance-based health disparities that we operationally define in our models as those that are generated or triggered by breaches or defects in the continuum of care that may negatively influence the quality, cost, safety, efficiency, or effectiveness of population health and health promotion

Table 3 (continued)

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, Bernal-Delgado E, Blomberg N, Bock C, Conesa A, Del Signore S, Delogne C, Devilee P, Di Meglio A, Eijkemans M, Flicek P, Graf N, Grimm V, Guchelaar HJ, Guo YK, Gut IG, Hanbury A, Hanif S, Hilgers RD, Honrado Á, Hose DR, Houwing-Duistermaat J, Hubbard T, Janacek SH, Karanikas H, Kievits T, Kohler M, Kremer A, Lanfear J, Lengauer T, Maes E, Meert T, Müller W, Nickel D, Oledzki P, Pedersen B, Petkovic M, Pliakos K, Rattray M, I Màs JR, Schneider R, Sengstag T, Serra-Picamal X, Spek W, Vaas LA, van Batenburg O, Vandelaer M, Varnai P, Villoslada P, Vizcaíno JA, Wubbe JP, Zanetti G. Making sense of big data in health research: Towards an EU action plan [9]	Addressing big data barriers in European countries To find the opportunities to creating the European Single Market for health, which will improve health and healthcare for all Europeans	Quanatative	Health Directorate of the Directorate-General for Research and Innovation at the European Commission (EC), the executive body of the EU, organized in Luxembourg a workshop entitled "Big data in health research: an EU action plan	6	Multiple, European countries	The potential benefits of big data for healthcare The challenges ahead for the effective use of big data in healthcare Data quality, acquisition, curation, and visualization Legal and regulatory aspects	Recommendations for an EU action plan: Launch pilot projects on the application of big data to inform health Leverage the potential of open and citizen science for the exploitation of big data in health Catalyze the involvement of all relevant stakeholders in projects	The digital revolution is underway. A number of industries have already transformed their activities or have now become inoperative. The driving forces are miniaturization, automation, and now increasingly the convergence of artificial intelligence, deep learning, and robotics. Healthcare will not escape these developments. In fact, big data as a driving force will play an even more important role than in most industries. In Europe, working across borders is the only way to master the challenges of this scientific, technological, and industrial revolution. The single most important factor is the workforce. Countries that are ahead in ICT competence and have an understanding of cultural differences and an ability and willingness to work together have the best chance to succeed

Table 3 (continued)

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Beckmann JS, Lew D. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. [56]	To define the challenges and opportunities for achieving clinical utility in precision medicine To find the opportunity of evidence-based precise medicine	Qualitative				Multiplicity of stakeholders and disciplines •Analyses of big data •Heterogeneity of complex, multilayered data types, and formats •Harmonization of data semantics (clinical, laboratory, and others): vocabularies, terminologies, classification and coding systems, ontologies •Standardization of data entry and storage •Integration of multiple data types (such as laboratory, clinical, behavioral, lifestyle, environmental) •Secure, sustainable, and effective data storage and sharing •Necessity for new analytic tools and algorithms •Multiplicity and lack of semantic and technical interoperability of electronic health record systems •Extremely dynamic and fast-changing field, with new tools constantly emerging •Training and education of the different stakeholders (medical staff, patients, and decision-makers) •Ethical, legal, social, and consent issues •Uberization of medicine	The role of clinical bioinformatics in precision medicine Aggregation of heterogeneous data sets into electronic health records Reconciling evidence-based medicine and precision medicine Citizen-centered medicine The economics of precision medicine Challenges of evidence-based precision medicine	We consider that evidence-based precision medicine rests on three pillars: (i) responsible inter-institutional sharing of large clinical and laboratory interoperable, harmonized data sets; (ii) data on vital signs and behavior collected by empowered citizens; and (iii) clinical bioinformatics required to convert this complex information into clinically useful knowledge, which will be returned by the medical practitioners to the individuals concerned.. The net outcome could be better clinical diagnosis or prognostication; this could facilitate clinical decision-making, improve medical care or treatment, and most importantly, could contribute to disease delay or even prevention. Clinical bioinformatics has a central role to play in this revolutionary person-centric effort by contributing to care delivery innovations and improved health preservation, and shifting the emphasis more and more from therapy to prevention, and from disease to wellness
Kumar S, Singh M. Big data analytics for healthcare industry: impact, applications, and tools. Big Data Mining and Analytics. 2018 Oct 15;2(1):48–57. [57]	To analyse Big data analytics for healthcare industry: impact, applications, and tools	Qualitative	Observation based study			The health industry sector has been confronted by the need to manage the big data being produced by various sources, which are well known for producing high volumes of heterogeneous data. Various big-data analytics tools and techniques have been developed for handling these massive amounts of data, in the healthcare sector		

Table 3 (continued)

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Balsam Alkouz, Zaher Al Aghbari, Jemal Hussien Abawajy. Tweetluenza: Predicting Flu Trends from Twitter Data. Big Data Mining and Anyalytics 2019, 2(4): 273–287. [58]	The authors develop a new Influenza prevalence prediction model, called Tweetluenza, to predict the spread of the Influenza in real time using cross-lingual data harvested from Twitter data streams with emphases on the United Arab Emirates (UAE). Based on the features of tweets, Tweetluenza filters the Influenza tweets, utilizing big data mining from social media account	Quantitative	To monitor the growth of Influenza, the reporting tweets were employed. Furthermore, a linear regression model leverages the reporting tweets to predict the Influenza-related hospital visits in the future. We evaluated Tweetluenza empirically to study its feasibility and compared the results with the actual hospital visits recorded by the UAE Ministry of Health		UAE	The experiments demonstrate the practicality of Tweetluenza, which was verified by the high correlation between the Influenza-related Twitter data and hospital visits due to Influenza. Furthermore, the evaluation of the analysis and prediction of Influenza shows that combining English and Arabic tweets improves the correlation results	To investigate the impact of Twitter data on the prediction of Influenza.ediction. To show that the combination of tweet counts and hospital visits counts improves the prediction of future hospital visits, we repeated the second experiment twice to predict the number of hospital visits: once using only the counts of tweets and once using only the counts of hospital visits. From Table 20, we note that the prediction by using either tweet counts only or the hospital visits counts only produces higher RMSE error than using them combined as compared to the result of	In this paper, author proposed Tweetluenza system that uses Twitter streams for Influenza surveillance and forecasting in a cross-lingual and cross-dialec. This show example the utlization of social media into predict health care system utlizing the big data infromation form the social media

Table 3 (continued)

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Gravili G, Manta F, Cristofaro CL, Reina R, Toma P. Value that matters: intellectual capital and big data to assess performance in healthcare. An empirical analysis on the European context. Journal of Intellectual Capital. 2020 Jul 30. [59]	The aim of this paper is to analyze and measure the effects of intellectual capital (IC), i.e. human capital (HC), relational capital (RC) and structural capital (SC), on healthcare industry organizational performance and understanding the role of data analytics and big data (BD) in healthcare value creation	Qualitative	The study has a twofold approach: in the first part, the authors operated a systematic review of the academic literature aiming to enquire the relationship between IC, big data analytics (BDA) and healthcare system, which were also the descriptors employed. In the second part, the authors built an econometric model analyzed through panel data analysis, studying the relationship between IC, namely human, relational and structural capital indicators, and the performance of healthcare system in terms of performance. The study has been conducted on a sample of 28 European countries, notwithstanding the belonging to specific international or supranational bodies, between 2011 and 2016		Multiple	The relationship between IC indicators and performance could be employed in other sectors, disseminating new approaches in academic research. Through the establishment of a relationship between IC factors and performance, the authors implemented an approach in which healthcare organizations are active participants in their economic and social value creation	The paper proposes a data-driven model that presents new approach to IC assessment, extendable to other economic sectors beyond healthcare. It shows the existence of a positive impact (turning into a mathematical inverse relationship) of the human, relational and structural capital on the performance indicator, while the physical assets (i.e. the available beds in hospitals on total population) positively mediates the relationship, turning into a negative impact of non-IC related inputs on healthcare performance. The result is relevant in terms of managerial implications, enhancing the opportunity to highlight the crucial role of IC in the healthcare sector	The authors provide a new holistic framework on the relationship between IC, BDA and organizational performance in healthcare organizations through a systematic review approach and an empirical panel analysis at a multinational level, which is quite a novelty regarding the healthcare

Table 3 (continued)

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Gu D, Li J, Li X, Liang C. Visualizing the knowledge structure and evolution of big data research in healthcare informatics. International journal of medical informatics. 2017 Feb 1;98:22–32. [60]	To explore the foundational knowledge and research hotspots of big data research in the field of healthcare informatics	Qualitative	A series of bibliometric analyses on the related literature, including papers' production trends in the field and the trend of each paper's co-author number, the distribution of core institutions and countries, the core literature distribution, the related information of prolific authors and innovation paths in the field, a keyword co-occurrence analysis		3	By conducting a literature content analysis and structure analysis, we found the following: (a) In the early stage, researchers from the United States, the People's Republic of China, the United Kingdom, and Germany made the most contributions to the literature associated with healthcare big data research and the innovation path in this field. (b) The innovation path in healthcare big data consists of three stages: the disease early detection, diagnosis, treatment, and prognosis phase, the life and health promotion phase, and the nursing phase. (c) Research hotspots are mainly concentrated in three dimensions: the disease dimension (e.g., epidemiology, breast cancer, obesity, and diabetes), the technical dimension (e.g., data mining and machine learning), and the health service dimension (e.g., customized service and elderly nursing)		This study will provide scholars in the healthcare informatics community with panoramic knowledge of healthcare big data research, as well as research hotspots and future research directions

Table 3 (continued)

Author and Title (Vancouver style)	Objective (s)	Research Approach	Study Design and Methods	# of Policies	# of Countries	Outcome Measure(s)/Variable (s)	Findings	Conclusions
Dhagarra D, Goswami M, Sarma PR, Choudhury A. Big Data and blockchain supported conceptual model for enhanced healthcare coverage. Business Process Management Journal. 2019 Oct 14. [61]	Despite recent progress in ensuring improved access to health care in past decade or so, disparities across gender, geography and socioeconomic status continue to persist. Fragmented and scattered health records and lack of integration are some of the primary causes leading to uneven healthcare service delivery. The devised framework is intended to address these challenges. The paper aims to discuss these issues	Qualitative	In this research a Big Data and blockchain anchored integrative healthcare framework is proposed focusing upon providing timely and appropriate healthcare services to every citizen of the country. The framework uses unique identification number (UID) system as formalized and implemented by the Government of India for identification of the patients, their specific case histories and so forth		1	A key component of our evolved framework is the Big Data analytics-based framework that seeks to provide structured health data to concerned stakeholders in healthcare services. The model entails all pertinent stakeholders starting from patients to healthcare service providers	The key characteristic of our proposed framework is that it provides easy access to secure, immutable and comprehensive medical records of patients across all treatment centers within the country. The model also ensures security and privacy of the medical records based upon the incorporation of biometric authentication by the patients for access of their records to healthcare providers	

Big data can improve public health surveillance and address disparities in the health care system. In the USA, the implementation of the Affordable Care Act and general insurance play an essential role in expanding the health care system and enhancing the accessibility of health care. The factors as mentioned above, help acquire extensive data from the population across different socioeconomic backgrounds to better understand additional population requirements and health status and formulate policy and strategy based on population requirements. This approach will help allocate resources and enhance the health care system by utilizing big data analysis.

Big data opportunities in health care

RQ2: What are the potential opportunities to enhance quality-of-care services through integrating big data in the health care system?

Geographic information systems (GISs) can help better understand the population at risk and health requirements and make the proper intervention based on real-time data through big data analysis. Another example Gamache provides the GIS to allocate vaccines targeting a particular population in response to an outbreak in a specific geographic area [21].

Social media data can be essential in understanding and monitoring population behavior and spreading infectious diseases. For example, Young et al. analyzed 553,186,061 tweets and found a correlation between the prevalence of HIV and the geographic location of HIV-related tweets [46]. This can be expanded by knowing and analyzing the current status and utilizing the data for future prediction of cases using big data modeling, analysis of population behavior, and linkage to social media data. Future projections can help make the proper intervention to reduce the spread of infections [47, 56, 58], reducing the cost of the disease burden on the community and health care system.

Zeng mentioned that big data modeling could be more accurate than traditional methods. In addition, big data modeling can include predictive modeling to forecast a disease occurrence and complications related to the particular disease [47], which can help in a better intervention, enhance the health outcome, and reduce the disease progression.

Challenges related to big data

RQ3: How to best understand the challenges of implementing and using big data technologies?

Many challenges are identified in the literature concerning the application of big data in the health care system. For example, the study performed by Auffray et al. [9] and Zeng [47] highlighted some of the key challenges concerning the development of utilizing big data. The most common areas of concern were data privacy [50], data storage [49, 56], data structure [57], data ownership, and governance [21].

Data security

Health care data are considered sensitive, and most literature agrees that big data raises a security issue. As several USA researchers mentioned, it is regarded as a challenge to process big data. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) prevents client data handling without prior consent. The same applies to UAE Article 379 of the UAE Penal Code, which requires prior permission to manage client's

data. A simple de-identification process was not the final solution, as Adibuzzaman et al. [1]. The De-identification still cannot protect client data within the ID process.

Further, it is easy to re-identify the person by other location or demographic information. Data related to health issues are considered highly sensitive and private [55]. This is why specific regulation regarding accessibility and the availability of the data with consideration to the customer is required and prior consent mandatory, even if the data were to be coded and de-identified.

Data storage

Data storage has been a concern in many studies due to data security. The legal issue within different countries and accessibility of data are limited to research and governments and is without a clear data storage and accessibility policy [51–54]. Big data cannot be stored by ordinary means, especially from different sources, including electronic medical records, monitoring devices, images, and lab tests.

Initiatives and solutions for data storage were discussed by Adibuzzaman et al. [1], including the platform Informatics for Integrating Biology and the Bedside (i2b2). It is a platform of more than 100 hospitals where patient data is de-identified and stored for research purposes since the hospital should use another software to transfer data. The author argues that this initiative does not permit patients to access their data [1]. Moreover, the hospital required much effort to de-identify the data and used a particular platform to transfer data. If the same system is to be applied, an additional budget is required, and the system will be limited to structured data. It is also mentioned by Dash et al. that data stored at the same time it is generated is less compared to if the data is transferred using another system [17]. Most of the evaluated authors agreed that structured data storage and analysis is more manageable than transferal to another platform. Cloud-based storage remains the ideal solution, yet the security issue remains a challenge to overcome.

Missing data and unstructured data

One of the common challenges in the literature review is missing data in the electronic health record, unstructured data, and free-text data, which is very difficult to process and analyze. For example, the study conducted by Zeng mentioned electronic health records lack social or behavioral data and that there was no standardization of the data format, leading to disparities in health information [47]. This can prevent or make big data analysis impossible. In addition, the lack of data standardization makes data transfer and acquisition impossible [23, 33, 50].

Most electronic health records are designed to make the diagnosis coding and billing more manageable and the information more explicit. Still, these are not yet advanced enough to make the analysis and data helpful linkage, as mentioned by Gamache [21]. One of the examples of the failure of EMR for data analysis is Medical Information Mart for Intensive Care (MIMIC III) which has collected data for more than 50,000 patients from Beth Israel Deaconess Hospital dating back to 2001. The researcher aimed to conduct studies to answer different questions, such as the drug–drug interaction between antihistamine

and antidepression. When they applied the selection criteria and checked the files, they had a minimal sample size, which is not representative. This issue can also apply to EMRs in the UAE. There are multiple EMR systems in the UAE with different software, and there is no standardization of medical record notes and fragmented systems between local and federal authorities [33]. This will make data transfer and analysis complicated, which should be considered.

Data ownership and governance

Health care information is considered sensitive data. Data ownership and sharing is an unclear and negotiable challenge among countries and has been mentioned by many studies [8, 9, 33]. The privacy policy across Europe varies, and there is no approach to big data sharing that can fit the existing policies in other countries. The EC for general data protection regulation (2012/0011COD) tried to synchronize the fragmented health care system to make the data available and useable among different European countries. One suggestion was to share data across a blockchain where all the transactions would be recorded and the data accessibility would be monitored; however, this is not yet adopted as security was a concern. This would not be the case in the UAE as the policy, and federal law is unified, and data control and accessibility could be maintained at the federal level.

Auffray et al. [9] mentioned that the USA is a "patient-driven economy" where patients own their data. This is a step forward and a promising approach for a fragmented health care system where different healthcare systems are present, as in the case of Europe and even the UAE. This would help the patient own their digital data, but it requires a digital infrastructure and storage system to ensure the data is transferred into a cloud. It also requires that it is in a structured format to access, understand, and easily analyze the data [59–61]. Nevertheless, this approach can help enhance health tourism and a health-driven economy.

Adibuzzaman et al. [1] mentioned data should be Findable, Accessible, Interoperable, and Reproducible (FAIR). In addition, the data would need to be stored as open source, where researchers, stakeholders, and even patients have access to those data while ensuring data protection and privacy. For example, data storage in a protected environment after properly de-identifying client's ID to maintain a privacy law.

Conclusion and future recommendations

The UAE government remains aware of the power of big data, as shown by the establishment of the UAE Strategy for Artificial Intelligence. The Dubai government policy framework is intended to develop and implement a culture of data sharing and evidence-based decision-making in Dubai [38]. The study concludes that the healthcare systems in the UAE can be enhanced through big data; however, the authorities within the UAE must acknowledge that the development of efficient frameworks for the performance and quality assessment of the new health care system is significant. The said goal can be achieved via integrating big data and health informatics with the help of IT specialists, health care managers, and stakeholders.

Recommendations to use big data in the health sector in the UAE

Specific recommendations for big data handling in the health care sector

- Formulate a unified EMR standardization for the medical note to be able to process medical note data and transfer it quickly.
- Incentivization of the health care provider to ensure they are following high standardization of EMR, as data entry is significant, and this was a significant challenge in previous studies.
- Public–private partnership is essential, and the private sector should incentivize to share their data. The UAE has the big challenge of population diversity, and most of the population is not seeking medical care in the government health care sectors. To overcome population discrepancy and ensure the data represents the whole population, data from the private sector should be accessible as well.
- Agree on data needed for real-time monitoring, such as infectious diseases or surveillance that requires intervention. The big data analysis will simplify and quicken the intervention decision, consequently offering a better response from authorities to an emergency.
- Utilize big data to take a proactive measurement and to ensure timely involvement of stakeholders to prevent a disease occurring in the case of non-communicable diseases. That will ultimately reduce the management cost and the burden of chronic diseases.

Conclusion and future research

The digital revolution has arrived, and it is impacting everyday life. There is a dire need to utilize existing health care system-related data, which is automatically generated daily. This big data can transform the healthcare system to improve patient care, proactively envisaging disease origins and implementing timely solutions to bridge the gaps in the existing healthcare system. It is time to think about innovative, technological solutions to link and analyze data faster, understand the disparities between community health and public health alongside enhancing the overall health care system, and implement United Nations SDG Goal 3, ensuring healthy lives and promoting well-being for all ages.

Author contributions

KH and IAM: Made a substantial contribution to all the sections and participated in the review, analysis, and interpretation. Involved in drafting the manuscript and revising it critically for important intellectual content. KH: Made a substantial contribution to study design. Participated in review, analysis, and interpretation. KH: Made a considerable contribution to background and method sections. KH: Made a significant contribution to the background and discussion sections. KH and IAM: Made a significant contribution to the discussion and conclusion sections. IAM: Made a substantial contribution to background, methods, and discussion sections. All authors give final approval for the version to be published and agree to be accountable for all aspects of the work. Both authors read an approved the final manuscript.

Funding

The authors disclose receipt of the following financial support for the authorship, research, and/or publication of this article: The authors would like to acknowledge Mohammed Bin Rashid School of Government (MBRSG), Dubai, UAE, and the Alliance for Health Policy and Systems Research at the World Health Organization for financial support as part of the Knowledge to Policy (K2P) Center Mentorship Program [BIRD Project].

Data availability

The authors confirm that the data supporting the findings of this study are available within the article via its supplementary materials.

Declarations

Ethics approval and consent to participate
Studies involving animal subjects: No animal studies are presented in this manuscript.
Studies involving human subjects: No human studies are presented in this manuscript.
Inclusion of identifiable human data: No potentially identifiable human images or data is presented in this study.

Competing interests
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 April 2022 Accepted: 25 December 2022
Published online: 01 February 2023

References

1. Adibuzzaman M, DeLaurentis P, Hill J, Benneyworth BD. Big data in healthcare–the promises, challenges, and opportunities from a research perspective: A case study with a model database. In: AMIA Annual Symposium Proceedings (Vol. 2017). American Medical Informatics Association. 2017, p. 384.
2. Adnan K, Akbar R. Limitations of information extraction methods and techniques for heterogeneous unstructured big data. Int J Eng Business Manag. 2019. https://doi.org/10.1177/1847979019890771.
3. Ahmadvand H, Dargahi T, Foroutan F, Okorie P, Esposito F. Big data processing at the edge with data skew aware resource allocation. In: 2021 IEEE conference on network function virtualization and software-defined networks (NFV-SDN) (pp. 81–86). IEEE. 2021.
4. Ahmadvand H, Foroutan F, Fathy M. DV-DVFS: merging data variety and DVFS technique to manage the energy consumption of big data processing. J Big Data. 2021;8(1):1–16.
5. Ahmadvand H, Goudarzi M. Using data variety for efficient progressive big data processing in warehouse-scale computers. IEEE Comput Archit Lett. 2016;16(2):166–9.
6. Ahmadvand H, Goudarzi M. SAIR: significance-aware approach to improve QoR of big data processing in case of budget constraint. J Supercomput. 2019;75(9):5760–81.
7. Ahmadvand H, Goudarzi M, Foroutan F. Gapprox: using gallup approach for approximation in big data processing. J Big Data. 2019;6(1):1–24.
8. AlMarzooqi FM, Moonesar IA, AlQutob R. Healthcare professional and user perceptions of eHealth data and record privacy in Dubai. Information. 2020;11(9):415.
9. Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, Bernal-Delgado E, Blomberg N, Bock C, Conesa A, Del Signore S. Making sense of big data in health research: towards an EU action plan. Genome Med. 2016;8(1):1–13.
10. Balador A, Bazzi A, Hernandez-Jayo U, de la Iglesia I, Ahmadvand H. A survey on vehicular communication for cooperative truck platooning application. Vehicular Commun. 2022;34:100460.
11. Bani-issa W, Eldeirawi K, Al Tawil H. Perspectives on the attitudes of healthcare professionals toward diabetes in community health settings in United Arab Emirates. J Diabetes Mellitus. 2014;5(01):1.
12. Barrett MA, Humblet O, Hiatt RA, Adler NE. Big data and disease prevention: from quantified self to quantified communities. Big data. 2013;1(3):168–75.
13. Blumenstock J, Cadamuro G, On R. Predicting poverty and wealth from mobile phone metadata. Science. 2015;350(6264):1073–6.
14. Catalyst NEJM. Healthcare big data and the promise of value-based care. NEJM Catalyst. 2018;4(1):89.
15. Coorevits P, Sundgren M, Klein GO, Bahr A, Claerhout B, Daniel C, Dugas M, Dupont D, Schmidt A, Singleton P, De Moor G. Electronic health records: new opportunities for clinical research. J Intern Med. 2013;274(6):547–60.
16. Dargahi T, Ahmadvand H, Alraja MN, Yu CM. Integration of blockchain with connected and autonomous vehicles: vision and challenge. ACM J Data Inform Quality (JDIQ). 2021;14(1):1–10.
17. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data. 2019;6(1):1–25.
18. Demchenko Y, Zhao Z, Grosso P, Wibisono A, De Laat C. Addressing big data challenges for scientific data infrastructure. In: 4th IEEE International Conference on Cloud Computing Technology and Science Proceedings (pp. 614–617). IEEE. 2012.
19. Effoe VS, Katula JA, Kirk J, et al. The use of electronic medical records for recruitment in clinical trials: findings from the Lifestyle Intervention for Treatment of Diabetes trial. Trials. 2016;17:496. https://doi.org/10.1186/s13063-016-1631-7.
20. Endriyas M, Alano A, Mekonnen E, Ayele S, Kelaye T, Shiferaw M, Misganaw T, Samuel T, Hailemariam T, Hailu S. Understanding performance data: health management information system data accuracy in Southern Nations Nationalities and People's Region. Ethiopia BMC Health Serv Res. 2019;19(1):1–6.
21. Gamache RK. Public and population health informatics: the bridging of big data to benefit communities. Yearb Med Inform. 2018;27(1):199.
22. Gao Y, Zhou Y, Zhou B, Shi L, Zhang J. Handling data skew in MapReduce cluster by using partition tuning. J Healthcare Eng. 2017;2017:56.
23. Goldstein BA, Navar AM, Pencina MJ, Ioannidis J. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc. 2017;24(1):198–208.
24. Laney D. 3D data management: Controlling data volume, velocity and variety. META Group Research Note. 2001;6(70):1.
25. Leyens L, Reumann M, Malats N, Brand A. Use of big data for drug development and for public and personal health and care. Genet Epidemiol. 2017;41(1):51–60.

26. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 2009;62(10):e1-34.
27. Mittelstadt BD, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. Sci Eng Ethics. 2016;22(2):303–41.
28. Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009;151(4):264–9.
29. Muni Kumar N, Manjula R. Role of Big data analytics in rural health care-A step towards svasth bharath. Int J Computer Sci Inform Technol. 2014;5(6):7172–8.
30. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ. 2021;372:67.
31. Pastorino R, De Vito C, Migliara G, Glocker K, Binenbaum I, Ricciardi W, Boccia S. Benefits and challenges of Big Data in healthcare: an overview of the European initiatives. Eur J Public Health. 2019;29(3):23–7.
32. Paxton C, Niculescu-Mizil A, Saria S. Developing predictive models using electronic medical records: challenges and pitfalls. AMIA … Annual Symposium proceedings. AMIA Symposium. 2013;2013:1109–15.
33. Sarabdeen J, Moonesar IA. Privacy protection laws and public perception of data privacy: The case of Dubai e-health care services. Benchmarking Int J. 2018;34:8.
34. Shelton T, Poorthuis A, Graham M, Zook M. Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of 'big data.' Geoforum. 2014;52:167–79.
35. Srikanth Thudumu PB. A comprehensive survey of anomaly detection techniques for high dimensional big data. J Big Data. 2020;7:42. https://doi.org/10.1186/s40537-020-00320-x.
36. Steele JE, Sundsøy PR, Pezzulo C, Alegana VA, Bird TJ, Blumenstock J, Bjelland J, Engø-Monsen K, De Montjoye YA, Iqbal AM, Hadiuzzaman KN. Mapping poverty using mobile phone and satellite data. J R Soc Interface. 2017;14(127):20160690.
37. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MD, Horsley T, Weeks L, Hempel S. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7):467–73.
38. UAE. U.ae. Retrieved 12 11, 2020. https://u.ae/en/about-the-uae/strategies-initiatives-and-awards/local-governments-strategies-and-plans/dubai-data-strategy. 2019.
39. UAE GOVERNMENT. Vision 2021 and health. Retrieved April 05, 2021, from https://u.ae/en/information-and-services/health-and-fitness/vision-2021-and-health. 2020.
40. UAE. The United Arab Emirates Government portal. https://u.ae/en/information-and-services/g2g-services/open-government-data. 2022.
41. Ukil A, Bandyoapdhyay S, Puri, C, Pal A. IoT healthcare analytics: The importance of anomaly detection. In: 2016 IEEE 30th international conference on advanced information networking and applications (AINA) (pp. 994–997). IEEE, 2016.
42. United Nations. SDG, The Sustainable Development Goals, Retrieved April 05, 2020, https://sdgs.un.org/goals. 2015.
43. United Nations. UN World Data Forum 2018 wraps up with launch of Dubai Declaration. Retrieved April 05, 2020. https://www.un.org/sustainabledevelopment/blog/2018/10/un-world-data-forum-2018-wraps-up-with-launch-of-dubai-declaration/. 2018.
44. Wang H. Towards felicitous decision making: an overview on challenges and trends of Big Data. Inf Sci. 2016;367–368:747–65.
45. Wu P, Lin. Unstructured big data analytics for retrieving e-commerce logistics knowledge. Telematics Inform. 2018;35(1):237–44.
46. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. Prev Med. 2014;63:112–5. https://doi.org/10.1016/j.ypmed.2014.01.024.
47. Zeng X, Zhang Y, Kwong JS, Zhang C, Li S, Sun F, Niu Y, Du L. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. J Evid Based Med. 2015;8(1):2–10.
48. Zhang Q, Hansen D. Approximate processing for medical record linking and multidatabase analysis. Int J Healthcare Inform Syst Inform (IJHISI). 2007;2(4):59–72.
49. Zhang X, Pérez-Stable EJ, Bourne PE, Peprah E, Duru OK, Breen N, Berrigan D, Wood F, Jackson JS, Wong DW, Denny J. Big data science: opportunities and challenges to address minority health and health disparities in the 21st century. Ethn Dis. 2017;27(2):95.
50. Ristevski B, Chen M. Big data analytics in medicine and healthcare. J Integr Bioinform. 2018;15:3.
51. Madanian S, Parry DT, Airehrour D, Cherrington M. mHealth and big-data integration: promises for healthcare system in India. BMJ Health Care Inform. 2019;26:1.
52. Murphy S, Castro V, Mandl K. Grappling with the future use of big data for translational medicine and clinical care. Yearb Med Inform. 2017;26(01):96–102.
53. Roca J, Tenyi A, Cano I. Paradigm changes for diagnosis: using big data for prediction. Clin Chem Lab Med (CCLM). 2019;57(3):317–27.
54. Thompson ME, Dulin MF. Leveraging data analytics to advance personal, population, and system health: Moving beyond merely capturing services provided. N C Med J. 2019;80(4):214–8.
55. Carney TJ, Kong AY. Leveraging health informatics to foster a smart systems response to health disparities and health equity challenges. J Biomed Inform. 2017;1(68):184–9.
56. Beckmann JS, Lew D. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. Genome Med. 2016;8(1):1–1.
57. Kumar S, Singh M. Big data analytics for healthcare industry: impact, applications, and tools. Big Data Mining Analyt. 2018;2(1):48–57.

58. Alkouz B, Al Aghbari Z, Abawajy JH. Tweetluenza: Predicting flu trends from twitter data. Big Data Mining Analyt. 2019;2(4):273–87.
59. Gravili G, Manta F, Cristofaro CL, Reina R, Toma P. Value that matters: intellectual capital and big data to assess performance in healthcare. An empirical analysis on the European context. J Intell Capital. 2020;34:56.
60. Gu D, Li J, Li X, Liang C. Visualizing the knowledge structure and evolution of big data research in healthcare informatics. Int J Med Informatics. 2017;1(98):22–32.
61. Dhagarra D, Goswami M, Sarma PR, Choudhury A. Big Data and blockchain supported conceptual model for enhanced healthcare coverage: The Indian context. Bus Process Manag J. 2019;67:7.

Publisher's Note